METRIC TOPOLOGY: A FIRST COURSE (FOR MATH 4450, SPRING 2011)

PAUL BANKSTON, MARQUETTE UNIVERSITY

Course Contents

The course is divided up into thirty roughly-hour-long lectures.

1. A Historical Introduction (p.2)
2. Sets and Set Operations (p.5)
3. Functions (p.8)
4. Equivalence Relations (p.11)
5. Countable and Uncountable Sets (p.13)
6. The Real Line (p.16)
7. Metric Spaces: Some Examples (p.20)
8. Open Sets and Closed Sets (p.24)
9. Accumulation Points and Convergent Sequences (p.28)
10. Interior, Closure, and Frontier (p.31)
11. Continuous Functions and Homeomorphisms (p.34)
12. Topologically Equivalent Metrics (p.38)
13. Subspaces and Product Spaces (p.40)
14. Complete Metric Spaces I (p.43)
15. Complete Metric Spaces II (p.46)
16. Quotient Spaces I (p.50)
17. Quotient Spaces II (p.54)
18. Separation Properties (p.56)
19. Introduction to Connectedness (p.59)
20. Connected Subsets of Euclidean Space (p.62)
21. Path Connectedness (p.65)
22. Local Connectedness (p.69)
23. Introduction to Compactness (p.71)
24. More on Compactness (p.74)
25. Other Forms of Compactness (p.77)
26. Countability Properties of Metric Spaces (p.80)
27. Introduction to Continua (p.84)
28. Irreducibility (p.88)
29. Cut Points (p.91)
30. Proper Subcontinua (p.94)

Further Reading (p.97)

Lecture 1: A Historical Introduction

Abstract. We give a brief historical introduction to topology, and focus on the development of Euler's famous theorem concerning spherical polyhedra: take the number of vertices, subtract the number of edges, add the number of faces, and the result is invariably 2.

Topology, the area of mathematics sometimes whimsically referred to as "rubber sheet geometry," is concerned with the study of properties of a geometric object that remain unaffected when the object is twisted, stretched, or folded (but not torn or punctured).

In high school geometry two objects are geometrically similar if one can be transformed into the other via a *geometric transformation*; i.e., a composition of linear translations, rotations, and dilations. Note that any such composition is invertible, and that its inverse is also a composition of linear translations, rotations, and dilations. Geometric transformations preserve such intuitive properties as angle measurement and straightness, but not size. For example any two circles are geometrically similar, as are any two isosceles right triangles or any two line segments. No triangle is geometrically similar to any polygon with more than three sides, however.

In topology what constitutes "similarity" is much more general, in the sense that what constitutes a "transformation" is much broader. Two objects are topologically similar if one can be continuously deformed into the other via a *topological transformation*; i.e., a continuous one-to-one correspondence whose inverse is also continuous. Another name for *topological transformation* is *homeomorphism*, and it is a goal of this course to make this idea mathematically sound and understandable.

Only a circle is geometrically similar to a given circle, but there are lots of other geometric objects that are topologically similar (i.e., homeomorphic) to it. Imagine your circle to be an elastic band. You can crimp it at various points to form a triangle or square; stretch it into a large oval; even cut it, tie it into a knot, and rejoin the ends. All these objects are homeomorphic to a circle, and all go under the heading of *simple closed curve*.

So what *isn't* homeomorphic to a circle? By the end of this course you will be able to prove mathematically that the following are not simple closed curves: line segments, figure-eight curves, spheres, and disks.

The word *topology* derives from Greek, and literally means "analysis of position." The corresponding Latin term is *analysis sitūs*, and was the more popular name for our subject early in the 20th century. How the Greek term ultimately gained prominence—from the 1920s on—is a subject for the historians of mathematics.

As an autonomous mathematical subject, topology did not really get off the ground until the late 19th century. By that time mathematics was entering its "modern" phase, characterized by being founded upon set theory. To a mathematician—especially one interested in foundations—no mathematical concept is clear until it can be framed in terms of sets. The term *set* itself, however, along with the notion of what it means for something to be a *member* of a given set, is an undefined—supposedly intuitively clear—concept. That said, if you look up *set* in the Oxford English Dictionary, you'll find several column inches devoted to its various meanings. In the context of this course, we offer the synonyms *class, family, collection, ensemble* and hope for the best. We'll have more to say about this later.

Historically, the first known topological result was proved almost four hundred years ago by none other than René Descartes of *Je pense donc je suis* fame. He was studying the classic polyhedra of antiquity—e.g., tetrahedra, cubes, octahedra, etc.—and discovered that the number F of polygonal faces, plus the number V of vertices exceeds the number E of edges by 2. (Try it out with a cube: there are six square faces and eight vertices, so $F + V = 14$. $E = 12$; *voilà*! Now try it with, say, an octahedron.)

Although we now recognize the result to be topological (for reasons to be given below), Descarte's original proof cannot be said to have been one of a topological flavor. The same assessment goes for the later (18th century) rediscovery of the result by Leonhard Euler, who formulated it as $V - E + F = 2$: both arguments made use of angle measure and relied on the straightness of the edges and the flatness of the faces.

The first truly topological version of this wonderful result is really due to Henri Poincaré in 1895. He realized its essential rubbery nature in the following way: Imagine that, on the surface of a more-or-less spherical balloon, you've marked off vertices, edges, and faces, much as if you were designing a soccer ball. The simplest soccer ball would have three vertices, each vertex would be joined to each other vertex, and there would be two curvy triangular faces (so $V = E = 3$, $F = 2$, and $V - E + F = 2$). The classic soccer ball is a bit more fancy: twelve (black) pentagonal faces and twenty (white) hexagonal faces (so $F = 32$). One easily checks that $V = 12 \times 5 = 60$ and that $E = (12 \times 5) + 30 = 90$. Thus, here too, $V - E + F = 60 - 90 + 32 = 2$. The essential observation is this: Every edge has two incident vertices (its end points) and forms part of the boundary of two faces. Suppose there's a polygonal face with $n > 3$ edges. We introduce a new vertex in the interior of that face, as well as edges connecting the new vertex to each vertex of that face. This means we've incremented V by 1, E by n, and F by $n - 1$; hence, by this process of triangulation, we have not changed the alternating sum $V - E + F$ at all. The upshot of this discussion is that we may assume, without loss of generality, that our original curvy polyhedron has triangular faces only.

Now suppose we remove one vertex and all n of its incident edges from this polyhedron with only triangular faces (e.g., tetrahedra, octahedra, icosahedra, but not cubes and standard soccer balls). Then what we have left is an n-sided polygonal "hole" in the balloon. Note that we have reduced V by 1 and both E and F by the same number n; hence $V - E + F$ has been reduced by 1. It suffices to show that this new alternating sum is 1. Now, since we have a rubbery surface with a polygonal hole in it, we can "squash" it down onto the plane. This gives us a polygon which has been subdivided into triangles, where each side of the polygon is a side of one of the triangles. Any one of these triangles has either $0, 1, 2$, or 3 edges that meet the complement of the polygon. If the number is 0, we call the triangle *interior*; otherwise it's called *exterior*. If triangle T is exterior with 3 edges bounding the complement, then it's the only triangle in the polygon, and we have $V - E + F = 3 - 3 + 1 = 1$. If T is exterior with 2 edges bounding the complement, then its removal decrements V by 1, E by 2, and F by 1; hence $V - E + F$ is decremented by $1 - 2 + 1 = 0$. If T is exterior with 1 edge bounding the complement, then its removal decrements V by 0, E by 1, and F by 1. Hence $V - E + F$ is decremented by $0 - 1 + 1 = 0$ in this case too. The process of removing exterior triangles must terminate eventually; hence whatever we started out with as the sum

$V - E + F$, we never altered it until we arrived at a single triangle, where the sum is 1. (The reader may recognize this method of argument as an informal version of the principle of mathematical induction.)

The alternating sum $V-E+F$ for spherical polyhedra is well known as the *Euler characteristic* $\chi(\mathbb{S}^2)$ of the sphere $\mathbb{S}^2$ (or of anything homeomorphic to the sphere). Poincaré defined this number for a wide variety of geometric objects, proving it to be a homeomorphism invariant.

Exercises 1. (1) What is the Euler characteristic of: (i) a line segment; (ii) a circle?

(2) A *disk* is the set of points on, or encircled by, a circle in the plane. That is, if the center is the origin and the radius is $r > 0$, then the disk of radius r, centered at the origin, is the set of real pairs $\langle x, y\rangle \in \mathbb{R}^2$ such that $x^2+y^2 \leq r^2$. Similarly, a *ball* is the set of points on, or girdled by, a sphere in real three-space $\mathbb{R}^3$. (In this case, the defining condition for triples is $x^2 + y^2 + z^2 \leq r^2$.) What is the Euler characteristic of: (i) a disk; (ii) a ball?

(3) If two geometric objects are homeomorphic, then they have the same Euler characteristic. Is the converse true?

(4) An *annulus* is the planar region consisting of points that lie between two concentric circles; i.e., if the common center is the origin and the two circles have radii $0 < r < R$, then the resulting annulus is $A = \{\langle x, y\rangle \in \mathbb{R}^2 : r^2 \leq x^2 + y^2 \leq R^2\}$. What is the Euler characteristic of an annulus?

(5) A *torus* is the bounding surface $\mathbb{T}^2$ of a donut; like an inner tube without the valve. It may also be regarded as the surface of revolution obtained when a circle lying in the upper half plane ($y > 0$) is revolved about the x-axis. Still another, less obvious, description is as the set of all real 4-tuples $\langle x, y, u, v\rangle \in \mathbb{R}^4$ such that $x^2 + y^2 = r^2$ and $u^2 + v^2 = R^2$ (where r and R are positive radii). What is the Euler characteristic of a torus?

Lecture 2: Sets and Set Operations

Abstract. We introduce the basic notions behind set theory, including some of the fundamental axioms. Boolean set operations, ordered pairs, and cartesian products are discussed.

Preliminaries. In the study of metric topology, the basic concepts are those of *metric space* and *continuous function*. This follows a familiar pattern in modern pure mathematics: one studies certain structured sets, along with "structure-respecting" functions between them. For example, in linear algebra the basic concepts are those of *vector space* and *linear transformation*. In this lecture we gather some of the basic set-theoretic notions that form the foundation for this approach.

We take the notion of *set*, and what it means for something to be an *element* (or *member*) of a set, as primitive; i.e., formally undefined. Synonyms for *set* are: *family*, *collection*, *aggregate*, and–if you want to get really fancy–*ensemble*. That said, we all have an intuitive idea of these notions; in particular we generally recognize the following statements as true:

- (Empty Set Axiom) There is a set that has no elements, it is denoted by the symbol $\emptyset$.
- (Extensionality Axiom) Two sets are equal if and only if they have exactly the same elements.
- (Pairing Axiom) Given any two sets, there is exactly one set whose elements are the two given sets.

If we accept the Extensionality Axiom, then any two sets with no elements must vacuously have the same elements; hence there is only one empty set. As for notation, if A and B are two sets (possibly the same), the set guaranteed in the Pairing Axiom is denoted $\{A, B\}$, the *unordered pair* with elements A, B. $\{A\}$ abbreviates $\{A, B\}$ when $A = B$; it is the *singleton* set whose sole element is A. (So while $\emptyset$ has no elements, $\{\emptyset\}$ has exactly one; hence $\emptyset \neq \{\emptyset\}$.) If a is an element of A, we write $a \in A$ to express this; the symbol $a \notin A$ expresses the fact that a is not an element of A.

Given sets A and B, the set $\{\{A\}, \{A, B\}\}$ is abbreviated $\langle A, B\rangle$. This set is called the *ordered pair* determined by A and B. The signal feature of ordered pairs is the following fact:

- (Ordered Pair Property) $\langle A, B\rangle = \langle C, D\rangle$ if and only if $A = C$ and $B = D$.

If we have a list $x_1, x_2, \ldots, x_n$ of the elements of A—in no special order—we may write $A = \{x_1, x_2, \ldots, x_n\}$. If we wish to highlight the set B consisting of those elements of A that satisfy a special property P, we also write $B = \{x : x \in A \text{ and } x \text{ has property } P\}$ (also denoted $\{x \in A : x \text{ has property } P\}$).

Boolean Set Operations. The basic Boolean—in honor of George Boole—set operations are union, intersection, and relative complement; parallelling the logical words, *or*, *and*, and *not*, respectively. That is, given two sets A and B, we have:

- (Union) The *union* of A and B, denoted $A \cup B$, is the set $\{x : x \in A \text{ or } x \in B\}$.
- (Intersection) The *intersection* of A and B, denoted $A \cap B$, is the set $\{x : x \in A \text{ and } x \in B\}$.
- (Relative Complement) The *complement* of B in A, denoted $A \setminus B$, is the set $\{x : x \in A \text{ and } x \notin B\} = \{x \in A : x \notin B\}$.

We say B is a *subset* of A (in symbols, $B \subseteq A$) if every element of B is also an element of A. Clearly this is equivalent to either of $A \cup B = A$ or $A \cap B = B$ holding. Note that $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$. Frequently when we wish to show two sets to be equal, we show—in two steps–that each is a subset of the other.

Two sets are said to be *disjoint* if they have empty intersection; i.e., they share no elements. The notion of disjointness of sets parallels the logical situation where two properties are mutually exclusive. For example, an integer can be odd or even, but not both.

Clearly, since every element of $\emptyset$ is also an element of any set A, it follows that $\emptyset \subseteq A$ always holds. Equally clear is the fact that A is always a subset of itself. B is then a *proper* subset of A (in symbols, $B \subset A$) if $B \subseteq A$ but $B \neq A$. (Note how this notation parallels that of $\leq$ and $<$ in the context of ordering points on the real line.)

Given a set X, the collection of subsets of X is denoted $\wp(X)$, the *power set* of X. Then $\wp(X) = \{A : A \subseteq X\}$. Since the union, intersection, and relative complement of two subsets of X is also a subset of X, the power set is endowed with algebraic structure, and is referred to as the *Boolean algebra* of subsets of X. Note that $\wp(\emptyset) = \{\emptyset\}$; and that if $X \neq \emptyset$, then $\wp(X)$ has at least two elements; i.e., $\emptyset$, X, plus all the nonempty proper subsets of X.

Here is a list of basic properties of the set operations and relations that we have considered so far.

Proposition 2.1. (1) *(Distributive Law 1)* $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.
(2) *(Distributive Law 2)* $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.
(3) *(De Morgan Law 1)* $A \setminus (B \cup C) = (A \setminus B) \cap (A \setminus C)$.
(4) *(De Morgan Law 2)* $A \setminus (B \cap C) = (A \setminus B) \cup (A \setminus C)$.

Proof. We prove (3), leaving the rest as exercises. Indeed, suppose $x \in A \setminus (B \cup C)$. Then $x \in A$, but $x \notin B \cup C$. This second condition means that x is in neither B nor C; hence both $x \notin B$ and $x \notin C$ hold. Thus $x \in A \setminus B$ and $x \in A \setminus C$, so $x \in (A \setminus B) \cap (A \setminus C)$. This tells us that $A \setminus (B \cup C) \subseteq (A \setminus B) \cap (A \setminus C)$.

For the reverse inclusion, we assume now that $x \in (A \setminus B) \cap (A \setminus C)$. Then x is in A, but is in neither B nor C; i.e., $x \in A \setminus (B \cup C)$. This shows that $(A \setminus B) \cap (A \setminus C) \subseteq A \setminus (B \cup C)$, so equality must hold.

□

Cartesian Products. If X and Y are two sets, then the *cartesian product* of these sets is denoted $X \times Y$, and is defined to be $\{\langle x, y\rangle : x \in X \text{ and } y \in Y\}$. For example, if $X = \{1, 2\}$ and $Y = \{a, b, c\}$, then $X \times Y = \{\langle 1, a\rangle, \langle 1, b\rangle, \langle 1, c\rangle, \langle 2, a\rangle, \langle 2, b\rangle, \langle 2, c\rangle\}$, while $Y \times X = \{\langle a, 1\rangle, \langle a, 2\rangle, \langle b, 1\rangle, \langle b, 2\rangle, \langle c, 1\rangle, \langle c, 2\rangle\}$. (Generally, $X \times Y$ and $Y \times X$ are unequal, but have the same number of elements.)

With $\mathbb{R}$ denoting the usual real line—we will have more to say later about the structure of $\mathbb{R}$—the cartesian product operation provides a general mechanism for obtaining the euclidean plane $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$.

We may extend the notion of ordered pair to that of *ordered n-tuple* for any positive natural number n as follows: Define the ordered triple (3-tuple) $\langle a_1, a_2, a_3\rangle$ to be the ordered pair $\langle\langle a_1, a_2\rangle, a_3\rangle$. In general, given that we know how to make ordered $(n-1)$-tuples, we define the ordered n-tuple $\langle a_1, a_1, \ldots, a_n\rangle$ to be the ordered pair $\langle\langle a_1, a_2, \ldots, a_{n-1}\rangle, a_n\rangle$. Now we have a general mechanism for obtaining euclidean n-space $\mathbb{R}^n$ as the n-fold cartesian product consisting of all n-tuples of real numbers. In general, if $X_1, X_2, \ldots, X_n$ are sets, their cartesian product $\prod_{i=1}^{n} X_i = X_1 \times X_2 \times \cdots \times X_n = \{\langle x_1, x_2, \ldots, x_n\rangle : x_i \in X_i \text{ for each } 1 \leq i \leq n\}$ is well defined.

Exercises 2. (1) Prove the Ordered Pair Property.

(2) Prove items 1,2, and 4 of Proposition 2.1.

(3) List the elements of $\wp(X)$, where $X = \{a, b, c, d\}$.

(4) If X has n elements, where $n = 0, 1, 2, \ldots$, how many elements does $\wp(X)$ have?

(5) Suppose $A \subseteq X$ and $B \subseteq Y$. Show that $A \times B = (A \times Y) \cap (X \times B)$.

(6) If $A, B \subseteq X$, show that $X \setminus (A \setminus B) = B \cup (X \setminus A)$.

(7) If A has m elements and B has n elements, where m and n are whole numbers, then how many elements does $A \times B$ have?

(8) Suppose X and Y are sets, each of which has at least two elements. Show that $X \times Y$ contains a subset that is *not* of the form $A \times B$ for any $A \subseteq X$, $B \subseteq Y$.

Lecture 3: Functions

Abstract. Functions and function composition are introduced. Also considered are images and inverse images of sets relative to a function.

The Concept of Function. In this lecture, we make more precise a notion that you have been familiar with since high school algebra, namely that of *function*. Every function consists of three pieces of data: the *domain*, or set of elements acted on by the function, or for which the function is defined; the *range* (sometimes called the *codomain*), or set of possible values the function takes; and the *rule*, or specification of the value taken by the function at each point of the domain. In symbols, $f : X \to Y$ denotes the function whose domain is X, whose range is Y, and whose rule of assignment is f. To each $x \in X$ there is a unique $y \in Y$ assigned to x by f, and we write $y = f(x)$.

In light of the notion of ordered pair and of cartesian product presented in Lecture 2, we see that the rule f consiste of ordered pairs $\langle x, y\rangle \in X \times Y$ such that $y = f(x)$. That is, a function from X to Y may be viewed as a subset $f \subseteq X \times Y$ satisfying the following two conditions:

- (Existence) For each $x \in X$, there is at least one $y \in Y$ such that $\langle x, y\rangle \in f$.
- (Uniqueness) If $\langle x, y_1\rangle$ and $\langle x, y_2\rangle$ are both in f, then $y_1 = y_2$.

The awkward and unintuitive notation $\langle x, y\rangle \in f$ may be replaced by the more natural $y = f(x)$ because: (1) every $x \in X$ has at least one $y \in Y$ assigned to it, i.e., f is *defined* at x (existence); and (2) f is never *multiply* defined at x (uniqueness).

Examples 3.1. (1) In calculus, when we write $f(x) = x^2$, the domain and range are implicitly understood to be the same set $\mathbb{R}$. Of course not every $y \in \mathbb{R}$ is of the form $f(x)$; no negative real is the square of a real. When we wish to specify exactly those $y \in Y$ that appear as values under f, this subset of Y is called the *image of X under the function f*. In symbols, the image of f is $\{y \in Y : y = f(x) \text{ for some } x \in X\}$. So in our squaring example, the image is the set of nonnegative reals, $[0, \infty)$.

(2) When we write $f(x) = \sqrt{x}$ in the context of functions of one real variable, the domain is no longer $\mathbb{R}$, but just the interval $[0, \infty)$. And since every $x > 0$ has exactly *two* square roots, one positive and one negative, we customarily choose the positive one to call $\sqrt{x}$, avoiding multiple valuedness. (After all, the other square root may be expressed as $-\sqrt{x}$.) However, the squaring operation may also be carried out for complex numbers, and it is now true that every nonzero complex number has exactly two complex square roots. Here it is still possible—but a bit trickier—to define $\sqrt{x}$.

When $f : X \to Y$ is a function and $A \subseteq X$, we denote by $f[A]$ the *image* of A under f; it is defined to be the subset $\{y \in Y : y = f(x) \text{ for some } x \in A\} = \{f(x) : x \in A\}$. (Note the difference in meaning between the notations $f[A]$ and $f(x)$.)

When $B \subseteq Y$, we denote by $f^{-1}[B]$ the *pre-image* of B under f; it is defined to be the subset $\{x \in X : f(x) \in B\}$.

Example 3.2. So in the case of $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$, we have: (1) $f[[-1,2)] = [0,4)$; (2) $f^{-1}[[-1,4)] = (-2,2)$; and (3) $f^{-1}[[1,2]] = [-\sqrt{2},-1] \cup [1,\sqrt{2}]$.

A function $f : X \to Y$ is termed *surjective* (or *onto*) if $Y = f[X]$; it is termed *injective* (or *one-to-one*) if no two elements of X are sent to the same value in Y. For $y \in Y$, we abbreviate $f^{-1}[\{y\}]$ by $f^{-1}[y]$. So when f is surjective, $f^{-1}[y] \neq \emptyset$ for each $y \in Y$; and when f is injective, $f^{-1}[y]$ has at most one element for each $y \in Y$. (Note that $f^{-1}[y] = \emptyset$ means that y is not in the image of f.)

If $f : X \to Y$ is both surjective and injective, it is called a *bijection*, or a *one-to-one correspondence*. A bijection $f : X \to Y$ provides a bi-unique pairing between the elements of X and those of Y. For example, X could be the set of fingers on a human hand and Y could be the set of points of a five-pointed star. There are lots (5!=120) of bijections between X and Y.

Proposition 3.3. *Let $f : X \to Y$ be a function, with $A_1, A_2 \subseteq X$ and $B_1, B_2 \subseteq Y$.*

(1) $f[A_1 \cup A_2] = f[A_1] \cup f[A_2]$.
(2) $f[A_1 \cap A_2] \subseteq f[A_1] \cap f[A_2]$*; the reverse inclusion holds if f is injective, but need not hold in general.*
(3) $f^{-1}[B_1 \cup B_2] = f^{-1}[B_1] \cup f^{-1}[B_2]$.
(4) $f^{-1}[B_1 \cap B_2] = f^{-1}[B_1] \cap f^{-1}[B_2]$.

Proof. We prove (2), leaving the rest for exercises. Suppose $y \in f[A_1 \cap A_2]$. Then $y = f(x)$ for some $x \in A_1 \cap A_2$. Since $x \in A_1$ and $x \in A_2$, we have $f(x) \in f[A_1]$ and $f(x) \in f[A_2]$, so $y = f(x) \in f[A_1] \cap f[A_2]$.

Now assume f is injective and $y \in f[A_1] \cap f[A_2]$. Since $y \in f[A_1]$, there is some $x_1 \in A_1$ such that $y = f(x_1)$. And since $y \in f[A_2]$, there is some $x_2 \in A_2$ such that $y = f(x_2)$. Since f is injective, we know $x_1 = x_2$, so there is some $x \in A_1 \cap A_2$ such that $y = f(x)$. Thus $y \in f[A_1 \cap A_2]$.

To show the reverse inclusion does not hold in general, we must produce a *counterexample*; i.e., an example of a function for which (2) fails. So let's take f to be the squaring function on the real line, let $A_1 = [-2,-1]$, and let $A_2 = [1,2]$. Then $f[A_1] \cap f[A_2] = [1,4] \cap [1.4] = [1,4]$. However, $A_1 \cap A_2 = \emptyset$; so $f[A_1 \cap A_2] = \emptyset$ as well. Hence $f[A_1 \cap A_2]$ can be a proper subset of $f[A_1] \cap f[A_2]$.

□

If we view a function $f : X \to Y$, as an ordered triple, then we may alter the function without altering the rule. For example, if $Z \subseteq Y$ contains the image $f[X]$, then the function $f : X \to Z$ results from the original function by *restriction of the range*. In particular, any injective function becomes a bijection when the range is restricted to the image.

Another way of altering the function without altering the rule is by *restricting the domain*. Starting with $f : X \to Y$, if $A \subseteq X$, then the restriction $f|A : A \to Y$ is defined by the rule $(f|A)(x) = f(x)$ for $x \in A$. (In setting up max/min problems in calculus, you often find that you want to optimize $f(x)$, subject to a "real-world" constraint; e.g., $x \geq 0$. So while $f(x)$ may be formally defined for negative x, the problem makes sense only when the function has restricted domain.)

Function Composition. If $f : X \to Y$ and $g : Y \to Z$, where the range of f equals the domain of g, then we may form the *composition* $g \circ f : X \to Z$, given by the rule of assignment $(g \circ f)(x) = g(f(x))$. (Recall the Chain Rule from calculus: it shows you how to differentiate the composition of two differentiable functions.) For any set X, the *identity function* $i_X : X \to X$ is defined by $i_X(x) = x$. If $f : X \to Y$ is a bijection, then there is a unique function $g : Y \to X$, the *function inverse* of f, such that $g \circ f = i_X$ and $f \circ g = i_Y$. In this situation, we denote g by f^{-1}. (Note that if f is bijective and $y = f(x)$, then $f^{-1}(y) = x$, while $f^{-1}[y] = \{x\}$.)

Exercises 3.

(1) Prove items 1,3, and 4 of Proposition 3.3.
(2) Let $f : X \to Y$, with $A \subseteq X$ and $B \subseteq Y$. Prove that $A \subseteq f^{-1}[f[A]]$ and that $f^{-1}[Y \setminus B] = X \setminus f^{-1}[B]$.
(3) If $f : X \to Y$, then $g : Y \to X$ is a *left inverse* for f if $g \circ f = i_X$; g is a *right inverse* for f is $f \circ g = i_Y$. Show that f is injective if and only if f has a left inverse and that f is surjective if and only f has a right inverse.
(4) Show that the composition of two injective (resp., surjective, bijective) functions is again injective (resp., surjective, bijective).
(5) If $f : A \to Y$ is a function and $A \subseteq X$, a function $F : X \to Y$ is an *extension* of f if $f = F|A$. Find two distinct extensions of the square root function from $A = [0, \infty)$ to $X = \mathbb{R}$.
(6) Let $f : X \to Y$, with $A \subseteq X$ and $B \subseteq Y$. Show that $f[A \cap f^{-1}[B]] = f[A] \cap B$.
(7) Let $f : X \to Y$, $g : Y \to Z$, with $B \subseteq Z$. Show that $(g \circ f)^{-1}[B] = f^{-1}[g^{-1}[B]]$.

Lecture 4: Equivalence Relations

Abstract. Binary relations and their main properties are introduced. Among the most important of binary relations on a set are partial orders and equivalence relations. An equivalence relation on a set induces—and is induced by—a partition of the set into pairwise disjoint nonempty subsets.

Binary Relations. Recall from Lecture 2 that we viewed the rule of assignment f of a function $f : X \to Y$ to be a subset of $X \times Y$. In general a subset R of $X \times Y$ is called a *relation* from X to Y. Since the notation $y = R(x)$ does not make sense in this more general context, we write xRy to indicate that $\langle x, y\rangle \in R$. A given $x \in X$ may belong to any number (zero included) of $y \in Y$. Often we write xR to denote the set of $y \in Y$ such that xRy; similarly, $Ry = \{x \in X : xRy\}$.

Examples 4.1. (1) X is a nonempty set, $Y = \wp(X)$, and $R = \{\langle x, A\rangle \in X \times Y : x \in A\}$.
(2) $X = Y = \mathbb{R}$, and $R = \{\langle x, y\rangle : x < y\}$.
(3) $X = Y = \mathbb{Z}$, where $\mathbb{Z}$ is the set of integers, and $R = \{\langle m, n\rangle : m - n \text{ is a multiple of } 3\}$.

Properties of Relations on a Set. If $R \subseteq X \times X$, R is called a *binary relation on* X. Here is a list of important properties that pertain to binary relations on a single set.

- (Reflexitivity) xRx for all $x \in X$.
- (Irreflexivity) If xRy, then $x \neq y$, for all $x, y \in X$.
- (Symmetry) For any $x, y \in X$, if xRy then yRx.
- (Anti-symmetry) For any $x, y \in X$, if xRy and yRx, then $x = y$.
- (Transitivity) For any $x, y, z \in X$, if xRy and yRz, then xRz.

$R \subseteq X \times X$ is a *partial order* if it satisfies reflexivity, anti-symmetry, and transitivity (e.g., the relation $\subseteq$ defined on $\wp(X) \times \wp(X)$). R is the *adjacency* relation for a *simple graph* if it satisfies irreflexivity and symmetry (e.g., elements of X are *vertices*, and when xRy we draw an "edge" joining x to y). Finally R is an *equivalence relation* if it satisfies reflexivity, symmetry, and transitivity.

Equivalence Relations and Partitions. While all kinds of binary relations come up in the study of topology, the most fundamental are the equivalence relations. Suppose $R \subseteq X \times X$ is one such. For $x \in X$, the sets xR and Rx are identical, due to symmetry, and together denote the *equivalence class* of x. So while an element determines its equivelance class, the equivalence class does not determine the element: xR may equal yR with x and y quite distinct. x is called a *representative* of the equivalence class yR if xRy.

The set of equivalence classes is denoted X/R. By reflexivity, we have $x \in xR$; and, by transitivity and symmetry, if $z \in xR \cap yR$, then $xR = yR$. (Exercise.) Thus X/R is a collection of nonempty subsets of X such that each element of X

lies in exactly one set in X/R. Such a family of subsets of a set is called a *partition* of the set.

So an equivelence relation R on a set naturally gives rise to a partition $\mathcal{P}_R = X/R$ on the set. Conversely, if $\mathcal{P} \subseteq \wp(X)$ is a partition of X, then we may define the relation $R_{\mathcal{P}}$ by the condition that $xR_{\mathcal{P}}y$ just in case x and y lie in the same member of $\mathcal{P}$. The relation $R_{\mathcal{P}}$ is easily shown (exercise) to be an equivalence relation on X; what is more, the functions $R \mapsto \mathcal{P}_R$ and $\mathcal{P} \mapsto R_{\mathcal{P}}$ define a bijection between equivalence relations on X and partitions of X. The following is a list of important equivenence relations in mathematics.

Examples 4.2. (1) On any set X there is the *coarsest* equivalence relation, namely the one that makes everything equivalent to everything else. In this case there is only one equivalence class, namely X itself. At the opposite end, there is the *finest* equivalence relation, namely the one that relates x to y just in case $x = y$. In this case the equivalence classes are the singletons $\{x\}$, for $x \in X$.

(2) As in Example 4.1(3) above, $X = \mathbb{Z}$ is the set of integers and mRn means that $m - n$ is a multiple of 3. This is an equivalence relation; the R-equivalence class of m is marked by the remainder—0,1, or 2—when m is divided by 3. Thus $\mathbb{Z}/R =$ $\{\{\dots,-6,-3,0,3,6,\dots\},\{\dots,-5,-2,1,4,7,\dots\},\{-4,-1,2,5,8,\dots\}\}$.

(3) $X = \mathbb{I}$ is the closed unit interval in the real line, and we define xRy just in case either $x = y$ or $\{x, y\} = \{0, 1\}$. The equivalence classes are the singletons $\{x\}$ for $0 < x < 1$, plus the doubleton $\{0, 1\}$. The end points 0 and 1 are said to be "identified," and the resulting geometric figure may be thought of as the result of "gluing" the ends of a piece of string to form a simple closed curve. We'll have more to say about this "cutting and pasting" process later.

Exercises 4. (1) Prove that two distinct equivalence classes are disjoint.

(2) Two people are siblings if they share both parents, and are cousins if they share one set of grandparents. Which of the relation properties listed above pertain to each of these kinship relations?

(3) For $x, y \in \mathbb{R}$ define xRy by the condition that $|x - y| < 1$. Which of the relation properties listed above pertain to R?

(4) For subsets A, B of X, define ARB just in case there is a bijection between A and B. Show this is an equivalence relation.

(5) If R is an equivalence relation on a set X, define the function $q : X \to X/R$ by the assignment $q(x) = xR$. q is called the *natural quotient function* associated with R; show that q is always surjective, and is injective just in case R is the equality relation on X.

(6) How many distinct equivalence relations are there on a set with two elements? Three elements? Four? (The general problem for n-element sets involves an application of the inclusion-exclusion principle of combinatorics.)

Lecture 5: Countable and Uncountable Sets

Abstract. Sets can be either finite or infinite; two sets are equinumerous if they can be put in one-to-one correspondence with each other. The lowest level of infinity for a set is to be countably infinite; i.e., to be equinumerous with the natural numbers. Sets that are infinite but not countable are called uncountable. We show that the sets of integers and of rational numbers are countably infinite.

The Natural Numbers. As mentioned in earlier lectures, modern mathematics interprets all "mathematical objects" as sets. What about numbers? The answer is that numbers too can be treated as sets, and here is how we do it. We start with the natural numbers, the ones we count with.

Although zero came rather late in the history of counting numbers—even the sophisticated ancient Romans had no symbol for zero—it is clearly the most basic: one interpretation of zero is "nothing;" a set with zero elements is empty. So we *define* zero to be the empty set; i.e., $0 = \emptyset$. Now, to define the rest of the numbers, we bootstrap ourselves from one number to its immediate successor by the inductive condition $n^+ = n \cup \{n\}$. I know this looks a bit weird, but let's look at some examples. Since we're grounded at $0 = \emptyset$, we proceed to define 1 as $1 = 0^+ = 0 \cup \{0\} = \emptyset \cup \{\emptyset\} = \{\emptyset\}$. Similarly, $2 = \{0, 1\} = \{\emptyset, \{\emptyset\}\}$, $3 = \{0, 1, 2\}$, etc. In general, each number is the *set* of its predecessors.

Finally, to complete the process, we need a set-forming axiom.

- (Axiom of Infinity) There is a set X such that $n \in X$ for each natural number n.

The unique set that contains each natural number *and nothing else* is denoted $\mathbb{N}$, and we write $\mathbb{N} = \{0, 1, 2, \dots\}$. The natural order on $\mathbb{N}$ is given by $0 < 1 < 2 < \dots$, and addition of natural numbers is given inductively. Briefly, $n + 0$ is defined to be n; and, given that we know how to define $n + m$, we define $n + (m^+)$ to be $(n + m)^+$. If $m \leq n$, then there is a unique $d \in \mathbb{N}$ such that $m + d = n$ (exercise), and we denote this d by $n - m$. The rest of the arithmetic of natural numbers is just a careful rewording of fourth grade mathematics.

The Integers. How we define the set of integers from the set of natural numbers parallels how one explains the notion of "signed number" to a high school algebra student; it's just a little more sophisticated.

First we settle on two special sets, $+$ and $-$, which we call *signs*. Then, for each positive natural number $n \in \mathbb{N} \setminus \{0\}$, we create the two ordered pairs $\langle +, n\rangle$ and $\langle -, n\rangle$. (In future we'll write n for $\langle +, n\rangle$ and $-n$ for $\langle -, n\rangle$. The pairs $\langle +, n\rangle$, for $n \in \mathbb{N} \setminus \{0\}$, are called *positive* integers, and we denote this set $\mathbb{Z}^+$. The pairs $\langle -, n\rangle$, for $n \in \mathbb{N} \setminus \{0\}$, are called *negative* integers, and we denote this set $\mathbb{Z}^-$. Finally, the set of *integers* is denoted $\mathbb{Z}$, and is the union $\mathbb{Z}^- \cup \{0\} \cup \mathbb{Z}^+$. $\mathbb{Z}$ acquires its natural order in the usual way; also its arithmetic: integers of the same sign—or zero—are added as if in $\mathbb{N}$, but with signs attached; adding $\langle +, m\rangle$ to $\langle -, n\rangle$ yields 0 if $m = n$, $\langle +, m - n\rangle$ if $m > n$, and $\langle -, n - m\rangle$ if $m < n$.

The Rational Numbers. In a calculus course you normally define a rational number to be a "fraction" m/n, where $m, n \in \mathbb{Z}$, $n \neq 0$; and this is a perfectly natural way to think of such a number, i.e., as a ratio, but this doesn't tell us what rational numbers *are* as sets. Again we resort to the cartesian product construction and—for the moment—represent a rational number as an ordered pair $\langle m, n\rangle \in \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$. We're only part way there when we do this, though, because the fractions $1/2$ and $2/4$ are equal ratios, while the pairs $\langle 1, 2\rangle$ and $\langle 2, 4\rangle$ are formally distinct. In order to account for this, we define an equivelence relation, using what we know about how we determine when $m/n = p/q$. So given $\langle m, n\rangle, \langle p, q\rangle \in \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$, we define $\langle m, n\rangle \sim \langle p, q\rangle$ just in case $mq = pn$ (cross multiplication). Letting m/n denote the equivalence class $\langle m, n\rangle \sim$, we then see that, say, $1/2$ formally consists of all pairs $\langle m, n\rangle \in \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ such that $n = 2m$. We denote by $\mathbb{Q}$ the set of such equivalence classes m/n; and in practice, we think of each "fraction" in the traditional way, as a point on a number line.

Defining the usual order and arithmetic on $\mathbb{Q}$ brings up the issue of relating equivalence classes on a set using representatives of those classes. Of course we decide whether $m/n < p/q$ by testing whether $mq < np$; the only thing to make sure of—and this is easy to do—is whether it is also true that $m'/n' < p'/q'$ whenever $\langle m, n\rangle \sim \langle m', n'\rangle$ and $\langle p, q\rangle \sim \langle p', q'\rangle$. We take similar care when we add and multiply fractions. The assignment $n \mapsto m/1$ shows how to think of integers as special rational numbers; those with sufficient algebraic background will recognize that the construction of $\mathbb{Q}$ as an expansion of $\mathbb{Z}$ is the template for the "field of fractions" construction that applies to any commutative ring with multiplicative identity.

Cardinality. If there is a bijection between sets X and Y, we say X and Y are *equinumerous*, or have the *same cardinality*, and write $X \equiv Y$. Clearly equinumerosity is an equivalence relation between sets. If $X \equiv n = \{0, 1, \ldots, n-1\}$, we say that X is *finite, with cardinality* n and write $|X| = n$. If $X \equiv \mathbb{N}$, we say X is *countably infinite*, and we customarily write $|X| = \aleph_0$ ("aleph-null," a notation introduced by Georg Cantor, the founder of set theory). X is termed *countable* if it is either finite or countably infinite, and we often state this by writing $|X| \leq \aleph_0$. X is *uncountable* if it is not countable; i.e., X cannot be put in one-to-one correspondence with a countable set.

We list two basic facts about cardinality, leaving their verification to the reader.

Proposition 5.1. (1) *Any subset of a finite set is finite; any subset of a countable set is countable.*

(2) *Any function image of a finite set is finite; any function image of a countable set is countable.*

Examples 5.2. (1) We show that $\mathbb{Z}$ is countably infinite. Indeed a bijection $f : \mathbb{N} \to \mathbb{Z}$ may be set up as follows: for each even $n \in \mathbb{N}$, define $f(n) = n/2$; and for each odd $n \in \mathbb{N}$, define $f(n) = -(n+1)/2$.

(2) We show that $\mathbb{Q}$ is countably infinite. First of all, since each integer has a canonical representation as a rational number, $\mathbb{Q}$ contains an infinite set and is therefore infinite. Since, by definition, $\mathbb{Q}$ is the image under the

quotient function associated with an equivalence relation on $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$, it suffices to show that the cartesian product of two countable sets is countable. So suppose $X = \{x_0, x_1, \dots\}$ and $Y = \{y_0, y_1, \dots\}$ are two countably infinite sets. We set up a bijection between $\mathbb{N}$ and $X \times Y$ using "small diagonals" as follows: For each $k \in \mathbb{N}$, there are $k+1$ pairs $\langle x_m, y_n \rangle$ such that $m+n = k$; so send 0 to $\langle x_0, y_0 \rangle$, 1 and 2 to the two pairs where $m+n = 1$, 3,4, and 5 to the three pairs where $m+n = 2$, etc. The reader is invited to put the details into this process.

The result that $\mathbb{Q}$ is countable may seem a bit surprising; indeed it was surprising to Cantor, the one to discover the result. One tends to think of countable sets as having a first element, then a second element, and so on. But when you look at the rational line, there's no way to get from one rational number to the "next" because there isn't a next: once you've skipped from one rational number to another , you've skipped over infinitely many in between. The way out of this conundrum is to discard the natural order on the rational line for the purposes of counting. If $\{q_0, q_1, \dots\}$ is a listing of the elements of $\mathbb{Q}$, there's no way of telling *a priori* how q_m and q_n are related in the natural order.

In the next lecture we will study the real line, showing—among other things—that it is uncountable.

Exercises 5. (1) Show that the exponential function $f(x) = e^x$ defines a bijection between $\mathbb{R}$ and the interval $(0, \infty)$.
(2) Show that the tangent function $f(x) = \tan x$ defines a bijection between $\mathbb{R}$ and the interval $(-\pi/2, \pi/2)$.
(3) Using the result of Exercise (2) above, show that if $a < b$ are real numbers, the interval (a, b) is equinumerous with the real line.
(4) Extending Exercise (3) above, show that any two nondegenerate (i.e., with more than one point) intervals in $\mathbb{R}$ are equinumerous.
(5) Show that if $X_1, X_2, \dots, X_n$ is a finite list of countable sets, then the cartesian product $X_1 \times X_2 \times \cdots \times X_n$ is countable.

Lecture 6: The Real Line

Abstract. We show that a set and its power set can never be equinumerous, and that the real line is equinumerous with the power set of the natural numbers. This shows the real line to be an uncountable set. We also look at the real line as a system, with order properties and algebraic properties.

Power Sets. We begin this lecture with a "negative" result, one that denies the existence of something; i.e., a surjection from any set to its own power set. From this it follows that a set can never be equinumerous with its power set.

Proposition 6.1. *Let X be a set, with $f : X \to \wp(X)$ a function. Then f is not a surjection. In particular, if X is countably infinite, then $\wp(X)$ is uncountable.*

Proof. For any $x \in X$, $f(x)$ is a subset of X; hence it makes sense to say that $x \in f(x)$ or $x \notin f(x)$. Let $A = \{x \in X : x \notin f(x)\}$. Then $A \in \wp(X)$; we show that A is not in the image of f, thereby showing that f cannot be surjective. Here is were we strive for a proof by *contradiction*, also called *reductio ad absurdum*; i.e., we assume the negation of what we're trying to prove and attempt to prove we're the Pope.

So assume A is in the image of f, so $A = f(a)$ for some $a \in X$. Then either $a \in A$ or $a \notin A$, so let's assume $a \in A$. Then since A consists of precisely those $x \in X$ such that $x \notin f(x)$, it follows that $a \notin f(a)$; i.e., $a \notin A$. But what if $a \notin A$? Then again, by the definition of A, we infer $a \in f(a)$; i.e., $a \in A$. This is a contradiction if ever there was one, so we infer that A is not of the form $f(a)$ at all.

Since the assignment $x \mapsto \{x\}$ gives an injective function from X to $\wp(X)$, it follows that $\wp(X)$ can never be finite without X being finite as well. Hence if X is countably infinite, $\wp(X)$ is infinite and cannot be put in one-to-one correspondence with a countable set. $\wp(X)$ is therefore uncountable.

□

In view of the facts that: (i) there is always a one-to-one function from a set X onto a subset of $\wp(X)$ (the assignment $x \mapsto \{x\}$), and (ii) no function from X to $\wp(X)$ can be surjective, it makes sense to think of the cardinality of $\wp(X)$ as being "strictly larger" than that of X. A great deal of modern set theory is concerned with comparing cardinalities of various sets.

The Cardinality of $\mathbb{R}$. We show that $\mathbb{R}$ may be put in one-to-one correspondence with $\wp(\mathbb{N})$—proved above to be an uncountable set—via the following bijection, which we describe informally in a few stages. First, using the tangent function, followed by one of the form $y = mx + b$ ($m \neq 0$), we show $\mathbb{R}$ is equinumerous with the open unit interval $(0, 1)$. Second, using well-known properties of the real line, we may express each $0 < t < 1$ as a binary expansion $.a_1a_2\cdots = \sum_{n=1}^{\infty} a_n 2^{-n}$, where each a_n is either 0 or 1. Moreover, this expansion is unique if we eliminate terminal strings of 1s; e.g., if we replace, say, $.0001111\ldots$ with $.001000\ldots$. Third, we note that these unique binary expansions $.a_1a_2\ldots$ can be put in one-to-one correspondence with sequences $\langle a_1, a_2, \ldots \rangle$ of infinite binary sequences with no terminal strings of 1s. Finally, each subset $A \subseteq \mathbb{N}$ is uniquely represented by its *characteristic function*; i.e., $f_A : \mathbb{N} \to \{0, 1\}$, defined by $f_A(n) = 1$ if and only if $n \in A$. The only hitch is that we must consider just those sets $A \subseteq \mathbb{N}$ whose characteristic functions have no terminal strings of 1s; i.e., sets not of the form $\{n, n+1, n+2, \dots\}$. But these sets are determined by their first elements; hence they are countable in number. There must be just as many subsets of $\mathbb{N}$ not of this form as there are subsets of $\mathbb{N}$.

The Structure of the Real Line. Of course in setting up the bijection above between $\mathbb{R}$ and $\wp(\mathbb{N})$, we have been assuming some basic facts about the real line that we have been given in calculus courses only at an intuitive level. The real story is much more interesting, and owes most of its inspiration to the nineteenth century German mathematician Richard Dedekind. While there are other ways to achieve a precise construction of the real line from the rational line, his is the most elegant.

First we define a *cut* on the rational line $\mathbb{Q}$ to be a pair $\langle A, B\rangle$ of subsets of $\mathbb{Q}$ such that:

- Both A and B are nonempty.
- A has no greatest element.
- $A \cap B = \emptyset$, $A \cup B = \mathbb{Q}$.
- If $a \in A$ and $c \leq a$, then $c \in A$.
- If $b \in B$ and $c \geq b$, then $c \in B$.

So if $\langle A, B\rangle$ is a cut, then each element of A is less than each element of B; cuts are designed to "fill the gaps" in the rational line.

If q is any rational number, then q determines the cut $\langle\{r \in \mathbb{Q} : r < q\}, \{r \in \mathbb{Q} : r \geq q\}\rangle$; however there are lots of cuts that do not arise in this way.

Example 6.2. There is no rational number whose square is 2. Like Proposition 6.1, this is also proved by contradiction. We start by assuming there is a rational $q \in \mathbb{Q}$ such that $q^2 = 2$. We may assume that $q = m/n$, represented using integers in lowest terms. That is, m and n have no common factors other than 1. Then we have $m^2 = 2n^2$, so m^2 is even. But the square of an odd number is odd; hence m must be even too. That means m^2 is divisible by 4; hence n^2 is even. But then n is even, contradicting our assumption that m and n are in lowest terms.

Now let us form the cut $\langle A, B\rangle$, where $A = \{q \in \mathbb{Q} : q \leq 0 \text{ or } q^2 < 2\}$ and $B = \{q \in \mathbb{Q} : q \geq 0 \text{ and } q^2 \geq 2\}$. This cut is how we define $\sqrt{2}$, an *irrational* real number.

So a real number is *defined* to be a cut as described above. How we tell whether $\langle A, B\rangle$ lies to the left of cut $\langle A', B'\rangle$ is to check whether $A \subset A'$ (equivalently, whether $B \supset B'$). Every rational number is a real number because it determines its own cut; how one extends the algebraic operations of addition and mutiplication from $\mathbb{Q}$ to $\mathbb{R}$ is also not terribly difficult. (The reader is invited to look this part up in the literature.)

What is important for our purposes is that the usual algebraic and order-theoretic structure of the real line makes it into a *complete Archimedean ordered field*. Let's consider briefly what these words mean before going further. First, the usual addition and multiplication operations on $\mathbb{R}$ satisfy the *field* axioms:

- The *associative* laws $(x+y)+z = x+(y+z)$ and $(xy)z = x(yz)$ hold.
- The *commutative* laws $x+y = y+x$ and $xy = yx$ hold.
- The *distributive* law, $x(y+z) = xy + xz$, of multiplication over addition holds.
- 0 and 1 are, respectively, the *additive* and *multiplicative identity elements*; i.e., the laws $x + 0 = x = x1$ hold.
- Every real number x has an *additive inverse* $-x$; i.e., the law $x + (-x) = 0$ holds.
- Every nonzero real number x has an *multiplicative inverse* x^{-1}; i.e., the law $xx^{-1} = 1$ holds.

Next, there are the *order* axioms, most conveniently given by saying there is a designated subset P of $\mathbb{R}$, the set of *positive* elements of the field, satisfying:

- $1 \in P$ and $0 \notin P$.
- If x and y are in P, then so are $x + y$ and xy.
- If $x \neq 0$, then either $x \in P$ or $-x \in P$.

We write $x > 0$ as an abbreviation for $x \in P$, and $x < y$ as an abbreviation for $y - x > 0$. Finally we have the two axioms for ordered fields that, in fact, characterize the real line:

- (Archimedean Property) Given any real number x, there is a natural number n such that $x \leq n$. Equivalently, given any $x > 0$, there is a natural number $n > 0$ such that $\frac{1}{n} < x$. (In the context of abstract ordered fields, it is more proper to say that $x \leq (1 + \cdots + 1)$, where we take the sum of n copies of the multiplicative identity element.)
- (Least Upper Bound (Completeness) Property) Given any nonempty subset $X \subseteq \mathbb{R}$, if X is *bounded above*; i.e., if there is some *upper bound* for X, an element b such that $x \leq b$ for all $x \in X$, then there is a *least* upper bound b_0 for X (i.e., b_0 is an upper bound for X, but no smaller real number is an upper bound for X).

Remark 6.3. (For the reader with some algebraic background.) The real line is unique as a complete Archimedean ordered field. That is, if F is any ordered field with addition $+_F$, multiplication $*_F$, identity elements 0_F and 1_F, and set P_F of positive elements, then there is a one-to-one correspondence φ between $\mathbb{R}$ and F such that both φ and φ^{-1} are *ordered field isomorphisms*. I.e., $\varphi(0) = 0_F$, $\varphi(1) = 1_F$, $\varphi(x+y) = \varphi(x) +_F \varphi(y)$, $\varphi(xy) = \varphi(x) *_F \varphi(y)$, and $\varphi(x) \in P_F$ for all $x > 0$.

Exercises 6. (1) Formulate a "greatest lower bound axiom," and show that it follows from—and is followed by—the least upper bound axiom above.

(2) Show that no $X \subseteq \mathbb{R}$ can have two distinct least upper bounds.

(3) The least upper bound of $X \subseteq \mathbb{R}$, when it exists, is denoted $\text{lub}(X)$. If $a < b$, what are $\text{lub}([a, b))$ and $\text{lub}([a, b])$?

(4) Show that the pair $\langle A, B \rangle$ in Example 6.2 is a cut.

(5) (For those with the algebraic background.) Show that the field of complex numbers is not orderable. I.e., given the algebraic operations of plus and times, there is no possibility of finding a subset P for which the order axioms above hold.

Lecture 7: Metric Spaces: Some Examples

Abstract. A metric space is a set endowed with a distance function satisfying certain basic properties. We introduce the concept and provide some familier—and some not-so-familiar—examples.

Measuring Distance. The fundamental idea behind the notion of metric space is that we have a measure of the distance between two elements—called *points*—of a set. This measure is a nonnegative real number, and is given as a real-valued function with pairs of points as arguments.

Definition 7.1. A *metric space* is a pair $\langle X, d \rangle$, where:

(1) X is a set of elements, called *points* of the metric space; the set X is termed the *underlying set* of the metric space.
(2) $d : X \times X \to \mathbb{R}$ is a designated *metric* (or *distance function*). All metrics are assumed to satisfy the following three conditions, for pairs from X:
   - (Positivity) $d(x, y) \geq 0$; $d(x, y) = 0$ if and only if $x = y$.
   - (Symmetry) $d(x, y) = d(y, x)$.
   - (Triangle Inequality) $d(x, z) \leq d(x, y) + d(y, z)$.

The basic intuition for metric spaces is that: (i) there are no negative distances; (ii) any point is distance zero from itself, and any two distinct points are a positive distance from each other; (iii) it's just as far from point x to point y as it is from y to x; and (iv) it's at least as far to go from point x to point z if you take a detour through a third point y as it is to go from x to z directly. The rest of this lecture is devoted to examples.

The Euclidean Metric. The most familiar metric spaces are the *euclidean spaces* $\mathbb{R}^n$, the ordered n-tuples of real numbers, equipped with the *euclidean metric*. This metric is best defined, as in a linear algebra course, using the natural inner product.

If $x = \langle x_1, \ldots, x_n \rangle$ and $y = \langle y_1, \ldots, y_n \rangle$, then their *inner* (or *dot*) product is given by the formula

$$x \cdot y = \sum_{i=1}^{n} x_i y_i.$$

The inner product gives rise to the *euclidean norm*

$$\|x\| = \sqrt{x \cdot x},$$

which in turn yields the *euclidean metric*

$$d_E(x, y) = \|x - y\|,$$

where $x \pm y$ is just vector sum/diference.

The euclidean metric, a.k.a. *euclidean distance*, is the one you see in calculus courses, one based on the classical Pythagorean Theorem.
Positivity and symmetry are obvious; as for the triangle inequality, we have the following trio of classical theorems.

Proposition 7.3. *Let $x = \langle x_1, \ldots, x_n \rangle$, $y = \langle y_1, \ldots, y_n \rangle$, and $z = \langle z_1, \ldots, z_n \rangle$ be points in $\mathbb{R}^n$.*

1. *(Cauchy-Schwarz Inequality) $|x \cdot y| \leq \|x\|\|y\|$.*
2. *(Minkowski's Inequality) $\|x + y\| \leq \|x\| + \|y\|$.*
3. *(Euclidean Triangle Inequality) $d_E(x, z) \leq d_E(x, y) + d_E(y, z)$.*

Proof. The Cauchy-Schwarz Inequality is a standard result of a multi-variable calculus or linear algebra course, and is left to the reader.

The Minkowski Inequality follows from Cauchy-Schwarz as follows:

$$\|x + y\|^2 = \sum_{i=1}^{n}(x_i + y_i)^2 = \sum_{i=1}^{n}(x_i^2 + 2x_iy_i + y_i^2) = \sum_{i=1}^{n} x_i^2 + 2\sum_{i=1}^{n} x_iy_i + \sum_{i=1}^{n} y_i^2 =$$

$$\|x\|^2 + 2(x \cdot y) + \|y^2\| \leq \text{ (Cauchy-Schwarz) } \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2.$$

We retain the inequality by taking square roots because we're dealing with nonnegative quantities.

As for the Euclidean Triangle Inequality, we have

$$d_E(x, z) = \|x - z\| = \|(x - y) + (y - z)\| \leq \text{ (Minkowski) } \|x - y\| + \|y - z\| = d_E(x, y) + d_E(y, z).$$

□

The Taxicab Metric. This metric is also defined on $\mathbb{R}^n$, and is best motivated by letting $n = 2$ and imagining the plane to be set out in a grid of "streets" running east-west and "avenues" running north-south. To get from one point to another, you have to follow a street-avenue path, meaning you can't cut across people's lawns. This metric, called the *taxicab metric*, is then given by the formula

$$d_T(x, y) = \sum_{i=1}^{n} |x_i - y_i|.$$

The proof that d_T is a metric is left to the reader.

The Max Metric. Our third metric on $\mathbb{R}^n$ is very simple to describe and calculate, but less easy to motivate naturally. Nevertheless, it is extremely useful, and is called the *max metric*, given by the formula

$$d_M(x, y) = \max\{|x_i - y_i|\}_{i=1}^n.$$

As above, positivity and symmetry are trivial to prove. As for the triangle inequality, it is key to note that that the maximum value of an indexed list of sums $a_i + b_i$ of nonnegative numbers is no larger than the maximum of the summands a_i, added to the the maximum of the summands b_i.

The Integral Metric. This metric measures how far two functions are apart, and is analogous to the taxicab metric. The underlying set is $\mathcal{C}[a, b]$, the set of of all continuous real-valued functions on the closed interval $[a, b] \subseteq \mathbb{R}$, and the metric is defined by the formula

$$\rho_I(f, g) = \int_a^b |f(x) - g(x)|\, dx.$$

If f and g are continuous functions from $[a, b]$ to $\mathbb{R}$, then so is $|f - g|$; hence—by the Fundamental Theorem of Calculus—the Riemann integral is defined. This is one case where the positivity axiom is the most difficult to prove; in particular, if $\rho(f, g) = 0$, how do we know that $f = g$? The trick is that if you know $f \neq g$, then $f(c) \neq g(c)$ for some $a \leq c \leq b$, hence $\epsilon = |f(c) - g(c)| > 0$. Finally, because f and g are continuous, there is some $\delta > 0$ such that if $a \leq x \leq b$ and $|x - c| \leq \delta$, then $|f(x) - g(x)| \geq \epsilon/2$. This tells us $\rho_I(f, g) > 0$.

The Uniform Metric. This is another metric on $\mathcal{C}[a, b]$, but is analogous to the max metric for $\mathbb{R}^n$. It is extremely useful on more advanced topology courses (metrization theory, in particular), and is defined by the formula

$$\rho_U(f, g) = \max\{|f(x) - g(x)| : a \leq x \leq b\}.$$

The fact that $\rho_U(f, g)$ is defined at all owes its existence to a basic topological result, one of the underpinnings of the calculus, called the Extreme Value Theorem: A continuous real-valued function defined on a closed bounded interval achieves a maximum (and a minimum) value. We'll have more to say about this important theorem later.

Discrete Metrics. For an arbitrary set X, define $d(x, y)$ to be 0 if $x = y$ and to be 1 otherwise. This is the *standard discrete metric* on X, and shows—among lots of other things—that any set can be assigned a metric. While it may seem to be of little interest in itself, it is of importance in defining other useful metrics.

Exercises 7. (1) For $x = \langle 1, 2\rangle$, $y = \langle 4, 6\rangle$ in $\mathbb{R}^2$, compute $d(x, y)$ when d is the euclidean (resp., taxicab, max, discrete) metric.

(2) Show that $d_M(x, y) \leq d_E(x, y) \leq d_T(x, y) \leq nd_M(x, y)$ always holds in $\mathbb{R}^n$. (In particular, they're all equal when $n = 1$.)

(3) Show that the discrete metrics always satisfy the metric axioms.

(4) If $\langle X, d\rangle$ is a metric space, with $x \in X$ and $A \subseteq X$ nonempty, define $d(x, A)$ to be $\text{glb}\{d(x, a) : a \in X\}$. This is the distance from a point to a set. Show it is well defined; i.e., it always makes sense. What is $d(x, A)$ when $X = \mathbb{R}^2$, $x = \langle 1, 1\rangle$, A is the unit disk $\{\langle a, b\rangle \in \mathbb{R}^2 : a^2 + b^2 \leq 1\}$, and d is the taxicab metric?

(5) If $\langle X, d\rangle$ is a metric space, with $A \subseteq X$ nonempty, we say A is *bounded* if $\{d(x, y) : x, y \in A\}$ is a bounded set of real numbers; and, in that case, we define the *diameter* $D_d(A)$ of A to be the least upper bound of this set of real numbers. Find $D_d(A)$, where A is the unit disk in $\mathbb{R}$ and d is the max metric.

(6) Let X be a set, and suppose $d : X \times X \to \mathbb{R}$ is a function satisfying, for all $x, y, z \in X$: (i) $d(x, y) = 0$ iff $x = y$; and (ii) $d(x, y) \leq d(x, z) + d(y, z)$. Prove that d is a metric on X.

Lecture 8: Open Sets and Closed Sets

Abstract. In a metric space, the set of points less than a given positive distance from a given point is called an open ball, centered at that point. We go from this metric notion to the topological notions of open set and of closed set. A set is closed just in case its complement is open; the signal feature of the family of all open subsets of a metric space is that it is closed under the taking of arbitrary unions and of finite intersections.

From Open Intervals to Open Sets. The use of the word *open* in this course broadens that which you will be familiar with from calculus. Recall that $I \subseteq \mathbb{R}$ is an *interval* if whenever $a, b \in I$ and $a \leq c \leq b$, then $c \in I$ as well. (Intervals are "closed under betweenness.") Because of the completeness property of the real line, intervals are defined by their end points. $(-\infty, \infty)$ is $\mathbb{R}$ itself, and is the only nonempty interval with no end points at all. An interval with one end point $a \in \mathbb{R}$ can either be *degenerate*, i.e., of the form $[a] = \{a\}$, or it can be a *ray*. Rays are either left-looking—of the form $(-\infty, a) = \{x \in \mathbb{R} : x < a\}$ or $(-\infty, a] = \{x \in \mathbb{R} : x \leq a\}$—or right-looking—of the form $(a, \infty) = \{x \in \mathbb{R} : x > a\}$ or $[a, \infty) = \{x \in \mathbb{R} : x \geq a\}$. The intervals with two end points $a < b$ in $\mathbb{R}$ are the *bounded nondegenerate* ones and come in four flavors, depending upon which end points are to be included: $(a, b) = \{x \in \mathbb{R} : a < x \text{ and } x < b\}$, $[a, b] = \{x \in \mathbb{R} : a \leq x \text{ and } x \leq b\}$, $[a, b) = \{x \in \mathbb{R} : a \leq x \text{ and } x < b\}$, and $(a, b] = \{x \in \mathbb{R} : a < x \text{ and } x \leq b\}$. (This should all be review.)

An interval is *open* if it contains none of its end points and is *closed* if it contains all of them. Thus the whole interval $(-\infty, \infty)$ is *both* open and closed since—vacuously—there are no end points either to contain or not to contain. (Likewise any interval of the form (a, a), since it is empty.) Each ray is either open or closed, but not both. Bounded intervals of the form $[a, b)$ or $(a, b]$, $a < b$, are neither open nor closed, and are sometimes facetiously referred to as "half-open."

We broaden the order-theoretic notion of *open interval* to the metric notion of *open ball* as follows. Let $\langle X, d \rangle$ be a metric space, with $a \in X$ and ϵ a positive real number. The *open d-ball of radius ϵ, centered at a* is denoted $B_d(a, \epsilon)$, and is defined to be $\{x \in X : d(a, x) < \epsilon\}$, the set of points of X that are less-than-ϵ distance from a. (The corresponding *closed ball* is denoted $B_d[a, \epsilon] = \{x \in X : d(a, x) \leq \epsilon\}$, but plays a less important role than that of the open ball.) We usually abbreviate $B_d(a, \epsilon)$ with $B(a, \epsilon)$ when the metric is understood.

Examples 8.1.
1. With the euclidean metric on $\mathbb{R}$, $B(a, \epsilon)$ and $B[a, \epsilon]$ are just $(a - \epsilon, a + \epsilon)$ and $[a - \epsilon, a + \epsilon]$, respectively.
2. With the euclidean metric on $\mathbb{R}^n$, $B(a, \epsilon)$ is the standard open n-ball $\{x \in \mathbb{R}^n : \sum_{i=1}^n (x_i - a_i)^2 < \epsilon^2\}$. (Note: open 1-balls are open intervals.)
3. With the taxicab metric on $\mathbb{R}^2$, $B(a, \epsilon)$, for say $a = \langle 0, 0 \rangle$ and $\epsilon = 1$, is the "open lozenge," defined by the inequality $|x| + |y| < 1$. It's bounded above by the line $y = 1 - x$ in the first quadrant, above by the line $y = 1 + x$ in the second quadrant, below by the line $y = -1 - x$ in the third quadrant, and below by the line $y = -1 + x$ in the fourth quadrant.

We now come to our first notions that are properly topological. If $\langle X, d\rangle$ is a metric space, $a \in X$ and $A \subseteq X$, we say A is a *neighborhood of* a if there is some $\epsilon > 0$ such that $B(a, \epsilon) \subseteq A$. So a neighborhood of $a \in X$ not only contains a as an element, but also contains all points "sufficiently close" to a. A set may be a neighborhood of some of its points, but not others; a set that is a neighborhood of each of its points is called an *open set*. The collection $\mathcal{T}_d = \{U \subseteq X : U \text{ is an open set }\}$ is called the *topology on* X *induced by the metric* d.

Proposition 8.2. *The following statements are equivalent for a subset* U *of a metric space* $\langle X, d\rangle$.

(a) U *is an open set.*
(b) *For each* $x \in U$*, there is a positive radius* ϵ *(usually depending on* x*) such that* $B(x, \epsilon) \subseteq U$.
(c) *For each* $x \in U$, $d(x, X \setminus U) > 0$.

Proof. The equivalence of (a) and (b) is down to the definition of *open set*. As for the equivalence with (c), recall the definition of $d(a, A)$ from Exercise 7(4) and assume (b) holds. If $B(x, \epsilon) \subseteq U$ and $y \in X \setminus U$, then $d(y, x) \geq \epsilon$. This says that ϵ is a lower bound for the set $\{d(y, x) : y \in X \setminus U\}$, so $d(x, X \setminus U)$, being the greatest lower bound, is $\geq \epsilon > 0$.

Now assume (c) holds and let's infer (b) from this. If $d(x, X \setminus U) = \epsilon > 0$, we show $B(x, \epsilon) \subseteq U$. Indeed, if $y \in B(x, \epsilon)$, then—by definition—$d(x, y) < \epsilon$. Now ϵ is a lower bound for the set $\{d(x, y) : y \in X \setminus U\}$. Hence, for y to be in $X \setminus U$, it must be the case that $d(x, y) \geq \epsilon$. This tells us that $B(x, \epsilon) \cap (X \setminus U) = \emptyset$; i.e., that $B(x, \epsilon) \subseteq U$.

□

In Lecture 2 we discussed the finitary Boolean operations; here we extend those to include possibly infinite unions and intersections. If $\mathcal{F}$ is a family of subsets of X (i.e., $\mathcal{F} \subseteq \wp(X)$), the *union* of that family is denoted $\bigcup \mathcal{F}$, and is defined to be the set $\{x \in X : x \in F \text{ for at least one } F \in \mathcal{F}\}$. If $\mathcal{F}$ is an *indexed family* $\langle F_i : i \in I\rangle$ (I is called the *index set*), then $\bigcup \mathcal{F}$ is often denoted $\bigcup_{i \in I} F_i$. (Or, in case the index set is $\mathbb{N}$, $\bigcup_{i=0}^{\infty} F_i$.)

The situation with intersections is analogous. $\bigcap \mathcal{F}$ is then defined to be $\{x \in X : x \in F \text{ for every } F \in \mathcal{F}\}$.

Proposition 8.3. *The family* $\mathcal{T} = \mathcal{T}_d$ *of open sets of a metric space* $\langle X, d\rangle$ *satisfies the following conditions:*

(1) $\emptyset$ *and* X *are open sets (i.e., members of* $\mathcal{T}$*).*
(2) *If* $\mathcal{U} \subseteq \mathcal{T}$ *is a finite collection of open sets then its intersection* $\bigcap \mathcal{U}$ *is an open set.*
(3) *If* $\mathcal{U} \subseteq \mathcal{T}$ *is an arbitrary collection of open sets then its union* $\bigcup \mathcal{U}$ *is an open set.*

Proof. (Ad (1)). The empty set is open because it is a neighborhood of each of its points—of which there are none. X itself contains all neighborhoods, and is hence a neighborhood of each point.

(Ad (2)). Suppose U and V are open sets and $x \in U \cap V$. Then there are positive real numbers ϵ_U and ϵ_V such that $B(x, \epsilon_U) \subseteq U$ and $B(x, \epsilon_V) \subseteq V$. Let $\epsilon = \min\{\epsilon_U, \epsilon_V\}$. Then $B(x, \epsilon) \subseteq B(x, \epsilon_U) \cap B(x, \epsilon_V) \subseteq U \cap V$. The intersection of $n + 1$ open sets is the intersection of the first n of them with the $n + 1$st; this provides the basis for an induction on the finite number of open sets.

(Ad (3)). Suppose $\mathcal{U}$ is a family of open sets, with $x \in \bigcup \mathcal{U}$. Then, for some $U \in \mathcal{U}$, $x \in U$. Since U is open, there is a ball neighborhood $B(x, \epsilon) \subseteq U$. Since $U \subseteq \bigcup \mathcal{U}$, $B(x, \epsilon) \subseteq \bigcup \mathcal{U}$, showing the union to be open.

□

Closed Sets. A subset C of a metric space $\langle X, d \rangle$ is a *closed set* if its complement $X \setminus C$ is an open set. (Warning! Do not read this as meaning that a set is closed whenever it isn't open. We've already seen examples of sets that are both open and closed.) The following is a direct consequence of Proposition 8.3, deducible using both the finitary and infinitary versions of the DeMorgan Laws.

Proposition 8.4. *The family $\mathcal{T}' = \mathcal{T}'_d$ of closed sets of a metric space $\langle X, d \rangle$ satisfies the following conditions:*

1. *$\emptyset$ and X are closed sets (i.e., members of $\mathcal{T}'$).*
2. *If $\mathcal{F} \subseteq \mathcal{T}'$ is a finite collection of closed sets then its union $\bigcup \mathcal{F}$ is a closed set.*
3. *If $\mathcal{F} \subseteq \mathcal{T}'$ is an arbitrary collection of closed sets then its intersection $\bigcap \mathcal{F}$ is a closed set.*

Exercises 8.

(1) For metric space $\langle X, d \rangle$, $a \in X$, and $\epsilon > 0$ show that the open ball $B(a, \epsilon)$ is an open set and that the closed ball $B[a, \epsilon]$ is a closed set.

(2) Show that a finite subset of a metric space is closed.

(3) Prove that a nonempty subset C of a metric space $\langle X, d \rangle$ is closed if and only if $d(x, C) > 0$ for each $x \notin C$.

(4) If A and B are nonempty subsets of metric space $\langle X, d \rangle$ and $x \in X$, show that $d(x, A \cup B) = \min\{d(x, A), d(x, B)\}$.

(5) Prove that every subset of a discrete metric space is both open and closed.

(6) Suppose x and y are distinct points of metric space $\langle X, d \rangle$. Show that there are open sets U and V with $x \in U$, $y \in V$, and $U \cap V = \emptyset$.

(7) Show that every nonempty open set in $\mathbb{R}^n$—where the metric is either euclidean, taxicab, or max—contains points all of whose coordinates are rational.

(8) Prove that Propositions 8.3 and 8.4 imply each other.

(9) If d is the standard discrete metric on X and $x \in X$, then what's the difference between $B(x, 1)$ and $B(x, 1.0001)$?

(10) If d is a metric on X and $\epsilon > 0$, the binary relation $E_d(\epsilon) := \{\langle x, y \rangle \in X^2 : d(x, y) < \epsilon\}$ is called an *entourage*. Which of the properties of relations in Lecture 4 pertain to the entourages of a metric?

(11) An *ultrametric* is a metric that satisfies the very strong triangle inequality $d(x,z) \leq \max\{d(x,y), d(y,z)\}$. Show that the entourages of an ultrametric are all equivalence relations.

Lecture 9: Accumulation Points and Convergent Sequences

Abstract. A point of a metric space is an accumulation point of a subset of the space if every open ball centered at the point intersects the subset in at least one other point. Here we relate this notion with those of open sets, closed sets, and convergent sequences. For example, a subset is closed if and only if it contains all its accumulation points; a point is an accumulation point of a subset just in case there is a sequence of distinct points of the subset that converges to the given point.

Accumulation Points. Another topological notion, just as fundamental as that of open (or closed) set, is the one where a point x of a metric space $\langle X, d\rangle$ is an *accumulation point* of a set $A \subseteq X$. This happens precisely when every neighborhood of x intersects A in a point other than x; i.e., when $B(x,\epsilon) \cap (A \setminus \{x\}) \neq \emptyset$ for all $\epsilon > 0$. The *derived set* $\text{der}(A)$ of $A \subseteq X$ is the set of all points $x \in X$ that are accumulation points of A.

Examples 9.1.
(1) Let our metric space be $\mathbb{R}$, with its euclidean metric, and let $A = \{1/n : n = 1, 2, \dots\}$. Then 0 is an accumulation point of A; in fact 0 is the only accumulation point of A. I.e., $\text{der}(A) = \{0\}$.
(2) If A is a finite subset of $\langle X, d\rangle$, then A has no accumulation points; i.e., $\text{der}(A) = \emptyset$.
(3) If our metric space is euclidean $\mathbb{R}^n$ and A consists of all n-tuples with rational coordinates, then $\text{der}(A) = \mathbb{R}^n$.

If x is an accumulation point of A, the intuition is that A "clusters around" x. It means that x is arbitrarily close not only to A, but to the smaller set $A \setminus \{x\}$. The proof of the following is left as an exercise.

Proposition 9.2. *Let $\langle X, d\rangle$ be a metric space, with $x \in X$ and $A \subseteq X$. Then x is an accumulation point of A if and only if $d(x, A \setminus \{x\}) = 0$.*

Proposition 9.3. *Let $\langle X, d\rangle$ be a metric space, with $A \subseteq X$. Then A is a closed set if and only if $\text{der}(A) \subseteq A$. (I.e., a set is closed just in case it contains all its accumulation points.)*

Proof. Suppose A is closed. Then $X \setminus A$ is open, by definition. If $x \notin A$, then $x \in X \setminus A$; hence there is a positive ϵ such that $B(x,\epsilon) \subseteq X \setminus A$. This means that $B(x,\epsilon) \cap A = \emptyset$, hence x is not an accumulation point of A. We've shown that $x \notin A$ implies $x \notin \text{der}(A)$, so therefore $\text{der}(A) \subseteq A$.

Suppose A is not closed. Then $X \setminus A$ is not open, which means that for some $x \notin A$, every $B(x,\epsilon)$, $\epsilon > 0$, intersects $A = A \setminus \{x\}$. Hence x is an accumulation point of A, which is not in A. Therefore $\text{der}(A) \not\subseteq A$.

□

Convergent Sequences. While the concept of accumulation point may be new to most of you, that of convergent sequence should ring some bells from your calculus courses. Recall that a *sequence* in a set X is a function $a : D \to X$, where $D \subseteq \mathbb{N}$ is an infinite set. The idea is that D has a beginning, but no end; and for the present discussion we take $D = \mathbb{N}$ and write $\langle a_0, a_1, \dots \rangle$, or $\langle a_n \rangle_{n=0}^{\infty}$, where $a_n = a(n)$, to indicate how sequences may be thought of as "infinity-tuples." Since a sequence is a function, it has an image in X, and we often refer to it as the *trace* $\{a_0, a_1, \dots\}$ of the sequence. Note that, while order counts in the notation $\langle a_0, a_1, \dots \rangle$, and that repetitions are possible, neither of these features holds for the notation $\{a_0, a_1, \dots\}$. Hence the trace of the sequence $\langle 1, -1, 1, -1, \dots \rangle = \langle (-1)^n \rangle_{n=0}^{\infty}$ is just $\{-1, 1\} = \{1, -1\}$. Sequences are infinite objects because they're functions with infinite domains; traces of sequences are subsets of the range set X and may well be finite.

If $\langle a_n \rangle$ is a sequence in a metric space $\langle X, d \rangle$, then we say the sequence *converges* to $x \in X$, and that x is a *limit* of the sequence, if for each $\epsilon > 0$ there is some $N \in \mathbb{N}$ such that $d(a_n, x) < \epsilon$ for every $n \geq N$. In colloquial terms, every neighborhood of x contains a *tail* of the sequence.

Proposition 9.4. *A sequence in a metric space cannot converge to more than one point of the space.*

Proof. Suppose $\langle a_n \rangle$ converges to both x and y in X, where $x \neq y$. Then $d(x, y) > 0$, so let $\epsilon = \frac{1}{2} d(x, y)$. Since the sequence converges to x, there is some $N_x \in \mathbb{N}$ such that $d(a_n, x) < \epsilon$ for all $n \geq N_x$. since the sequence converges to y, there is some $N_y \in \mathbb{N}$ such that $d(a_m, y) < \epsilon$ for all $m \geq N)y$. So let $N = \max\{N_x, N_y\}$. Then $N \geq N_x$, so $d(a_N, x) < \epsilon$. But $N \geq N_y$ also, so $d(a_N, y) < \epsilon$. By the triangle inequalilty, $d(x, y) \leq d(a_N, x) + d(a_N, y) < \epsilon + \epsilon = d(x, y)$. No number is strictly less than itself, so we have a contradiction: It can never happen that a sequence converges to two or more points.

□

Because limits of sequences are unique when they exist, we write $x = \lim_{n\to\infty} \langle a_n \rangle$, or $\langle a_n \rangle \to x$, to indicate that x is *the* limit of $\langle a_n \rangle_{n=0}^{\infty}$. If A is a subset of X, a *sequence in* A is just a sequence in X whose trace is a subset of A. A *sequence of distinct points of* A is a sequence in A that is also one-to-one as a function from $\mathbb{N}$ to X.

Proposition 9.5. *Let* $\langle X, d \rangle$ *be a metric space, with* $A \subseteq X$.

1. *A point* $x \in X$ *is an accumulation point of* A *if and only if there is a sequence of distinct points of* A *that converges to* x.
2. A *is a closed set in* X *if and only if* A *contains the limit of every convergent sequence in* A.

Proof. (Ad (1)). Suppose $\langle a_n \rangle$ is a sequence of distinct points of A that converges to $x \in X$. To show x is an accumulation point of A, fix $\epsilon > 0$. Then there is some $N \in \mathbb{N}$ such that $a_n \in B(x, \epsilon)$ for all $n \geq N$. Since the sequence consists of distinct points, there must be some $n \geq N$ such that $a_n \neq x$. Hence $B(x, \epsilon) \cap (A \setminus \{x\}) \neq \emptyset$.

Suppose $x \in X$ is an accumulation point of A. We need to construct a sequence of distinct points of A, which converges to x. We do this by an inductive process;

to get started we pick $a_0 \in B(x, 1) \cap (A \setminus \{x\})$, guaranteed to exist since x is an accumulation point of A. For the second step, we let $\epsilon_1 = \min\{2^{-1}, d(x, a_0)\}$ and pick $a_1 \in B(x, \epsilon_1) \cap (A \setminus \{x\})$. Then a_1, being strictly closer to x than a_0, must be distinct from a_0. In the inductive step, we assume we have found $a_0, a_1, \ldots, a_n \in A$, where $0 < d(x, a_m) < \min\{2^{-m}, d(x, a_{m-1})\}$, for all $1 \le m \le n$. We then let $\epsilon_{n+1} = \min\{2^{-(n+1)}, d(x, a_n)\}$ and pick $a_{n+1} \in B(x, \epsilon_{n+1}) \cap (A \setminus \{x\})$. This gives us the sequence we want.

(Ad (2)). This is proved much as is Proposition 9.3, and is left as an exercise.

□

Exercises 9. (1) Prove Proposition 9.2.
(2) Prove Proposition 9.5(2).
(3) A sequence $\langle a_n \rangle$ is *eventually constant* if there is some $N \in \mathbb{N}$ such that the terms a_n, $n \ge N$, are all equal to one another. Show that every eventually constant sequence in $\langle X, d \rangle$ is convergent; and, if d is the discrete metric, show every convergent sequence is eventually constant.
(4) If x is the limit of a sequence $\langle a_n \rangle$ of distinct points in $\langle X, d \rangle$, show that x is also an accumulation point of the trace of the sequence. Does the same necessarily hold for all sequences?
(5) Prove that if a is an accumulation point of A, then every neighborhood of a intersects A in an infinite set.

Lecture 10: Interior, Closure, and Frontier

Abstract. In a metric (or topological) space, the interior of a subset is the union of all open sets contained in the subset; the closure of that subset is the intersection of all closed sets containing the subset. Points of the interior of a subset are "well within" it; points of the closure of a subset—but not in the set itself—are accumulation points that are "hanging onto" the subset. These points form the frontier of the subset, namely the closure of the subset intersected with the closure of the complement of the subset.

The Interior Operation. Let $\langle X, d\rangle$ be a metric space, $x \in X$, and $A \subseteq X$. Define x to be an *interior point* of A if $U \subseteq A$ for some open set U containing x; i.e., if A is a neighborhood of x. This simple paraphrase gives us the notion of the *interior* of a set $A \subseteq X$ as the set of all interior points of A. The interior of A, denoted $\text{int}(A)$, is a new set in X, obviously a subset of A. The assignment $A \mapsto \text{int}(A)$ defines a function $\text{int} : \wp(X) \to \wp(X)$, called the *interior operation* on $\langle X, d\rangle$.

Examples 10.1. Suppose $\langle X, d\rangle$ is the real line with the euclidean metric.

(1) For $a, b \in \mathbb{R}$, $a < b$, we have $\text{int}([a, b]) = \text{int}([a, b)) = \text{int}((a, b]) = \text{int}((a, b)) = (a, b)$.
(2) Every finite set has empty interior.
(3) Both the sets of rational points and of irrational points have empty interior.

Proposition 10.2. *Let $\langle X, d\rangle$ be a metric space, $A, B \subseteq X$.*

(1) *$\text{int}(A)$ is an open set, and is the union of all open sets contained in A. A is an open set if and only if $A = \text{int}(A)$.*
(2) *$\text{int}(\emptyset) = \emptyset$; $\text{int}(X) = X$.*
(3) *If $A \subseteq B$, then $\text{int}(A) \subseteq \text{int}(B)$.*
(4) *$\text{int}(A \cap B) = \text{int}(A) \cap \text{int}(B)$.*
(5) *$\text{int}(A \cup B) \supseteq \text{int}(A) \cup \text{int}(B)$; the reverse inclusion need not hold.*

Proof. (Ad (1, 2)). This is obvious, and left as an exercise.

(Ad (3)). Suppose $A \subseteq B$. If $x \in \text{int}(A)$, then A is a neighborhood of x. Immediate from the definition is the fact that any superset of a neighborhood of a point is also a neighborhood of that point; hence B is a neighborhood of x. Thus $x \in \text{int}(B)$.

(Ad (4)). Since both $A \cap B \subseteq A$ and $A \cap B \subseteq B$ hold we have—from (3)—that $\text{int}(A \cap B) \subseteq \text{int}(A)$ and $\text{int}(A \cap B) \subseteq \text{int}(B)$ both hold. Hence $\text{int}(A \cap B) \subseteq \text{int}(B) \cap \text{int}(B)$. For the reverse inclusion, suppose $x \in \text{int}(A) \cap \text{int}(B)$. Then there are open sets U and V with $x \in U \subseteq A$ and $x \in V \subseteq B$. Since $U \cap V$ is also open, we have $x \in U \cap V \subseteq A \cap B$; hence $x \in \text{int}(A \cap B)$.

(Ad (5)). Since $A \subseteq A \cup B$ and $B \subseteq A \cup B$, we use (2), plus basic facts about set union, to conclude that $\text{int}(A) \cup \text{int}(B) \subseteq \text{int}(A \cup B)$. As for the failure of equality, all we need is a single example where the left-hand side is a proper subset of the right. We need look no further than $\mathbb{R}$, with the euclidean metric: Set $A = [0,1]$ and $B = [1,2]$. Then $\text{int}(A) \cup \text{int}(B) = (0,1) \cup (1,2)$, while $\text{int}(A \cup B) = \text{int}([0,2]) = (0,2)$. These sets are not equal; the latter contains 1, which is not in the former.

□

The Closure Operation. Let $\langle X, d\rangle$ be a metric space. We define the *closure operation* $\text{cl} : \wp(X) \to \wp(X)$ by the rule $\text{cl}(A) = A \cup \text{der}(A)$. That is, to form the closure of A, we add to A all those points of X that are "stuck" to A.

Examples 10.3. Suppose $\langle X, d\rangle$ is the real line with the euclidean metric.

(1) For $a, b \in \mathbb{R}$, $a < b$, we have $\text{cl}((a,b)) = \text{cl}([a,b)) = \text{cl}((a,b]) = \text{cl}([a,b]) = [a,b]$.
(2) Every finite set equals its own closure, since its derived set is empty.
(3) The closures of both the sets of rational points and of irrational points are equal to $\mathbb{R}$.

The following does for closure what Proposition 10.2 did for interior.

Proposition 10.4. *Let $\langle X, d\rangle$ be a metric space, $A, B \subseteq X$.*

(1) *$cl(A)$ is a closed set, and is the intersection of all closed sets containing A. A is a closed set if and only if $A = cl(A)$.*
(2) *$cl(\emptyset) = \emptyset$; $cl(X) = X$.*
(3) *If $A \subseteq B$, then $cl(A) \subseteq cl(B)$.*
(4) *$cl(A \cup B) = cl(A) \cup cl(B)$.*
(5) *$cl(A \cap B) \subseteq cl(A) \cap cl(B)$; the reverse inclusion need not hold.*

Proof. (Ad (1)). If $x \in X \setminus \text{cl}(A)$, then—since $x \notin \text{der}(A)$—there is an open set U containing x such that $U \cap (A \setminus \{x\}) = \emptyset$. Since $x \notin A$, though, we know $U \cap A = \emptyset$ too. It's also true that $U \cap \text{der}(A) = \emptyset$; otherwise U would have to intersect A. Thus $U \subseteq X \setminus \text{cl}(A)$, proving $\text{cl}(A)$ to be closed. If C is any closed set containing A, then C must contain $\text{der}(A)$ as well. (Why?) Since the intersection of any family of closed sets is closed, and $\text{cl}(A)$ is a closed set containing A, $\text{cl}(A)$ must therefore be the intersection of all closed sets containing A. From this it is clear that A is a closed set if and only if $A = \text{cl}(A)$.

The proofs of the other items are left as exercises.

□

The Frontier Operation. Let $\langle X, d\rangle$ be a metric space. We define the *frontier operation* $\text{fr} : \wp(X) \to \wp(X)$ by the rule $\text{fr}(A) = \text{cl}(A) \cap \text{cl}(X \setminus A)$. That is, a point x is on the frontier of set A just in case is is in the closures of both A and the complement of A. So clearly a set and its complement have the same frontier; the following result is left to the reader to prove as an exercise.

Proposition 10.5. *Let $\langle X, d\rangle$ be a metric space, $x \in X$, and $A \subseteq X$. The following statements are equivalent.*

(a) *$x \in fr(A)$.*

(b) $x \in cl(A) \setminus int(A)$.
(c) *Every open set containing* x *intersects both* A *and* $X \setminus A$.
(d) *Every neighborhood of* x *intersects both* A *and* $X \setminus A$.
(e) $d(x, A) = d(x, X \setminus A) = 0$.

Remark 10.6. Many modern authors use the word *boundary* instead of *frontier* in talking about the points that adhere both to a set and to its complement. We chose in these lectures to use *frontier* for three reasons: first, there is the fact of precedence, that there is a sizeable number of authors who use the word; second, the word is not used in any other topological context; and third, the word *boundary* does have another meaning in topology. For example, if $\mathbb{I} = [0, 1]$ is the closed unit interval in $\mathbb{R}$, then its frontier is $\{0, 1\}$. However, we can view $\mathbb{I}$ as the closed unit interval in the x-axis of $\mathbb{R}^2$. In that case it is its own frontier: it is a closed set in $\mathbb{R}^2$, and the closure of its complement in $\mathbb{R}^2$ is all of $\mathbb{R}^2$. On the other hand, the boundary of $\mathbb{I}$, as a one-dimensional manifold, is $\{0, 1\}$, regardless of how the interval is viewed as a subset of an ambient metric space.

Exercises 10. (1) Prove Proposition 10.2(1,2).
(2) Let $\mathbb{R}$ have the euclidean metric. Prove that $\text{int}(\mathbb{Q}) = \emptyset$ and $\text{cl}(\mathbb{Q}) = \mathbb{R}$. What is $\text{fr}(\mathbb{Q})$?
(3) Prove Proposition 10.4(2,3,4,5).
(4) Prove Proposition 10.5.
(5) Show that in any metric space, $\text{int}(\text{int}(A)) = \text{int}(A)$ and $\text{cl}(\text{cl}(A)) = \text{cl}(A)$.
(6) Show that in a discrete metric space, every subset has empty frontier.
(7) Show that in any metric space, $\text{cl}(B(x, \epsilon)) \subseteq B[x, \epsilon]$. Give an example where equality does not hold.

Lecture 11: Continuous Functions and Homeomorphisms

Abstract. The definition of continuity if functions in calculus is broadened to apply to functions from one metric (or topological) space to another. A one-to-one function from one space onto another is a homeomorphism if both the function and its inverse are continuous. Two spaces are homeomporphic—i.e., topologically indistinguishable—if there is at least one homeomorphism between them.

Continuity. Recall from calculus that a function $f : \mathbb{R} \to \mathbb{R}$ is continuous at a point $a \in \mathbb{R}$ if, whenever $\epsilon > 0$, there is a $\delta > 0$ such that if $|x - a| < \delta$, then $|f(x) - f(a)| < \epsilon$. This is admittedly a confusing criterion, but is unfortunately unavoidable. The intuition is simple: "If x is close to a, then $f(x)$ must be close to $f(a)$." This says, roughly, that continuous functions don't tear things apart. The famous—or infamous—epsilon-delta criterion for continuity was formulated in the 1820s by Augustin-Louis Cauchy in order to make the definition more precise, to meet the needs of tightening standards of rigor in mathematical discourse. The problem, as you can probably guess, lies with the word *close*: there simply is no objective characterization of what this means.

The big breakthrough came with the realization that one didn't need an *absolute* notion of closeness, only a *relative* one. So, by making x close enough to a in response to a given standard of closeness to $f(a)$, one could guarantee that $f(x)$ would meet that standard.

The idea can be likened to a problem in archery. Suppose you're trying to aim an arrow to hit a circular target that's four feet across and one hundred feet down range. Assuming the existence of an ideal aim that will send the arrow to the exact center of the target, how much could you be off that aim and still not miss the target entirely?

Where the epsilon ϵ comes in is as the standard of closeness you have to meet, getting the arrow within $\epsilon = 2$ feet of dead center; where the delta δ comes in is as a measure of how far off perfect aim you can be and still meet the ϵ-standard. So, given that you're within δ degrees of the angle of perfect aim, your result will be within ϵ feet of dead center.

Now let's try this idea in the context of metric spaces, recalling the fact that the quantities $|x - a|$ and $|f(x) - f(a)|$ are just the euclidean distances between x and a, and $f(x)$ and $f(a)$, respectively.

Definition 11.1. Let $\langle X, d\rangle$ and $\langle Y, e\rangle$ be metric spaces, $f : X \to Y$ a function, and $a \in X$. then f is *continuous at* a if whenever $\epsilon > 0$, there is a $\delta > 0$ such that $e(f(x), f(a)) < \epsilon$ whenever $d(x, a) < \delta$.

As in the real context, the value of δ may depend on ϵ, as well as on the particular point a at which you're trying to show continuity. The δ response is like a winning move in a two move game when the second player has a sure winning strategy: no matter how cleverly the first player chooses a (very small, maybe) $\epsilon > 0$, the second player can always respond with an appropriate $\delta > 0$ to make the required continuity condition hold. On the other hand, if f is not continuous at a, then the epsilon player can play an ϵ so small, that no matter how small the delta player chooses δ, it is still possible to find some $x \in X$ such that $e(f(x), f(a)) \geq \epsilon$ despite the fact that $d(x, a) < \delta$. In this case, the epsilon player has a winning strategy.

Examples 11.2. (1) Let $f : \mathbb{R} \to \mathbb{R}$ be $x \mapsto x^2$, and suppose $a = 2$. To show f is continuous at a, you need to be assured that $|x^2 - 4| < \epsilon$ as long as x is close enough to 2. Doing a little algebra, we have $|x^2 - 4| = |x + 2||x - 2|$. If we make sure that $\delta \leq 1$, we can maximize what $|x + 2|$ can be; i.e., $1 < x < 3$, so $|x + 2| < 5$. This tells us we should pick $0 < \delta \leq \min\{1, \epsilon/5\}$. Thus, if x is such that $|x - 2| < \delta$, then we have $|x^2 - 4| < 5(\epsilon/5) = \epsilon$, as desired. Here, the choice of δ is how the delta player applies her winning strategy.

(2) Let $g : \mathbb{R} \to \mathbb{R}$ be constantly 0 for $x < 0$ and to be constantly 1 for $x \geq 0$. We expect the function to be discontinuous at $a = 0$, so let's find a winning strategy for the epsilon player. Indeed, it looks like the function takes a leap of 1 unit as x goes from negative to positive, so it looks like setting ϵ to be any positive number ≤ 1 should work to foil the delta player. Indeed, given any $\delta > 0$, there are negative numbers within δ of $a = 0$. Hence, if x is any one of these, then $g(x) = 0$, while $g(0) = 1$. Thus $|g(x) - g(0)| = 1 \geq \epsilon$.

Proposition 11.3. *Let $\langle X, d\rangle$ and $\langle Y, e\rangle$ be metric spaces, with $f : X \to Y$ and $a \in X$. The following are equivalent.*

(a) *f is continuous at a.*
(b) *For every open $V \subseteq Y$ with $f(a) \in V$, there is an open $U \subseteq X$ with $x \in U$ and $f[U] \subseteq V$.*
(c) *If $\langle a_n\rangle$ is a sequence in X converging to a, then $\langle f(a_n)\rangle$ is a sequence in Y converging to $f(a)$.*
(d) *If $A \subseteq X$ and $a \in cl(A)$ then $f(a) \in cl(f[A])$.*

Proof. (Ad ($a \implies b$)). Assume f is continuous at a, and that $V \subseteq Y$ is open and contains $f(a)$. Then there is some $\epsilon > 0$ such that $B_e(f(a), \epsilon) \subseteq V$. Since f is continuous at a, there is some $\delta > 0$ such that $d(x, a) < \delta$ implies $e(f(x), f(a)) < \epsilon$. In other words, $f[B_d(a, \delta)] \subseteq B_e(f(a), \epsilon)$. So choose $U = B_d(a, \delta)$, well known to be an open set. Then $f[U] \subseteq B_e(f(a), \epsilon) \subseteq V$, as desired.

(Ad ($b \implies c$)). Suppose $\langle a_n\rangle \to a$ in $\langle X, d\rangle$. If V is open in Y and contains $f(a)$, we find U open in X such that $a \in U$ and $f[U] \subseteq V$. Since $\langle a_n\rangle \to a$, there is some $N \in \mathbb{N}$ such that $a_n \in U$ for all $n \geq N$. Thus $f(a_n) \in V$ for all $n \geq N$; this shows $\langle f(a_n)\rangle \to f(a)$.

(Ad ($c \implies d$)). Suppose $A \subseteq X$ and $a \in \mathrm{cl}(A)$. If $a \in A$, then $f(a) \in f[A] \subseteq \mathrm{cl}(f[A])$, and we're done. Otherwise $a \in \mathrm{der}(A)$, and we invoke Proposition 9.5(1) to infer that there is a sequence $\langle a_n\rangle$ in A that converges to a. Thus, by (c), $\langle f(a_n)\rangle \to f(a)$. This tells us that $f(a) \in \mathrm{cl}(f[A])$.

(Ad ($d \implies a$)). (Exercise.)

$\square$

$f : X \to Y$ is called *continuous*, or a *map* (or *mapping*), if f is continuous at each point $a \in X$.

Proposition 11.4. *Let $\langle X, d\rangle$ and $\langle Y, e\rangle$ be metric spaces, with $f : X \to Y$. The following are equivalent.*

(a) *f is continuous.*
(b) *For every open $V \subseteq Y$, $f^{-1}[V]$ is open in X.*
(c) *For every closed $C \subseteq Y$, $f^{-1}[C]$ is closed in X.*
(d) *For every $A \subseteq X$, $f[cl(A)] \subseteq cl(f[A])$.*

Proof. (Ad ($a \implies b$)). Let $V \subseteq Y$ be open, with $a \in f^{-1}[V]$. Since $f(a) \in V$ and V is open, we may find $\epsilon > 0$ such that $B_e(f(a), \epsilon) \subseteq V$. Since f is continuous at a, there is some $\delta > 0$ such that $f[B_d(a, \delta)] \subseteq B_e(f(a), \epsilon) \subseteq V$. This tells us that $B_d(a, \delta) \subseteq f^{-1}[V]$, and hence that $f^{-1}[V]$ is open in X.

(Ad ($b \implies c, c \implies d$). (Exercise.)

(Ad ($d \implies a$). Pick an arbitrary point $a \in X$; we aim to show f is continuous at a. From Proposition 11.3 ($d \implies a$), it suffices to show that if $A \subseteq X$ is such that $a \in \mathrm{cl}(A)$, then $f(a) \in \mathrm{cl}(f[A])$. But if A is such a set, then, assuming (d) holds, $f[\mathrm{cl}(A)] \subseteq \mathrm{cl}(f[A])$. This is enough to get what we need.

□

Homeomorphisms. A continuous bijection whose inverse is also continuous is called a *homeomorphism*. The existence of a homeomorphism—if there's one, there are many—between two metric spaces tells us that the two spaces are *homeomorphic*, or *topologically indistinguishable*: either one may be continuously transformed into the other. This is what we hinted at in Lecture 1, using the metaphor "rubber sheet geometry."

Examples 11.5. (1) If $\langle X, d\rangle$ and $\langle Y, e\rangle$ are metric spaces and $f : X \to Y$ is a surjection such that $e(f(x), f(y)) = d(x, y)$ for all $x, y \in X$, we call f an *isometry* and say that the two metric spaces are *isometric*. It is easy to show that isometries are homeomorphisms: if $x \neq y$, then $d(x, y) > 0$, hence $e(f(x), f(y)) > 0$, so $f(x) \neq f(y)$; continuity works by setting $\delta = \epsilon$ every time. When we're studying the topology of metric spaces, it's homeomorphisms we care about most; we care about isometries when we're studying the *geometry* if metric spaces.

(2) The exponential function $y = e^x$ provides a homeomorphism between the euclidean real line $\mathbb{R}$ and the euclidean ray $(0, \infty)$. Its inverse is the natural logarithm; it is not an isometry since $|e^1 - e^0| = e - 1 \neq 1 = |1 - 0|$.

Exercises 11. (1) Prove the remaining implication in Proposition 11.3.

(2) Prove the remaining two implications in Proposition 11.4.

(3) A point a of metric space $\langle X, d\rangle$ is *isolated* if $a \notin \mathrm{der}(X)$. Show that any function is continuous at each isolated point of its domain.

(4) Let $f : X \to Y$ be a function between metric spaces. Show f is continuous if and only if $f^{-1}[\text{int}(B)] \subseteq \text{int}(f^{-1}[B])$ for every $B \subseteq Y$.
(5) Suppose $f : X \to Y$ and $g : Y \to Z$ be continuous functions between metric spaces. Show that the composite function $g \circ f : X \to Z$ is continuous.
(6) Show that if $f : X \to Y$ is a function and X has the discrete metric, then f is continuous.

Lecture 12: Topologically Equivalent Metrics

Abstract. We examine the situation when two metrics on the same set give rise to the same family of open sets; i.e., when they are topologically equivalent. A very useful result is that every metric is topologically equivalent to a metric that is bounded.

When Two Metrics Give the Same Topology. Recall from Lecture 11 that two metric spaces $\langle X, d\rangle$ and $\langle Y, e\rangle$ are *homeomorphic* if there is a homeomorphism $f : X \to Y$; i.e., a continuous bijection whose inverse is also continuous. In the special case where $X = Y$ and f is the identity function i_X, the two metrics d and e are said to be *topologically equivalent*; in other words, they yield the same collections of open sets.

Proposition 12.1. *Let d and e are two metrics with underlying set X. The following are equivalent.*

(a) *d and e are topologically equivalent.*
(b) *For every $a \in X$ and $\epsilon > 0$ there are $\delta, \eta > 0$ such that both $B_d(a, \delta) \subseteq B_e(a, \epsilon)$ and $B_e(a, \eta) \subseteq B_d(a, \epsilon)$ hold.*

Proof. (Ad ($a \implies b$). Since d and e are topologically equivalent, their associated topologies $\mathcal{T}_d$ and $\mathcal{T}_e$ are equal. So pick $a \in X$ and $\epsilon > 0$. Then $B_e(a, \epsilon) \in \mathcal{T}_d$, so there is some $\delta > 0$ such that $B_d(a, \delta) \subseteq B_e(a, \epsilon)$. Likewise $B_d(a, \epsilon) \in \mathcal{T}_e$, so there is some $\eta > 0$ such that $B_e(a, \eta) \subseteq B_d(a, \epsilon)$.

(Ad ($b \implies a$). Suppose $U \in \mathcal{T}_d$. To show $U \in \mathcal{T}_e$, let $a \in U$. Then there is some $\epsilon > 0$ such that $B_d(a, \epsilon) \subseteq U$. Now, by (b), there is some $\eta > 0$ such that $B_e(a, \eta) \subseteq B(a, \epsilon) \subseteq U$. This tells us that every point of U is interior to U with respect to the metric e; hence that $U \in \mathcal{T}_e$. Likewise we show that every e-open set is d-open, so the two metrics are topologically equivalent.

□

Proposition 12.2. *Let d and e are two metrics with underlying set X, and suppose there are positive real numbers α and β such that, for all $x, y \in X$, $d(x, y) \le \alpha e(x, y)$ and $e(x, y) \le \beta d(x, y)$. Then d and e are topologically equivalent.*

Proof. In view of Proposition 12.1, all we need to show is that condition 12.1(b) holds. Indeed, given $\epsilon > 0$, let $\delta = \epsilon/\beta$ and $\eta = \epsilon/\alpha$. If $a \in X$ is fixed and $x \in B_d(a, \delta)$, then $d(a, x) < \delta$, so $e(a, x) \le \beta d(x, a) < \beta\delta = \epsilon$. Hence $x \in B_e(a, \epsilon)$. Thus $B_d(a, \delta) \subseteq B_e(a, \epsilon)$. Similarly, $B_e(a, \eta) \subseteq B_d(a, \epsilon)$.

□

Example 12.3. From Exercise 7(2), in $\mathbb{R}^n$ we have the following inequality holding among the max metric, the euclidean metric, and the taxicab metric:

$$d_M(x, y) \le d_E(x, y) \le d_T(x, y) \le n d_M(x, y).$$

From this, plus Proposition 12.2, it follows that all three of these metrics give rise to the same open sets in $\mathbb{R}^n$, even though the various ϵ-balls have completely different geometric shape.

Bounded Metrics. The metric d with underlying set X is *bounded* if there is some positive real number M such that $d(x, y) \leq M$ for all $x, y \in X$. Clearly the max, euclidean, and taxicab metrics on $\mathbb{R}^n$ are not bounded, as $d(x, y)$ can get as large as we like when d is one of these. On the other hand, the standard discrete metric on a set is bounded because $d(x, y) \leq 1$, no matter what. The following result may seem a bit surprising at first; it is of tremendous use in metrization theory—whatever *that* is—and basically shows that boundedness is not a topological property of a metric space.

Proposition 12.4. *Let $\langle X, d\rangle$ be a metric space. Then there is a bounded metric on X that is topologically equivalently to d.*

Proof. The bounded metric we show equivalent to d is defined by $e(x, y) = \min\{d(x, y), 1\}$. It is an easy to show that e is a legitimate metric; notice that if d is unbounded to begin with, then d and e do not satisfy the hypothesis of Proposition 12.2. So once we show d and e to be equivalent, we will also have an example telling us that the hypothesis of Proposition 12.2 is not necessary for the conclusion to hold.

By Proposition 12.1, it suffices to show: For every $a \in X$ and $\epsilon > 0$ there are $\delta, \eta > 0$ such that both $B_d(a, \delta) \subseteq B_e(a, \epsilon)$ and $B_e(a, \eta) \subseteq B_d(a, \epsilon)$ hold.

First we set $\delta = \epsilon$. For indeed, $e(x, y) \leq d(x, y)$; hence if $d(a, x) \leq \epsilon$, so too we have $e(a, x) \leq \epsilon$. Thus $B_d(a, \epsilon) \subseteq B_e(a, \epsilon)$.

Now set $\eta = \min\{\epsilon, 1\}$. If $x \in B_e(a, \eta)$, then $e(a, x) < \eta \leq 1$. If $d(a, x) \leq 1$, then $e(a, x) = d(a, x)$, so $d(a, x) < \eta \leq \epsilon$. Hence, in this case, $x \in B_d(a, \epsilon)$. If $d(a, x) > 1$, then $e(a, x) = 1$, a contradiction. Thus $B_e(a, \eta) \subseteq B_d(a, \epsilon)$, as desired. □

Exercises 12.

(1) Show that any two closed bounded intervals, of positive length, in the real line are homeomorphic. When are they isometric?
(2) Show that any two metric spaces with the standard discrete metric are isometric if and only if they are equinumerous.
(3) If $\langle X, d\rangle$ is a metric space, let $e(x, y) = \dfrac{d(x, y)}{1 + d(x, y)}$, and show that e is a bounded metric topologically equivalent to d.
(4) If $a, b \in \mathbb{R}^n$ (euclidean metric), show there is an isometry from $\mathbb{R}^n$ to itself that takes a to b.

Lecture 13: Subspaces and Product Spaces

Abstract. By restricting the metric of a metric space to pairs in a subset of the space, we create a metric subspace. We show that—for example—the resulting open (resp., closed) sets in the subspace are intersections of open (resp., closed) sets in the larger space with the given subset.

We also look at how to define metrics on cartesian products, given we have metrics on the factor spaces.

Subspaces. If $\langle X, d\rangle$ is a metric space and $A \subseteq X$, then the restriction $d_A = d|(A \times A)$ of $d : X \times X \to \mathbb{R}$ to $A \times A$ defines a metric on A, called the *subspace metric*, or the metric *induced on A by d*. This extremely simple construct has enormous consequences.

Examples 13.1. (1) The subspace metric idea turns *any* subset of a metric space into another metric space. For example, in the euclidean plane, let A be the three-quarter circle $\{\langle x, y\rangle : x^2 + y^2 = 1$ and either $x \le 0$ or $y \le 0\}$. (So we remove all pairs with positive coordinates.) Then the induced euclidean metric on A measures the straight line distance between two points, not the distance along the shortest arc. For example, the distance between $\langle 1, 0\rangle$ and $\langle 0, 1\rangle$ using the subspace metric is $\sqrt{2}$; whereas the distance along the shortest arc is $3\pi/2$. This difference, where the measure of distance is along a *geodesic*, or minimal curve in the subspace, as against an as-the-crow-flies distance, is at the basis of a lot of science fiction: When we see a distant star, we're actually viewing the light from the star after it has traveled along a geodesic path in the physical universe. But what if the universe were "folded" on itself in such a way that you could travel to the star in a "worm hole" that is exterior to the universe. Going back to our three-quarter circle A, suppose a "star" at $\langle 1, 0\rangle$ sends light to "earth" at $\langle 0, 1\rangle$. Then this light has to stay in A, so must travel along the only path it can, and this path has length $3\pi/2$. Meanwhile, the "worm hole" distance between the star and earth is merely $\sqrt{2}$.

(2) Recall from Exercise 7(5) that a subset A of $\langle X, d\rangle$ is *bounded* if there is some real M such that $d(x, y) \le M$ for all $x, y \in A$. If A and B are subsets of X that are homeomorphic as subspaces, and if A is bounded, is B bounded too? The answer is no, because the tangent function defines a homeomorphism between $(-\pi/2, \pi/2)$ and $(-\infty, \infty)$. What if we replace *bounded* with other words, say, *closed* or *open*? The answers to both these questions are equally no, but there are restricted situations where the answer is a resounding yes. For example, if $\langle X, d\rangle$ is euclidean n-space, and if A is closed and bounded, then any $B \subseteq \mathbb{R}^n$ homeomorphic to A is also closed and bounded (the Heine-Borel Theorem). The same is true if *closed and bounded* is replaced by *open* (Brouwer's Invariance of Domain Theorem).

The next fact is easy to prove—but well worth noting—and is left as an exercise.

Proposition 13.2. *Let $\langle X, d\rangle$ be a metric space, with $A \subseteq X$. Then a subset B of A is open (closed) relative to the subspace metric d_A if and only if there is an open (closed) set $G \in \mathcal{T}_d$ such that $B = A \cap G$.*

Product Spaces. In Lecture 2 we defined the set-theoretic construction of cartesian product; here we extend that idea to the definition of a cartesian product *metric*. We work with two spaces here, but the construction extends naturally to products of any finite number of spaces.

So suppose $\langle X, d\rangle$ and $\langle Y, e\rangle$ are two metric spaces. What is the most natural way to define a metric on $X \times Y$? Well, following the example if taking the product of two real lines to form the plane, we could define a "euclidean" product metric $(d \times_E e)(\langle x, y\rangle, \langle u, v\rangle) = \sqrt{d(x, u)^2 + e(y, v)^2}$, a "taxicab" product metric $(d \times_T e)(\langle x, y\rangle, \langle u, v\rangle) = d(x, u) + e(y, v)$, or a "max" product metric $(d \times_M e)(\langle x, y\rangle, \langle u, v\rangle) = \max\{d(x, u), e(y, v)\}$. So many choices, which to choose?

The good news is that, while these three candidates for a product metric give you differing distances between a given pair of points in $X \times Y$ (i.e., pair of pairs), there is no difference among them topologically. The inequalities corresponding to the ones in Example 12.3 ($n = 2$) still hold, so all three product metrics give rise to the same open sets in $X \times Y$.

So let us choose the simplest one to deal with, namely the "max" version. That is, for the purposes of these lectures, we choose $(d \times e)(\langle x, y\rangle, \langle u, v\rangle)$ to be $\max\{d(x, u), e(y, v)\}$. One advantage of this choice is that ball neighborhoods look like "squares."

Proposition 13.3. *Let $\langle X, d\rangle$ and $\langle Y, e\rangle$ be metric spaces, with the product metric $\rho = d \times e$ as defined above. If $\langle a, b\rangle \in X \times Y$ and $\epsilon > 0$, then $B_\rho(\langle a, b\rangle, \epsilon) = B_d(a, \epsilon) \times B_e(b, \epsilon)$.*

Proof. Suppose $\langle x, y\rangle \in B_\rho(\langle a, b\rangle, \epsilon)$. Then $\rho(\langle x, y\rangle, \langle a, b\rangle) < \epsilon$, so it is the case that $\max\{d(x, a), e(y, b)\} < \epsilon$; i.e., that $d(x, a) < \epsilon$ and $e(y, b) < \epsilon$ both hold. Hence $\langle x, y\rangle \in B_d(a, \epsilon) \times B_e(b, \epsilon)$.

The reverse inclusion is just as easy, and is left as an exercise.

□

Examples 13.4. Let $\mathbb{I} = [0, 1] = \{x \in \mathbb{R} : 0 \leq x \leq 1\}$ be the closed unit interval in the euclidean line, and let $\mathbb{S}^1 = \{\langle x, y\rangle \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ be the unit circle in the euclidean plane, where both $\mathbb{I}$ and $\mathbb{S}^1$ are given the subspace metric.

(1) $\mathbb{I}^2 = \mathbb{I} \times \mathbb{I} \subseteq \mathbb{R}^2$ is the closed unit square; in general, $\mathbb{I}^n$, the n-fold product of $\mathbb{I}$, is the closed unit n-cube. It is an important advanced topological fact that no m-cube can be homeomorphic to a subspace of $\mathbb{R}^n$ for $n < m$.

(2) $\mathbb{I} \times \mathbb{S}^1 \subseteq \mathbb{R}^3$, a cylindrical tube of unit height and cross-sectional radius, is homeomorphic to the planar annulus $\{\langle x, y\rangle \in \mathbb{R}^2 : 1 \leq x^2 + y^2 \leq 2\}$. A is not homeomprphic to any subspace of $\mathbb{R}$, as we will soon develop the technology to prove.

(3) $\mathbb{S}^1 \times \mathbb{S}^1 \subseteq \mathbb{R}^4$ is a "circle of circles," or the *standard torus* $\mathbb{T}^{\nvDash}$. The torus is homeomorphic to a surface of revolution in $\mathbb{R}^3$, namely where the circle $(x - R)^2 + z^2 = r^2$ in the xz-plane, $0 < r < R$, is revolved about

the z-axis. The resulting "inner tube" is defined by the implicit equation $(R-\sqrt{x^2+y^2})^2+z^2=r^2$. Cubes, tubes, and tori will be some of the topics we take up later in the course.

Exercises 13. (1) Prove Proposition 13.2.

(2) Prove the second half of Proposition 13.3.
(3) Let $\langle X, d\rangle$ be a metric space, and let $A \subseteq X$ be equipped with the subspace metric. Show the *inclusion map* $i : A \to X$, given by $i(x) = x$, $x \in A$, is continuous.
(4) Let $\langle X_1, d_1\rangle, \ldots, \langle X_n, d_n\rangle$ be n metric spaces. Define the "euclidean," the "taxicab," and the "max" metrics on the n-fold product, as an extension of what we defined above. Check that these are all legitimate metrics. (Hint. It's the triangle inequality every time.)
(5) Prove that if $\langle X_1, d_1\rangle$ is homeomorphic to $\langle Y_1, e_1\rangle$ and $\langle X_2, d_2\rangle$ is homeomorphic to $\langle Y_2, e_2\rangle$, then $\langle X_1 \times X_2, d_1 \times d_2\rangle$ is homeomorphic to $\langle Y_1 \times Y_2, e_1 \times e_2\rangle$.
(6) Let $\langle X, d\rangle$ and $\langle Y, e\rangle$ have the standard discrete metrics. What are the values of $\rho(\langle x, y\rangle, \langle u, v\rangle)$ for distinct points in $X \times Y$, for the three product metrics described above? Which of these—if any—gives you the standard discrete metric on $X \times Y$?
(7) Prove that if U is open in $\langle X, d\rangle$ and V is open in $\langle Y, e\rangle$, then the "open rectangle" $U \times V$ is open in $\langle X, d\rangle \times \langle Y, e\rangle$.
(8) Assume $f : X \to Y$ is a continuous function between metric spaces. Show that f, considered as a subset of $X \times Y$, is closed. This fact is the easy half of what is known as the Closed Graph Theorem. Under special circumstances; i.e., when the spaces are Banach spaces and f is a linear transformation, the converse is true. Show the converse to be false in general by exhibiting a noncontinuous function from $\mathbb{R}$ to itself, whose graph is closed.
(9) Let $f : [0, 2\pi) \to \mathbb{S}^1$ map the half-open interval onto the unit circle by the rule $t \mapsto \langle \cos t, \sin t\rangle$. Show that f is a continuous bijection, but not a homeomorphism.

Lecture 14: Complete Metric Spaces I

Abstract. A Cauchy sequence is one where, given any positive distance, there is a tail of the sequence, all of whose entries are closer to one another than that given distance. If a sequence is convergent, then it is Cauchy; a metric space is complete if the converse holds. The main result is Cauchy's criterion, which states that the euclidean metrics are complete.

Cauchy Sequences. It is a rare calculus course that includes Cauchy sequences in its study of convergence because nobody ever uses the concept in engineering applications. However, in more advanced pure mathematics courses, the distinction between a Cauchy sequence and a convergent sequence assumes a much larger role. Recall from Lecture 9 what it means for a sequence $\langle a_n \rangle$ in a metric space $\langle X, d \rangle$ to *converge* to a point $a \in X$. Intuitively, as n gets "large," a_n gets "close" to its target a. With Cauchy sequences, we no longer talk of a target point outside the sequence; rather we make a statement about how "crowded together" the tails of the sequence must be.

Definition 14.1. Let $\langle a_n \rangle$ be a sequence in metric space $\langle X, d \rangle$. The sequence is *Cauchy*, if for all $\epsilon > 0$, there is an $N \in \mathbb{N}$ such that $d(a_m, a_n) < \epsilon$ for all $m, n \geq N$.

In other words, with Cauchy sequences, the diameters of the tails tend to zero.

Proposition 14.2. *Every convergent sequence in a metric space is Cauchy.*

Proof. Suppose $\langle a_n \rangle$ is a sequence in $\langle X, d \rangle$, with limit a. For $\epsilon > 0$, we find $N \in \mathbb{N}$ such that $d(a_n, a) < \epsilon/2$ for $n \geq N$. Thus, if $m, n \geq N$, we have $d(a_m, a_n) \leq d(a_m, a) + d(a_n, a) < \epsilon/2 + \epsilon/2 = \epsilon$.

□

A sequence in a meteric space is called *bounded* if the trace (i.e., the set of points) of the sequence is a bounded set in the space.

Proposition 14.3. *Every Cauchy sequence in a metric space is bounded.*

Proof. Pick $N \in \mathbb{N}$ such that $|a_n - a_N| < 1$, for all $n \geq N$. Then, setting $M = \max\{|a_0 - a_N|, \cdots, |a_{N-1} - a_N|, 1\}$, we see that $2M$ is at least as great as the diameter of the trace of our sequence.

□

A metric space in which all Cauchy sequences converge to an element of the space is called *complete*. First we present some easy examples of metric spaces that are not complete.

Examples 14.4. (1) Let $\langle X, d \rangle$ be the open ray $(0, \infty)$ of positive real numbers, with the inherited euclidean metric. Then $\langle 1/n \rangle_{n=1}^{\infty}$ is a sequence in

X that converges to 0 *in* $\mathbb{R}$; however it converges to no point in X itself. It is still a Cauchy sequence in X, though; hence $\langle X, d\rangle$ is not a complete metric space.

(2) Let $\langle X, d\rangle$ be the rational points in $\mathbb{R}$, again with the inherited euclidean metric. Using the fact that any real number may be approximated by rational numbers, let $\langle a_n\rangle$ be a sequence in $\mathbb{Q}$ that converges in $\mathbb{R}$ to, say, the irrational number $\sqrt{2}$. (For example, we could have the sequence $\langle 1, 1.4, 1.41, 1.414, \cdots\rangle$.) Then we have a Cauchy sequence of rational numbers that converges to no rational number. (Note that for both these examples, we use the fact that no sequence has more than one limit.)

The following theorem is a basic result of nineteenth century analysis, though the current phraseology is more up-to-date.

Proposition 14.5 (Cauchy's Criterion). *Euclidean space is complete.*

Proof. Suppose, for the moment, that the dimension is 1 and that $\langle a_n\rangle$ is a Cauchy sequence in $\mathbb{R}$. As in the proof of Proposition 14.3, we find $N_0 \in \mathbb{N}$ such that $|a_n - a_{N_0}| < 2^{-0} = 1$ for all $n \geq N_0$, but now we put in a new twist. Let $\alpha_0 = \text{glb}\{a_n : n \geq N_0\}$ and $\beta_0 = \text{lub}\{a_n : n \geq N_0\}$. Then $\alpha_0 \leq \beta_0$ and $\beta_0 - \alpha_0 \leq 2$. Next, we use the fact that the tail $\langle a_n\rangle_{n=N_0}^{\infty}$ is a Cauchy sequence in the interval $[\alpha_0, \beta_0]$ to find $N_1 \geq N_0$ in $\mathbb{N}$ such that $|a_n - a_{N_1}| < 2^{-1}$ for all $n \geq N_1$. Let $\alpha_1 = \text{glb}\{a_n : n \geq N_1\}$ and $\beta_1 = \text{lub}\{a_n : n \geq N_1\}$. Then $\alpha_0 \leq \alpha_1 \leq \beta_1 \leq \beta_0$ and $\beta_1 - \alpha_1 \leq 2(2^{-1}) = 1$. Proceeding by induction, we develop $N_0 \leq N_1 \leq \cdots$ in $\mathbb{N}$ such that, for each $k \in \mathbb{N}$, $|a_n - a_{N_k}| < 2^{-k}$ for $n \geq N_k$. Then, setting $\alpha_k = \text{glb}\{a_n : n \geq N_k\}$ and $\beta_k = \text{lub}\{a_n : n \geq N_k\}$, we have $\alpha_0 \leq \alpha_1 \leq \cdots, \cdots \leq \beta_1 \leq \beta_0$ and, for each $k \in \mathbb{N}$, $\beta_k - \alpha_k \leq 2(2^{-k}) = 2^{1-k}$.

The increasing sequence $\alpha_0 \leq \alpha_1 \leq \cdots$ is bounded above by β_0, so there is a least upper bound a. Since every β_k also serves as an uppor bound for the α_n, we know $a \leq \beta_k$ for each $k \in \mathbb{N}$; so a is a lower bound for the decreasing sequence $\beta_0 \geq \beta_1 \geq \cdots$. Let b be the greatest lower bound of the β_k. Then we have the infinite chain of relations

$$\alpha_0 \leq \alpha_1 \leq \cdots \leq a \leq b \leq \cdots \leq \beta_1 \leq \beta_0$$

holding.

Since $[a, b]$ is nonempty and contained in $[\alpha_k, \beta_k]$ for all $k \in \mathbb{N}$, and since $\beta_k - \alpha_k \leq 2^{1-k}$, we know that $b - a \leq 2^{1-k}$ for every $k \in \mathbb{N}$. It follows from the Archimedean property of the real line, that $b - a = 0$; i.e., $a = b$.

If $\epsilon > 0$, we pick k large enough so that $[\alpha_k, \beta_k] \subseteq (a - \epsilon, a + \epsilon)$. Then, for $n \geq N_k$ we have that $|a_n - a| < \epsilon$, showing that $a = \lim_{n\to\infty} a_n$.

The proof for $\mathbb{R}^p$, $p \geq 1$, flows quite naturally from the one-dimensional case. If $\langle a_n\rangle$ is a Cauchy sequence in $\mathbb{R}^p$, then each a_n is a p-tuple $\langle a_n^1, \cdots, a_n^p\rangle$, and for $1 \leq i \leq p$, $\langle a_n^i\rangle_{n=0}^{\infty}$ is a Cauchy sequence in $\mathbb{R}$. (Why?) By the proof above, then, we have, for each $1 \leq i \leq p$, a limit $a^i = \lim_{n\to\infty} a_n^i$. Then the p-tuple $\langle a^1, \cdots, a^p\rangle$ is the limit of the original sequence of p-tuples.

□

Some Topological Consequences of Completeness. If d is a complete metric on X and e is a metric on X that is topologically equivalent to d (i.e., $\mathcal{T}_e = \mathcal{T}_d$), is it true that e is a complete metric too? Recalling Proposition 12.4, if $e(x, y) = \min\{d(x, y), 1\}$ is the standard truncation of d to a bounded metric, then one metric is complete if and only if the other is. (Exercise.) But this fact is more the exception than the rule; in general it is quite possible to have two topologically equivalent metrics, only one of them complete.

Example 14.6. Recall from Proposition 14.5 that the euclidean real line $\mathbb{R}$ is complete; however, by Example 14.4(1), the open subspace $(0, \infty)$ is not. Now the exponential function $x \mapsto 2^x$ defines a homeomorphism between $\mathbb{R}$ and $(0, \infty)$, and converts the euclidean metric on $(0, \infty)$ to a metric $e(x, y) = |2^x - 2^y|$ on $\mathbb{R}$ that is equivalent to the euclidean metric but nevertheless incomplete: the divergent sequence $\langle -n \rangle_{n=1}^{\infty}$ is Cauchy in the space $\langle \mathbb{R}, e \rangle$.

In light of the discussion above, define a metric space $\langle X, d \rangle$ to be *topologically complete* if there is a complete metric e on X that is topologically equivalent to d. The next question one might well ask is: “Are all metric spaces topologiclly complete?” After all, with *bounded* replacing *complete*, the answer—by Proposition 12.4—is yes. So it is pointless to define a metric space to be “topologically bounded,” as this does not give us anything new. However, in this case we have something new. We will show that, for example, the rational line with the euclidean metric is not only incomplete as shown in Example 14.4(2), it is topologically incomplete as well.

Exercises 14.

(1) Let d be a metric on X, with $e(x, y) = \min\{d(x, y), 1\}$ the standard truncation of d. Show that d is complete if and only if e is complete.

(2) Prove that if $\langle X, d \rangle$ is a complete metric space and $A_0 \supseteq A_1 \supseteq A_2 \supseteq \cdots$ is a *nested sequence* of closed sets such that the sequence $\langle D(A_n) \rangle$ if diameters converges to 0, then $\bigcap_{n=0}^{\infty} A_n$ has exactly one point. (This is called Cantor’s Intersection Theorem.)

(3) Show by example that in Problem (2), $\bigcap_{n=0}^{\infty} A_n$ can be empty if the requirement that that the diameters go to zero is deleted, or even replaced with the weaker condition that the sets be bounded.

Lecture 15: Complete Metric Spaces II

Abstract. A subset of a metric space is residual if it contains a dense open set. Baire's theorem states that, in a complete metric space, the intersection of countably many residual sets is dense.

We also outline a proof that any metric space may be isometrically embedded as a dense subspace of a complete metric space. This process of completion generalizes how one obtains the real line from the rational line.

Baire's Theorem. One of the most profound and far-reaching consequences of completeness is Baire's theorem, also known as the Baire Category Theorem. First define a subset A of a metric space $\langle X, d\rangle$ to be *dense* if $\mathrm{cl}(A) = X$; i.e., if every nonempty open subset of X intersects A. The rational line is dense in the real line, for example. In some sense dense subsets of a space are "large" in the space, but not quite large enough: the complement of a dense set can also be dense—witness the irrational line in $\mathbb{R}$—so two dense sets can, for example, be disjoint.

A stronger notion of "large subset" is that of residual set. $A \subseteq X$ is *residual* in X if A contains a dense open set.

Proposition 15.1. *In any metric space, the intersection of a finite number of residual subsets is residual.*

Proof. The proof is by induction on the number n of residual sets; the inductive step is easy once we prove the assertion for $n = 2$.

So suppose A and B are residual, say $A \supseteq U$ and $B \supseteq V$, where U and V are dense open sets. Then $A \cap B \supseteq U \cap V$, an open set; hence we are done once we show $U \cap V$ is dense: any superset of a dense set is automatically dense. If W is any nonempty open set, then—because U is dense open—$W \cap U$ is a nonempty open set. And since V also is dense, we have $\emptyset \neq (W \cap U) \cap V = W \cap (U \cap V)$. Thus $A \cap B$ contains a dense open set, and is hence residual.

□

So residual sets are, in a believable sense, large subsets of a metric space. Hence it makes sense to think of complements of residual sets as "small." Officially, a subset of a metric space is called *nowhere dense* if its complement is residual. It is easy to show, and left as an exercise, that $A \subseteq X$ is nowhere dense if and only if $\mathrm{int}(\mathrm{cl}(A)) = \emptyset$. A rephrasing of Proposition 15.1, then, is that a finite union of nowhere dense sets is nowhere dense. For example, a singleton set is nowhere dense if and only if it's not open; i.e., just in case its single point is not isolated. Hence a finite set of nonisolated points is nowhere dense.

We now shift our attention to infinitary Boolean operations and ask the question whether countable intersections of residual sets are residual. The answer is definitely no in general.

Example 15.2. Let our metric space be the rational line $\mathbb{Q}$, with the usual euclidean metric. Since $\mathbb{Q}$ is a countable set, let $q_0, q_1, \cdots$ be a list of all the rational numbers; and, for each $n \in \mathbb{N}$, let $A_n = \mathbb{Q} \setminus \{q_n\}$. Then, since singletons are always closed in a metric space, and no point is isolated in $\mathbb{Q}$, each A_n is a dense open set.

However $\bigcap_{n=0}^{\infty} A_n = \emptyset$, about as far from being residual as you can get.

Baire's theorem tells us that, in the presence of completeness, countable intersections of residual subsets may not be residual, but at least they're "reasonably large."

Proposition 15.3 (Baire's Theorem). *Let $\langle X, d\rangle$ be a complete metric space, and for each $n \in \mathbb{N}$, let $A_n \subseteq X$ be residual. Then $\bigcap_{n=0}^{\infty} A_n$ is a dense subset of X.*

Proof. For each $n \in \mathbb{N}$, let $A_n \supseteq U_n$, where U_n is dense open. It suffices to show $\bigcap_{n=0}^{\infty} U_n$ is dense. Let V be a fixed nonempty open set; we wish to show $V \cap \bigcap_{n=0}^{\infty} U_n \neq \emptyset$.

Now if W is an open set with $x \in W$, then for any $\epsilon > 0$ with $B(x, \epsilon) \subseteq W$, we have the closed ball $B[x, \epsilon/2] \subseteq B(x, \epsilon)$. Hence we have $\text{cl}(B(x, \epsilon/2)) \subseteq B(x, \epsilon)$.

With this in mind, let B_0 be an open ball, of radius < 1, such that $\text{cl}(B_0) \subseteq V \cap U_0$. For the next step, we may pick an open ball B_1, of radius $< 1/2$, such that $\text{cl}(B_1) \subseteq B_0 \cap U_1$. For the third step, we pick open ball B_2, of radius $< 1/3$, such that $\text{cl}(B_2) \subseteq B_1 \cap U_2$. This process may be continued indefinitely, with the result that we have a nested sequence $V \supseteq \text{cl}(B_0) \supseteq B_0 \supseteq \text{cl}(B_1) \supseteq B_1 \supseteq \text{cl}(B_2) \supseteq B_2 \supseteq \cdots$, where, for all $k \in \mathbb{N}$, we have $\text{cl}(B_k) \subseteq V \cap \bigcap_{n=0}^{k} U_n$.

At this point you may either invoke Cantor's Intersection Theorem (Exercise 14(2)) or simply note that the centers x_n of the balls B_n form a Cauchy sequence since the diameters of the balls tend to zero. If $x = \lim_{n\to\infty} x_n$, then we have $x \in V \cap \bigcap_{n=0}^{\infty} U_n$, as desired. □

Note that Proposition 15.3 deals only with topological concepts, and only requires the existence of a complete metric equivalent to a given one. Thus we could actually replace *complete metric space* in the hypothesis with *topologically complete metric space* and the argument would still go forward. As an immediate corollary (exercise), we know that the space of rational numbers, as a subspace of $\mathbb{R}$, is not topologically complete.

It is easy to prove (exercise) that a subspace of a complete metric space is complete in the subspace metric if and only if the subspace is closed. However, the question arises as to which subspaces of a complete metric space are topologically complete. The answer is fairly subtle—check Wikipedia if you're interested—but the short answer is that a subspace A of a complete metric space $\langle X, d\rangle$ is topologically complete if and only if $A = \bigcap_{n=0}^{\infty} U_n$, where each U_n is open in X. (Such subsets are called G_δ *sets* in the literature. All open sets are obviously G_δ sets, by definition. It is also true, however, that in metric spaces, all closed sets are G_δ sets too (exercise).)

Completion. There is a completeness theorem—due mainly to Felix Hausdorff (1914)—that shows how we may view incomplete metric spaces as non-closed sets of points in a complete space. So rather than produce metric spaces as subspaces of given ones, this process produces a "superspace" that is complete, and in which the original space sits as a dense subspace. This process parallels the development of the reals from the rationals, but using Cauchy sequences rather than the cuts talked about in Lecture 6. The problem in the general metric case is that there is no linear order to fall back on.

In Lecture 11 we defined *isometry* as a perticularly "rigid" kind of homeomorphism. If $f : \langle X, d\rangle \to \langle Y, e\rangle$ is an isometry onto a subspace of $\langle Y, e\rangle$, we call f an *isometric embedding.* This notion is key if we are to start with one space and construct another for which the first space is a subspace.

Proposition 15.4 (Completion Theorem). *Let $\langle X, d\rangle$ be a metric space. Then there is a complete metric space $\langle Y, e\rangle$ and an isometric embedding $f : X \to Y$ for which $f[X]$ is a dense subspace of Y. The space $\langle Y, e\rangle$ is "unique," in the sense that if $\langle Y', e'\rangle$ is complete and $f' : X \to Y'$ is an isometric embedding for which $f'[X]$ is a dense subspace of Y', then there is an isometry $H : Y \to Y'$ such that $H \circ f = f'$. If d is a complete metric to begin with, then f is an isometry, and we get nothing new.*

Proof. We present an outline of the proof; for while a detailed proof is quite lengthy, its basic idea is straightforward.

Step 1. Beginning with the metric space $\langle X, d\rangle$, let $\mathcal{C}$ be the set of all Cauchy sequences in X. For $\langle x_n\rangle$ and $\langle y_n\rangle$ in $\mathcal{C}$, define $\langle x_n\rangle \sim \langle y_n\rangle$ just in case the sequence $\langle d(x_n, y_n)\rangle$ of real numbers has limit 0.

Step 2. Verify that $\sim$ is an equivalence relation on $\mathcal{C}$, and let Y be $\mathcal{C}/\sim$, the set of $\sim$-equivalence classes $[\langle x_n\rangle]$.

Step 3. For $\langle x_n\rangle$ and $\langle y_n\rangle$ in $\mathcal{C}$, the real seqence $\langle d(x_n, y_n)\rangle$ is Cauchy (exercise), and hence converges. This allows us to define $e([\langle x_n\rangle], [\langle y_n\rangle]) = \lim_{n\to\infty} d(x_n, y_n)$. We then verify that this definition is independent of choice of equivalence class representatives, and hence makes sense as a function from $Y \times Y$ to $\mathbb{R}$.

Step 4. Verify that e defines a complete metric on Y, and that the function $f : X \to Y$, defined by taking a point $x \in X$ to $[\langle x, x, \cdots\rangle]$, is an isometric embedding onto a dense subset of $\langle Y, e\rangle$, which is an isometry (i.e., onto) just in case d itself is complete.

Step 5. Prove uniqueness as follows: Let $f' : \langle X, d\rangle \to \langle Y', e'\rangle$ be an isometric embedding onto a dense subspace of a complete metric space. Then, for each $x \in X$, define $h(f(x)) = f'(x)$. This clearly gives an isometry from $f[X]$ onto $f'[X]$. We extend h to $H : Y \to Y'$ by noticing that each $y \in Y \setminus f[X]$ is the limit of a sequence $\langle y_n\rangle$ of members of $f[X]$ that converges to y. Then $\langle h(y_n)\rangle$ is a Cauchy sequence in $f'[X]$, which converges to a unique $y' \in Y'$. $H(y)$ is now defined to be this y', and gives us the isometry we want. Clearly $H \circ f = f'$; i.e., $H(f(x)) = f'(x)$ for all $x \in X$.

□

Exercises 15. (1) Show that $A \subseteq X$ is nowhere dense—i.e., the complement in X of a residual set—if and only if $\text{int}(\text{cl}(A)) = \emptyset$.

(2) Show by example that, even for a complete metric space, the intersection of countably many residual sets needn't be residual.
(3) Show that the space of rationals—euclidean metric—is not topologically complete.
(4) Let $\langle X, d\rangle$ be a complete metric space, with $A \subseteq X$. Show that the subspace metric d_A is complete if and only if A is closed in X.
(5) Prove that in a metric space, each closed set is a countable intersection of open sets.
(6) If $\langle x_n\rangle$ and $\langle y_n\rangle$ are two Cauchy sequences in $\langle X, d\rangle$, show that the real sequence $\langle d(x_n, y_n)\rangle$ converges.

Lecture 16: Quotient Spaces I

Abstract. The intuition behind forming quotients of metric (or topological) spaces is found in how we glue the ends of a strip of paper to form either a tube or a Möbius band. Formally the process involves placing an equivalence relation on a space and then finding the proper way to define the open sets on the set of equivalence classes. Another approach is to define an identification map to be a continuous onto map with the property that only the open sets in the range have open pre-images in the domain. The natural mapping from a space to a quotient space is an identification map; conversely, the image of a space under an identification map is homeomorphic to a quotient of the space.

Cut-and-Paste. The idea of quotient space provides a theoretical backbone to the very intuitive notion of "cut-and-paste" constructions. Let's start with an example.

Example 16.1. Take a rectangular strip of paper, the closed unit square $\mathbb{I}^2$, say, roll it into a tube, and glue the top and bottom edges together. Somewhat more explicitly, we "identify" each top-edge point $\langle t, 1\rangle$ with its corresponding bottom-edge point $\langle t, 0\rangle$, $0 \leq t \leq 1$. This provides a construction of a *standard band*, an example of what is known in the trade as a compact 2-manifold with boundary consisting of two disjoint simple closed curves.

Alternatively, prior to gluing the top and bottom edges together, give the strip a 180-degree twist. This amounts to identifying each top-edge point $\langle t, 1\rangle$ with the bottom-edge point $\langle 1-t, 0\rangle$, and results in a *Möbius band*, a compact 2-manifold with boundary consisting of one simple closed curve.

The process just described may be made mathematically precise as follows. First, start with a metric space X; in the example above, X is the closed unit square. Next, specify an equivalence relation R on X; in the standard band example, two points are identified (equivalent) if and only if they're either equal, or one is a top-edge point and the other is a bottom-edge point with the same first coordinate. Third, consider the equivalence classes $[x]$, $x \in X$, as elements of the quotient set X/R, *à la* Lecture 4. Finally—and this is the key step— we describe how to make X/R into a quotient *space*.

Recall from Proposition 8.3 that the topology $\mathcal{T}_d$ induced by a metric d on set X satisfies some Boolean conditions that are not metric-specific: Both X and $\emptyset$ are open; finite intersections of open sets are open; arbitrary unions of open sets are open. A family of subsets of a set satisfying these three conditions is generally called a *topology*, regardless of whether there is a metric involved. With this in mind, let us define a collection $\mathcal{B} \subseteq X/R$ of equivalence classes to be *open* in the quotient space just in case the union of the equivalence classes constituting $\mathcal{B}$ is open in X. In other words, the open sets of X/R are precisely those families $\mathcal{B} \subseteq X/R$ such that $\bigcup \mathcal{B}$ is open in X. If we let $\mathcal{Q}$ be the collection of these open sets, it is straightforward to prove that $\mathcal{Q}$ satisfies the Boolean conditions that define what it means to be a topology.

The question naturally arises as to whether there is an appropriate metric on X/R whose topology is $\mathcal{Q}$. The answer to this is, in general, no; however, in the most important cases we consider in this course, those where X is a closed bounded subspace of euclidean space and each equivalence class is closed in X, the answer is yes: if $[x]$ and $[y]$ are two distinct equivalence classes—necessarily disjoint—the set of distances $\{d(x', y') : x' \in [x], y' \in [y]\}$ has a minimum value, which is positive. Let this value be the distance between $[x]$ and $[y]$ in the quotient. An explanation of why this works is deferred until we study compactness in Lecture 24. What is important for us here is that we know what the open sets are in the quotient space, not what matric may or may not product them.

The assignment taking a point $x \in X$ to the unique equivalence class $[x] \in X/R$ containing x defines a surjection $q : X \to X/R$, known as the *quotient map* associated with X and R (see Exercise 4.5). This function is deserving of the designation *map* because if $\mathcal{B} \subseteq X/R$, then $q^{-1}[\mathcal{B}] = \bigcup \mathcal{B}$. By the definition of what it means to be an open set in the quotient space, we see that not only is it the case that q pulls open sets back to open sets, but it pulls non-open sets back to non-open sets.

Every function $f : X \to Y$ determines an equivalence relation R_f on X, namely we write $xR_f y$ just in case $f(x) = f(y)$. This equivalence relation is often called the *kernel* of f.

Proposition 16.2 (Transgression Theorem). *Let $q : X \to X/R$ be a quotient map, with $f : X \to Y$ an arbitrary mapping. Then there is a unique continuous $g : X/R \to Y$ with $g \circ q = f$ if and only if $R \subseteq R_f$.*

Proof. Suppose such a $g : X/R \to Y$ exists, and that xRy. Then $[x] = [y]$, and so $f(x) = g(q(x)) = g([x]) = g([y]) = g(q(y)) = f(y)$. Hence $xR_f y$. Conversely, suppose $R \subseteq R_f$; and for each $x \in X$, define $g([x]) = f(x)$. Does this depend upon the choice of equivalence class representative x? Well no, because if $[x] = [x']$ then xRx'; hence $xR_f x'$ and thus $f(x) = f(x')$. This tells us there is a function $g : X/R \to Y$ with $g \circ q = f$. The uniqueness of g is apparent; as for continuity, suppose $V \subseteq Y$ is open. Then $f^{-1}[V]$ is open in X, since f is continuous. But $f = g \circ q$, so $f^{-1}[V] = (g \circ q)^{-1}[V] = q^{-1}[g^{-1}[V]]$. Since this set is open in X and q is a quotient map, we infer that $g^{-1}[V]$ is open in X/R.

□

Identification Maps. Recall from Proposition 11.4 that a function's being continuous is characterized by the condition that pre-images of open (resp. closed) subsets of the range space are open (resp., closed) in the domain. If $f : X \to Y$ is a continuous surjection, f is called an *identification* map if the only way $f^{-1}[B]$ is open in X is for B to be open in Y. Clearly the quotient maps $q : X \to X/R$ described above are identification maps. Thus $f^{-1}[B]$ is open (resp., closed) in X if and only if B is open (resp., closed) in Y. (We may interchange *open* and *closed* in the definition because inverse images are so well behaved with regard to the Boolean set operations.)

Note that in the statement of Proposition 16.2, we could replace $q : X \to X/R$ by any idenfification map $p : X \to Z$, and the proof would go through, with $R = R_p$. This brings us to the following corollary of Proposition 16.2, the main "cut and paste" result we use.

Proposition 16.3. *Let $q : X \to X/R$ be a quotient map, with $f : X \to Y$ an identification mapping such that $R = R_f$. Then there is a unique homeomorphism $g : X/R \to Y$ such that $g \circ q = f$.*

Proof. We apply Proposition 16.2, first to obtain the map $g : X/R \to Y$, and second, to obtain a map $h : Y \to X/R$. Then $h \circ g : X/R \to X/R$ is the unique mapping satisfying Proposition 16.2 with $f = q$. Since the identity map works, we know $h \circ g$ is the identity map on X/R. Similarly, we know $g \circ h$ is the identity map on Y; hence g and h are homeomorphisms that are inverse to one another.

□

An important source of identification maps consists of the continuous surjections that are either open or closed. We say that a function $f : X \to Y$ is *open* (resp., *closed*) if $f[A]$ is open (resp., closed) in Y whenever A is open (resp., closed) in X. Note that while continuous functions "pull" open sets back to open sets (or closed sets to closed sets), open functions and closed functions "push" open sets over to open sets and closed sets to closed sets. Since forward images can fail to preserve complementation, the notions of *open functin* and *closed function* are distinct.

Proposition 16.4. *If a continuous surjection is either open or closed, then it is an identification map.*

Proof. Suppose $f : X \to Y$ is a continuous surjection that is also an open map. If $B \subseteq Y$ is such that $f^{-1}[B]$ is open in X, then $B = f[f^{-1}[B]]$ is open in Y. If f is assumed to be a closed map, just repeat the above with *closed* in place of *open*.

□

Examples 16.5.

(1) Given X and Y, we let $p : X \times Y \to X$ and $q : X \times Y \to Y$ be given by $\langle x, y\rangle \mapsto x$ and $\langle x, y\rangle \mapsto y$, respectively. These are the *standard coordinate projection* maps. Continuity follows because—e.g., $p^{-1}[U] = U \times Y$—open sets get pulled back to open rectangles. These maps are also open. To see this, note that if U and V are open in X and Y respectively, then $U \times V$ is open in $X \times Y$. Then $p[U \times V] = U$, so p pushes open rectangles to open sets. But every open set in $X \times Y$ is a union of open rectangles. And since (see Proposition 3.3) forward images of functions preserve unions, we infer that p—and similarly q—is an open map.

Coordinate projections are identification maps, but they are not necessarily closed. For example, let $C \subseteq \mathbb{R}^2$ be the set $\{\langle x, y\rangle : xy = 1\}$. This is the graph of a hyperbola in the plane, definitely a closed set (exercise). But $p[C] = \mathbb{R} \setminus \{0\}$, decidedly not closed in $\mathbb{R}$.

(2) If $f : X \to Y$ is continuous and X is a closed bounded subspace of euclidean n-space for some n, then f is a closed map. We'll see why this is true when we study compactness.

(3) (For those with some experience with the algebra of complex numbers.) If the euclidean plane is endowed with complex multiplication, then the algebraic structure satisfies the field axioms—see Lecture 6—and so it makes sense to form limits of difference quotients, $f'(z) = \lim_{\Delta z \to 0} \frac{f(z+\Delta z)-f(z)}{\Delta z}$, as in the basic calculus definitioin of derivative. If $\Omega \subseteq \mathbb{R}^2$ is, say, an open disk or open annulus, and $f : \Omega \to \mathbb{R}^2$ is such that $f'(z)$ exists for all

$z \in \Omega$, then we call f *holomorphic* on Ω. The Open Mapping Theorem of complex analysis says that nonconstant holomorphic functions are not only continuous, but opan as well. Since even constant maps are identification maps onto their images, we infer the same for all holomorphic functions.

(4) It is easy to show that the composition of any two open maps is an open map; likewise for closed maps. However, the composition of a closed map and an open map, while still an identification map, may fail to be either open or closed. Indeed, suppose $p : \mathbb{R}^2 \to \mathbb{R}$ is projection onto the first coordinate. We saw in (1) above that p is an open surjection that is not closed. Now let $g : \mathbb{R} \to [0, \infty)$ take all negative reals to 0 and fix everything else. Then g is a closed surjection that is not open, and $f = g \circ p : \mathbb{R}^2 \to [0, \infty)$ is an identification map that is neither open nor closed: If U is any nonempty open set in the left half-plane, then $f[U] = \{0\}$; not closed. And if $C = \{\langle x, y\rangle : x > 0, y > 0, \text{ and } xy = 1\}$, then C is closed in $\mathbb{R}^2$ but $f[C] = (0, \infty)$ is not closed in $[0, \infty)$.

Exercises 16. (1) Let $f : X \to Y$ be an identification map. Show that if $g : Y \to Z$ is any function such that $g \circ f$ is continuous, then g is continuous.

(2) Show that any map having a continuous right inverse (see Exercise 3.3) is an identification map.

(3) Let $X = \mathbb{I}$ be the closed unit interval in the euclidean real line, and declare xRy just in case either $x = y$ or $\{x, y\} = \{0, 1\}$. Show that X/R is homeomorphic to the closed unit circle $\mathbb{S}^1 = \{\langle x, y\rangle \in \mathbb{R}^2 : x^2 + y^2 = 1\}$.

(4) Let $X = \mathbb{I}^2$ be the closed unit square in the euclidean plane, and declare $\langle x, y\rangle R\langle u, v\rangle$ just in case $\langle x, y\rangle = \langle u, v\rangle$ or $\{y, v\} = \{0, 1\}$ and $x = u$. Show that X/R is homeomorphic to the standard band $\mathbb{S}^1 \times \mathbb{I}$.

(5) Give an example of a continuous surjection that is not an identification map.

Lecture 17: Quotient Spaces II

Abstract. The cone over a space is formed in two steps: first take the cartesian product of the space with the closed unit interval; then identify the top of the product—all points with second coordinate 1—with a point. A line segment is the cone over a two-point space; a disk is the cone over a circle. We also describe other classic "cut-and-paste" constructions—e.g., suspensions, attaching spaces—using quotients.

Identifying a Set to a Point. An important class of quotient spaces arises from the identification of a subset of a space to a point. More precisely, if X is a metric space and $A \subseteq X$ is a subset, define the equivalence xR_Ay to mean that either $x = y$ or both x and y are members of A. The resulting quotient space is denoted X/A instead of X/R_A; the equivalence classes are A itself, plus the singletons $\{x\}$ for $x \in X \setminus A$.

Example 17.1. The *cone* $\mathsf{C}(X)$ over a space X is defined to be the quotient space $(X \times \mathbb{I}/(X \times \{1\})$. That is, we form the product $X \times \mathbb{I}$ and pinch the top $X \times \{1\}$ to a point. The mapping $x \mapsto [\langle x, 0\rangle]$, assigning a point of X to the corresponding singleton equivalence class in the base of $\mathsf{C}(X)$ is a homeomorphism onto a closed subset of the cone. In this way we view any space as canonically embedded in its cone.

We may generalize slightly the discussion in the first paragraph by starting with a family $\mathcal{A} = \{A_i : i \in I\}$ of pairwise disjoint subsets of X, and identifying two points just in case they are either equal or both in one of the sets A_i. The resulting quotient space is denoted $X/\mathcal{A}$, and in the special case $\mathcal{A} = \{A\}$, we write X/A to abbreviate $X/\{A\}$.

Example 17.2. The *suspension* $\mathsf{S}(X)$ over a space X is defined to be the quotient space $(X \times [-1, 1])/\{X \times \{-1\}, X \times \{1\}\}$. That is, we form the product $X \times [-1, 1]$ and pinch the top to one point, the bottom to another. In this case we use the assignment $x \mapsto [\langle x, 0\rangle]$ to treat X as a closed subspace of its suspension. In the exercises, you are asked to describe the cones and suspensions of familiar spaces.

Attaching of Spaces. Here we describe an important "cut-and-paste" construction that arises from a map $f : A \to Y$, where $A \subseteq X$. We call it *attaching* X *to* Y *via* f. First, we may assume that the spaces X and Y are disjoint; otherwise we may make them disjoint by identifying X with its homeomorph $X \times \{0\}$ and Y with its homeomorph $Y \times \{1\}$. That said, we form the *free union* $X + Y$ to be the disjoint union $X \cup Y$, where $U \subseteq X \cup Y$ is declared *open* in the free union just in case $U \cap X$ and $U \cap Y$ are open in X and Y, respectively. If d and e are metrics on X and Y, respectively, then we may form a metric on $X \cup Y$ as follows: first, because of Proposition 12.4, we may assume both d and e are bounded, say, by 1; next we define $\rho(x, y)$, for $x, y \in X \cup Y$, to be $d(x, y)$ if $x, y \in X$, to be $e(x, y)$ if $x, y \in Y$, and to be 2 otherwise. It is straightforward to show that the open sets defined for the free union constitute the topology $\mathcal{T}_\rho$. Since X and Y are disjoint, they keep their own metric/topological structure as subspaces of the free union that are both open and closed therein.

So let $f : A \to Y$ be a given continuous map, with A a (closed, in practice) subset of X. Then we form the free union $X + Y$, and, in there, identify each $a \in A$ with its image $f(a) \in Y$. So distinct points in X are identified together just in case they are both in A and are in the kernel of f; no two distinct points in Y are identified together; and $x \in X$ is identified with $y \in Y$ just in case $x \in A$ and $y = f(x)$. The resulting quotient space is denoted $X \cup_f Y$; f is called the *attaching map*. The proof of the following is left as an exercise.

Proposition 17.3. *Let $A \subseteq X$ be closed, $f : A \to Y$ continuous, and $q : X + Y \to X \cup_f Y$ the quotient map. Then $q|Y : Y \to X \cup_f Y$ embeds Y as a closed subspace of $X \cup_f Y$ and $q|(X \setminus A) : X \setminus A \to X \cup_f Y$ embeds $X \setminus A$ as an open subspace of $X \cup_f Y$.*

Examples 17.4.
(1) If $A \subseteq X$, then X/A may be viewed as an attachment $X \cup_f Y$, where Y is a singleton and $f : A \to Y$ is the constant map. This shows us how to view cone spaces via attaching.
(2) If $A, B \subseteq X$ are disjoint nonempty sets, then $X/\{A, B\}$ may be viewed as an attachment $X \cup_f Y$, where Y is the discrete two-point space, and $f : A \cup B \to Y$ takes A to one of the points, B to the other. This shows us how to view suspension spaces via attaching.
(3) If X and Y are two metric spaces, with $x_0 \in X$ and $y_0 \in Y$, then the *join* $\langle X, x_0\rangle \vee \langle Y, y_0\rangle$ is the attaching space $X \cup_f Y$, where $f : \{x_0\} \to Y$ is defined by $f(x_0) = y_0$.

Exercises 17.
(1) What are the cone and suspension over a one-point space? A two-point (discrete) space? An n-point (discrete) space? (The cone over a 3-point discrete space is called a *simple triod*. When $n > 3$, the word *n-od* is often used.)
(2) Describe $\mathsf{C}(X)$ and $\mathsf{S}(X)$, where X is the closed unit interval. The unit circle.
(3) Show that a continuous function $f : X \to Y$ gives rise to continuous maps $\mathsf{C}(f) : \mathsf{C}(X) \to \mathsf{C}(Y)$ and $\mathsf{S}(f) : \mathsf{S}(X) \to \mathsf{S}(Y)$, defined by the assignment $\langle x, t\rangle \mapsto \langle f(x), t\rangle$.
(4) Show that $\mathsf{S}(X)$ and $\mathsf{C}(X)/X$ are homeomorphic.
(5) Let $\langle X, d\rangle$ and $\langle Y, e\rangle$ be metric spaces, with X and Y disjoint, and both d and e bounded by 1. Show that the function ρ defined on $X \cup Y$ as above is a metric, whose open sets consist of all $U \subseteq X \cup Y$ such that $U \cap X \in \mathcal{T}_d$ and $U \cap Y \in \mathcal{T}_e$.
(6) Prove Proposition 17.3.

Lecture 18: Separation Properties

Abstract. We introduce the basic separation axioms; e.g, a space is Hausdorff if for any two points, there are two disjoint open sets, each containing exactly one of the points. The space is normal if the same can be said for two disjoint closed sets as well. The main result is that metric spaces are normal.

Recall that a family $\mathcal{T}$ of subsets of X satisfying the Boolean conditions of Proposition 8.3 is called a *topology* on X. If a metric d produces $\mathcal{T}$, in the manner described in Lecture 8, then $\mathcal{T} = \mathcal{T}_d$; however topologies that do not arise in this way abound, and form the subject matter of the area of mathematics called *general topology*. In this course we are focusing on the topology of metric spaces, but many arguments turn on what it means for a set to be *open* and do not require the metric assumption. Furthermore there are topological constructions—the quotient topology, for instance—that do not necessarily output a metric topology, even when metric topologies are input.

A pair $\langle X, \mathcal{T}\rangle$, where $\mathcal{T}$ is a topology on X, is called a *topological space*. Topological spaces whose topologies arise from metrics are called *metrizable* spaces; note the difference between the terms *metric space* and *metrizable space*. In an instance of the former term, a metric is specified and the topology is implicit. In an instance of the latter term, only the topology is specified.

A class/property of metric spaces is called *topological* if it is invariant under homeomorphism. In this lecture we concentrate on what are called the *separation properties*; ways in which disjoint subsets may be said to be "separated" from one another. Let's first talk about disjoint singleton subsets; i.e., distinct points. The following three properties are the most basic.

Definition 18.1. Let $X = \langle X, \mathcal{T}\rangle$ be a topological space. For each $x \in X$, denote by $\mathcal{T}_x$ the family of open neighborhoods of x.

- X is a T_0 (or *Kolmogorov*) *space* if $\mathcal{T}_x \neq \mathcal{T}_y$ whenever $x \neq y$.
- X is a T_1 (or *Fréchet*) *space* if $\mathcal{T}_x \setminus \mathcal{T}_y \neq \emptyset \neq \mathcal{T}_y \setminus \mathcal{T}_x$ whenever $x \neq y$.
- X is a T_2 (or *Hausdorff*) *space* if $U \cap V = \emptyset$ for some $U \in \mathcal{T}_x$, $V \in \mathcal{T}_y$ whenever $x \neq y$.

Proposition 18.2. (1) *Every T_2 space is a T_1 space, and every T_1 space is a T_0 space.*
(2) *A space is a T_1 space if and only if each finite subset of the space is closed.*

Proof. Ad (1): Let X be a T_2 space, with $x, y \in X$ distinct points. Then there are $U \in \mathcal{T}_x$ and $V \in \mathcal{T}_y$ with $U \cap V = \emptyset$. In particular, $U \notin \mathcal{T}_y$ and $V \notin \mathcal{T}_x$. This says the space is T_1. That T_1 implies T_0 is immediate from the definition.

Ad (2): Assume each finite subset of X is closed, and let $x, y \in X$ be distinct points. Since $\{y\}$ is closed, there is a neighborhood U of x that misses $\{y\}$; i.e., $U \in \mathcal{T}_x \setminus \mathcal{T}_y$. Since $\{x\}$ is closed, there is a neighborhood V of y that misses $\{x\}$; i.e. $V \in \mathcal{T}_y \setminus \mathcal{T}_x$. Thus X is a T_1 space.

Conversely, if $x \in X$ is fixed and $y \neq x$ is any other point, then the T_1 axiom tells us there is an open neighborhood of y that misses x. Hence $\{x\}$ is a closed set. Since finite unions of closed sets are closed, we infer that finite subsets—being finite unions of singletons—of a T_1 space are closed.

□

As the T-notation suggests, there is a hierarchy of separation properties, becoming more restrictive as the subscript increases. This notation is due to Heinrich Tietze. Because our focus is the world of metrizable topologies, we skip T_3—the interested reader is encouraged to look it up—and move right on to the next level of the hierarchy.

Definition 18.3. A topological space X is a T_4 (or *normal*) *space* if it is T_1, and whenever $H, K \subseteq X$ are disjoint closed subsets, there are disjoint open subsets U and V with $H \subseteq U$ and $K \subseteq V$.

So the T_4 axiom is like the T_2 axiom, only with *two distinct points* replaced with *two disjoint closed sets.* Adding the T_1 axiom to the definition ensures that singletons are closed, and hence that there are enough closed sets to make T_4 imply T_2.

In general topology, there are examples of T_0 spaces that are not T_1, T_1 spaces that are not T_2, and so forth. The next result, though, tells that none of these counterexamples can be metrizable.

Proposition 18.4 (Metric Normality Theorem). *All metric spaces are normal.*

Proof. The theorem really says that all *metrizable* topological spaces are normal, but we want to fix a metric d on X before we proceed.

First, if $x, y \in X$ are distinct and $\epsilon = \frac{1}{2}d(x,y)$, then the ball neighborhoods $B_d(x,\epsilon)$ and $B_d(y,\epsilon)$ are disjoint; hence metrizable spaces are easily seen to be Hausdorff, even stronger than T_1.

Now let H and K be disjoint closed subsets of X. For each $a \in H$, since $a \notin K$ and K is closed, there is a ball neighborhood $B(a,\epsilon_a)$ missing K. Similarly, for each $b \in K$ there is a ball neighborhood $B(b,\epsilon_b)$ missing H. Now define $U = \bigcup\{B(a,\frac{1}{2}\epsilon_a) : a \in H\}$ and $V = \bigcup\{B(b,\frac{1}{2}\epsilon_b) : b \in K\}$. Then U and V are open sets, and clearly $H \subseteq U$ and $K \subseteq V$. It remains to show that $U \cap V = \emptyset$. Suppose not, say $x \in U \cap V$. Then there are $a \in H$ and $b \in K$ with $x \in B(a,\frac{1}{2}\epsilon_a) \cap B(b,\frac{1}{2}\epsilon_b)$. This tells us that $d(x,a) < \frac{1}{2}\epsilon_a$ and $d(x,b) < \frac{1}{2}\epsilon_b$. Let $\epsilon = \max\{\epsilon_a,\epsilon_b\}$. Then $d(a,b) \leq d(a,x) + d(x,b) < \frac{1}{2}\epsilon + \frac{1}{2}\epsilon = \epsilon$. Hence either $d(a,b) < \epsilon_a$ or $d(a,b) < \epsilon_b$; i.e., either $b \in B(a,\epsilon_a)$ or $a \in B(b,\epsilon_b)$. But $K \cap B(a,\epsilon_a) = H \cap B(b,\epsilon_b) = \emptyset$, a contradiction.

□

Exercises 18.

(1) Prove that a quotient space X/R is T_1 if and only if every R-equivalence class is closed in X.

(2) Prove that a T_1 space X is normal if and only if, for each closed $H \subseteq X$ and open $U \supseteq H$, there is an open $V \supseteq H$ with $\mathrm{cl}(V) \subseteq U$.

(3) Prove that a T_1 space X is normal if and only if, for each pair H, K of disjoint closed subsets of X, there exist open sets $U \supseteq H$ and $V \supseteq K$ such that $\mathrm{cl}(U) \cap \mathrm{cl}(V) = \emptyset$.
(4) Show that every subspace of a T_n space is T_n, where $n = 0, 1, 2$.
(5) Show that every closed subspace of a normal space is normal, and that every subspace of a metrizable space is normal.

Lecture 19: Introduction to Connectedness

Abstract. A space is connected if it "comes in one piece;" i.e., if it cannot be partitioned into two nonempty closed sets. This is one of the most intuitively compelling of the basic topological properties, but also one of the trickiest to analyze. We show here that connectedness is preserved by continuous images. Also: if a subset of a space is connected, then so is its closure; and if all the members of a collection of connected subsets have a point in common, then the union of that collection is connected. Hence each point of a space is contained in a unique largest connected subset, the component of the point, and each component is closed.

Connectedness is a topological property of fundamental importance; it says of a space that it has "one piece." How this is made precise is a marvel of elegant simplicity.

Disconnections. If X is a metric—or topological—space, a *disconnection* of X is a partition of X into two nonempty closed sets. Thus it is a two-set family $\{H, K\}$, where $H \neq \emptyset \neq K$, H and K are both closed, $H \cup K = X$, and $H \cap K = \emptyset$. H and K are each called a *piece* of the disconnection $\{H, K\}$. X is called *disconnected* if it has a disconnection, and *connected* if it hasn't.

Note that if $\{H, K\}$ is a disconnection of X, then H is open because K, the complement of H, is closed. Likewise K is open as well as closed. Subsets that are both closed and open are—somewhat whimsically—referred to as *clopen*.

Examples 19.1.

(1) Let X be any nonempty subset of $\mathbb{R}$ that is not convex; i.e., there are real numbers $a < b < c$ with $a, c \in X$ and $b \notin X$ (see Lecture 6). Then $\{(-\infty, b) \cap X, (b, \infty) \cap X\}$ is a disconnection of X.

(2) Let $X = \mathbb{Q}$, the subspace of rational real numbers. If $a, b \in \mathbb{Q}$, say $a < b$, let $r \in \mathbb{R}$ be irrational, with $a < r < b$. Then $\{(-\infty, r) \cap \mathbb{Q}, (r, \infty) \cap \mathbb{Q}\}$ is a disconnection of $\mathbb{Q}$ such that a is in one piece and b is in the other.

To show a space to be disconnected, it suffices to produce a disconnection. This may be difficult in some cases, but it is quite a direct and concrete enterprise in general. On the other hand, to show a space to be connected, one is required to demonstrate that no disconnection is possible. Clearly any space with only one point is connected; oftentimes what is done in less trivial circumstances is to construct a proof by contradiction: assume there is a disconnection and try to infer some statement that is known to be false (e.g., $0 = 1$). In the next lecture we will do just this, showing that not only is the real line connected, but so is any subspace that is convex as a subset (i.e., an interval).

General Facts about Connectedness.

Proposition 19.2. *The following statements are equivalent about a space X:*

(a) *X is connected.*
(b) *There are no proper nonempty clopen subsets of X.*
(c) *Every continuous map from X to a discrete space is constant.*

(d) *Every proper nonempty subset if X has nonempty frontier.*

Proof. The normal "round robin" method of proving several statements to be equivalent might go like: (a) $\implies$ (b) $\implies$ (c) $\implies$ (d) $\implies$ (a). This gives us four arguments to devise, rather than the twelve one might expect to need (see Exercise 19.1).

We'll do part of the job of proving Proposition 19.2; you're asked to finish it off.

Ad ((b) $\implies$ (c)). Assume (c) is false, that there is a continuous nonconstant map $f : X \to D$, where D is discrete (say, having the standard discrete $\{0, 1\}$-valued metric). Let $a, b \in f[X]$ be distinct, with $A = f^{-1}(a)$ and $B = f^{-1}(b)$. Then since $\{a\}$ is closed and f is continuous, A is closed too. A is nonempty because $a \in f[X]$; A is proper because $b \in f[X] \setminus f[A]$. But singleton subsets of discrete spaces are open too. Hence A is a proper nonempty clopen subset of X.

Ad ((a) $\implies$ (d)). Suppose $A \subseteq X$ is proper nonempty, and that its frontier $\text{cl}(A) \cap \text{cl}(X \setminus A)$ is empty. Since $A \cup (X \setminus A) = X$ and sets are subsets of their closures, we have $\text{cl}(A) \cup \text{cl}(X \setminus A) = X$. This tells us that $\text{cl}(A)$ and $\text{cl}(X \setminus A)$ are nonempty and complements of one another. Also they're closed; hence they form a disconnection of X.

□

If X is a space and $A \subseteq X$, we say A is a *connected subset* of X if A is connected as a sub*space* of X.

Proposition 19.3. *Continuous images of connected spaces are connected.*

Proof. Let $f : X \to Y$ be a continuous map, and assume X is connected. To show that $f[X]$ is a connected subset of Y, it suffices to assume f is onto and to show that Y is connected. Indeed, suppose Y is disconnected, and that $\{H, K\}$ is a disconnection of Y. Then, because f is continuous, it is easy to check that $\{f^{-1}[H], f^{-1}[K]\}$ is a disconnection of X.

□

The following result tells about how we may infer connectness in some subsets, given we know connectedness in others.

Proposition 19.4. (1) *If a family of connected subsets has nonempty intersection, then the union of that family is connected.*
(2) *If $A \subseteq B \subseteq$ cl(A) and A is connected, then so is B.*

Proof. Ad (1). Suppose $\{A_i : i \in I\}$ is a family of connected subsets of X, and suppose further that there is some point $a \in \bigcap_{i \in I} A_i$. Let $A = \bigcup_{i \in I} A_i$, and suppose—for the sake of obtaining a contradiction—that A is disconnected. This means (see Exercise 19.4) that there are open sets $U, V \subseteq X$ such that $U \cup V \supseteq A$, $A \cap U \neq \emptyset \neq A \cap V$, and $(A \cap U) \cap (A \cap V) = A \cap U \cap V = \emptyset$. Fix $i \in I$. Then $U \cup V \supseteq A_i$ and $A_i \cap U \cap V = \emptyset$; hence, since A_i is connected, we know either

$A_i \subseteq U$ or $A_i \subseteq V$. Let $I_U = \{i \in I : A_i \subseteq U\}$ and $I_V = \{i \in I : A_i \subseteq V\}$. Then $I_U \cup I_V = I$. If I_U, say, were empty, then $A \cap U$ would also be empty, a contradiction. Hence I_U (and likewise I_V) is nonempty. Pick $i, j \in I$ such that $A_i \subseteq U$ and $A_j \subseteq V$. Then $a \in A_i \cap A_j \subseteq A \cap U \cap V = \emptyset$, also a contradiction. Hence no disconnection of A is possible, and A is connected.

Ad (2). Suppose $A \subseteq B \subseteq \text{cl}(A) \subseteq X$, and that B is disconnected. We show that A must also be disconnected. Indeed, as above, there are open sets $U, V \subseteq X$ such that $U \cup V \supseteq B$, $B \cap U \neq \emptyset \neq B \cap V$, and $B \cap U \cap V = \emptyset$. Since $A \subseteq B$, we have that $U \cup V \supseteq A$ and $A \cap U \cap V = \emptyset$; so it remains to show that $A \cap U \neq \emptyset \neq A \cap V$. Suppose $b \in B \cap U$. Then $b \in \text{cl}(A)$, and so—by definition of closure—U, an open neighborhood of b, must intersect A. Hence $A \cap U \neq \emptyset$. Similarly, $A \cap V \neq \emptyset$, and the proof is complete.

□

Proposition 19.4 is crucial for the following important concept. If X is any space and $x \in X$, then the *component* $C(x)$ of x in X is the union of the family of all connected subsets of X that contain x. By Proposition 19.4(1), $C(x)$ is connected; it is the largest (under the subset relation) connected subset of X containing the point x. By Proposition 19.4(2), $C(x)$ is closed; otherwise $\text{cl}(C(x))$ would be a larger connected subset of X containing x.

Exercises 19.

(1) Suppose you're trying to set up a proof of a proposition that states the equivalence of $n \geq 2$ conditions. How many "subproofs" to you need to devise if you use a "round robin" strategy? How many if you take the statements pairwise and prove equivalence for each pair?
(2) Finish the proof of Proposition 19.2.
(3) Suppose $\{A_i : i \in I\}$ is a family of connected subsets of X, and that B is a connected subset of X such that $B \cap A_i \neq \emptyset$ for all $i \in I$. Show that $B \cup \bigcup_{i \in I} A_i$ is connected.
(4) Prove that $A \subseteq X$ is disconnected if and only if there are open sets $U, V \subseteq X$ such that: $U \cup V \supseteq A$; $A \cap U \neq \emptyset \neq A \cap V$; and $(A \cap U) \cap (A \cap V) = A \cap U \cap V = \emptyset$.
(5) Show that the components of a space partition the space.
(6) A space X is called *totally disconnected* if all of its components are singletons. Show that $\mathbb{Q}$ is totally disconnected.
(7) Suppose X and Y are connected spaces, $A \subseteq X$ is nonempty, and $f : A \to Y$ is a mapping. Show that the attaching space $X \cup_f Y$ is connected.

Lecture 20: Connected Subsets of Euclidean Space

Abstract. The connected subsets of the real line are precisely the intervals. We prove that products of connected spaces are connected; hence that higher-dimensional "boxes" are also connected. As applications of connectedness, we prove: (i) the intermediate value theorem from calculus, that a continuous real-valued function defined on an interval takes on all values between two given ones; and (ii) the Brouwer fixed point theorem in dimension one, that any continuous function from a closed bounded interval to itself has to map at least one point to itself.

Connected Subsets of the Real Line. In Example 19.1(1), we showed that any nonempty subset of $\mathbb{R}$ that is not an interval is disconnected. Here we show the converse; that any interval in $\mathbb{R}$ is connected.

Proposition 20.1. *Let $A \subseteq \mathbb{R}$ be a nonempty subset. Then A is connected if and only if A is an interval.*

Proof. We already showed non-intervals are disconnected; it remains to show that intervals are connected.

We start with the closed unit interval $[0, 1]$. Assume—in the interests of proving you're the Pope—there is a disconnection $\{H, K\}$ of $[0, 1]$; without loss of generality, assume $0 \in H$. Since H is a nonempty subset of $\mathbb{R}$ that is bounded above, we infer (see Lecture 6) that H has a least upper bound c. And since 1 is an upper bound for H, we know $c \in [0, 1]$. Now, if $c = 0$, then either $K = (0, 1]$ or $K = [0, 1]$. The first assertion can't hold because K is closed; the second can't hold because H is nonempty.

So $0 < c \leq 1$, and we may pick $0 < r < c$. Then r is not an upper bound for H because c is the least such. Hence there is some $h \in H$ such that $r < h$. Since c is an upper bound for H, we know $h \in (r, c]$. This tells us that every neighborhood of c intersects H, that therefore c is in the closure of H. Since H is already closed, we infer that $c \in H$.

Suppose first that $c \neq 1$. Then, because K is closed and $c \notin K$, we know there exist $0 < r < c < s < 1$ with $(r, s) \cap K = \emptyset$; i.e., with $(r, s) \subseteq H$. But then we have members of H lying to the right of c, contradicting the fact that c is an upper bound for H.

So we conclude that c, the least upper bound for H, must be 1, hence $1 \in H$. With that in mind, we shift our attention to K. By the same argument as above, we know that the least upper bound d of K is a member of K that can't be in H; in particular, $0 < d < 1$. But now we argue as in the last paragraph to draw another contradiction; hence there can be no disconnection of $[0, 1]$.

We now use the fact that all closed bounded intervals $[a, b] \subseteq \mathbb{R}$ are homeomorphic. Since connectedness is a topological property, we know that such intervals are connected. Finally we use Proposition 19.4(1), plus the fact that any interval in $\mathbb{R}$ can be represented as an increasing union of closed bounded intervals (see Exercise 20.1) to conclude that all intervals in the real line are connected.

□

Connectedness in Higher Dimensions. It should be intuitively clear that euclidean space $\mathbb{R}^n$ is connected for $n \geq 1$. We have the $n = 1$ case, so we need a result that gives us connectedness in products.

Proposition 20.2. *Let $X_1, \ldots, X_n$ be spaces. Then $X_1 \times \cdots \times X_n$ is connected if and only if each X_i is connected.*

Proof. Let the product be Π, and let $p_i : \Pi \to X_i$ be projection onto the ith coordinate. Since each p_i is a continuous surjection, we infer connectedness in X_i from connectedness in Π by Proposition 19.3.

For the converse, assume $n = 2$ for simplicity. Once we show the proposition true for this case, we may use induction to obtain the proposition in general.

So assume X and Y are connected but that $X \times Y$ is not; let $\{H, K\}$ be a disconnection. We pick $\langle a, b\rangle \in H$ and $\langle c, d\rangle \in K$. Let $A = X \times \{b\}$ and $B = \{c\} \times Y$. Then it is easy to show that A is homeomorphic to X and that B is homeomorphic to Y—just use the restricted projection maps. Thus both A and B are connected subsets of $X \times Y$. And since $\langle c, b\rangle \in A \cap B$, we conclude that $C = A \cup B$ also is connected, by Proposition 19.4(1). Since $\langle a, b\rangle \in A$ and $\langle c, d\rangle \in B$, we conclude that $\{H \cap C, K \cap C\}$ is a disconnection of C, a contradiction. Thus $X \times Y$ is connected whenever both X and Y are.

□

The following is now an immediate corollary.

Proposition 20.3. *Given a finite number n of subintervals of $\mathbb{R}$, their cartesian product, a multidimensional box, is connected. This includes $\mathbb{R}^n$ itself.*

Although it is now easy to identify the connected subsets of $\mathbb{R}$, the corresponding problem for $\mathbb{R}^n$, $n \geq 2$, is immeasureably more complicated. We will present several "exotic" examples in later lectures; for now, let's take up the issue of when $\mathbb{R}^n$ and $\mathbb{R}^m$ are homeomorphic. This may seem like a trivial question, it should not be the case that they're homeomorphic if $m \neq n$, but this is far from trivial to prove. In fact we will only have the wherewithall in this introductory course to show that no $\mathbb{R}^n$, $n \geq 2$, is homeomorphic to $\mathbb{R}$. For this we introduce a new and important notion: define $c \in X$ to be a *cut point* if X is connected and $X \backslash \{x\}$ is not.

Proposition 20.4. *If $n \geq 2$, $\mathbb{R}^n$ is not homeomorphic to $\mathbb{R}$.*

Proof. We first state (see Exercise 20(2)) the easy-to-prove fact that a homeomorphism takes cut points to cut points and noncut points to noncut points. That said, we note that every point of $\mathbb{R}$ is a cut point; so all we need to show that for $n \geq 2$, $\mathbb{R}^n$ has no cut points at all. But this too is easy (see Exercise 20(3)).

□

Some Applications. Here is a theorem you've seen in calculus. It's always stated without proof—and for good reason.

Proposition 20.5 (Intermediate Value Theorem). *Suppose* $f : [a, b] \to \mathbb{R}$ *is continuous and* y *is a point lying between* $f(a)$ *and* $f(b)$. *Then there is some* $c \in [a, b]$ *with* $y = f(c)$.

Proof. $f[a, b]$ is an interval containing $f(a)$ and $f(b)$, by Propositions 19.3 and 20.1. □

A *self mapping* is a continuous function whose domain and range are equal. A *fixed point* of a self mapping is a point of the domain that the function takes to itself. For example, the self mapping on $\mathbb{I}$ given by $x \mapsto x^2$ has only 0 and 1 as fixed points.

There are many theorems in mathematics that go under the heading of "fixed point theorem," the most famous of these in topology is due to the Dutch mathematician/logician L. E. J. Brouwer.

Proposition 20.6 (Brouwer's Fixed Point Theorem). *Every self mapping on the* n*-cube* $\mathbb{I}^n$, $n = 1, 2, \ldots,$ *has a fixed point.*

Proof. While this theorem is true for all finite dimensions n, the only case within the scope of this course is when $n = 1$. So suppose $f : \mathbb{I} \to \mathbb{I}$ is continuous, and define $g : \mathbb{I} \to \mathbb{R}$ by $g(x) = x - f(x)$. Then clearly g is continuous because f is. Assuming 0 and 1 are not fixed points of f—if either is, we're done—we see that $g(0) < 0 < g(1)$. Hence, by the Intermediate Value Theorem, there is some $c \in \mathbb{I}$ with $g(c) = 0$; i.e., with $f(c) = c$. □

Spaces for which every self mapping has a fixed point are said to satisfy the *fixed point property*.

Exercises 20.
(1) Prove that any interval A in $\mathbb{R}$ may be written as $A_1 \cup A_2 \cup \ldots$, where $A_1 \subseteq A_2 \subseteq \ldots$ is an increasing chain of closed bounded intervals.
(2) Show that if $c \in X$ is a cut point of X and $h : X \to Y$ is a homeomorphism, then $h(c)$ is a cut point of Y.
(3) Prove that every point of $\mathbb{R}^n$, $n \geq 2$ is a noncut point.
(4) Give an example of a connected subset of $\mathbb{R}^2$ that is not of the form $A \times B$, where A and B are intervals.
(5) Give a complete proof of the fact that if $f(x)$ is a polynomial and $a, b \in \mathbb{R}$ are such that $f(a) < 0 < f(b)$, then there is a root of $f(x)$ lying in (a, b).
(6) Does every self mapping on $\mathbb{R}$ have a fixed point? How about the unit circle $\mathbb{S}^1$?
(7) Show that satisfying the fixed point property is a topological invariant.

Lecture 21: Path Connectedness

Abstract. A space is path connected just in case you can get from any one point to any other via a path; i.e., each two of its points lie within the image of a continuous mapping whose domain is a closed bounded interval in the real line. Because such intervals are connected, and connectedness is preserved by continuous images, path connectedness is a stronger property than is connectedness. We construct an example of a subspace of the plane, the "topologist's sine curve," which is connected but not path connected. We also show that any connected open subset of euclidean space is path connected.

Many results about connectedness have parallels for path connectedness. For example, each point of a space is contained in a unique largest path connected subset, the path component of the point. Path components needn't be closed sets, however.

Paths. Perhaps the most intuitive notion of connectedness in a space is to say that you can "go from any one point to any other via a continuous path." Of course all these words in quotes need to be explained.

First define a *path* in space X to be a continuous function $\alpha : \mathbb{I} \to X$. If $a, b \in X$ are two points, we say α is a path *from a to b* if $\alpha(0) = a$ and $\alpha(1) = b$. Clearly, if α is a path from a to b, then the *reverse path* $\overline{\alpha}$, defined by $\overline{\alpha}(t) = \alpha(1 - t)$, is a path from b to a. So it makes sense to say that points a and b can be *joined* by a path if there is some path from one of them to the other. We define a space to be *path connected* if any two of its points may be joined by a path.

Proposition 21.1. *Path connectedness implies connectedness.*

Proof. Suppose X is path connected but not connected, and let $\{H, K\}$ be a disconnection of X. Pick $a \in H$ and $b \in K$. Then there is a path $\alpha : \mathbb{I} \to X$ with $\alpha(0) = a$ and $\alpha(1) = b$. But then $C = \alpha[\mathbb{I}]$, the image of α, is a connected subset of X that intersects both pieces of a disconnection of X, and hence itself is disconnected. Contradiction.

□

If u and v are points in $\mathbb{R}^n$, we may treat the points as vectors in a vector space and define paths using linear algebra. For example, $\alpha : \mathbb{I} \to \mathbb{R}^n$, defined by $\alpha(t) = (1 - t)u + tv$, is a straight line path from u to v, its image is a line segment with u and v as end points. This tells us the unsurprising fact that euclidean n-space is path connected.

There is a way of "appending" one path to another which is of basic importance in the study of path connectedness. The key ingredient is the following lemma, justifying how, in calculus, we often define continuous functions by cases.

Proposition 21.2 (Gluing Lemma). *Suppose $X = A \cup B$, where A and B are closed subsets of X, and let $f : A \to Y$ and $g : B \to Y$ be continuous. If $f|(A \cap B) = g|(A \cap B)$ (agreement on the overlap), then there is a unique continuous $h : X \to Y$ such that $h|A = f$ and $h|B = g$.*

Proof. We call h a *joint extension* of f and g; we may think of h as being defined via cases:

$$h(x) = \begin{cases} f(x) & \text{if } x \in A \\ g(x) & \text{if } x \in B. \end{cases}$$

This is a good function definition because there is no ambiguity when $x \in A \cap B$, as $f(x) = g(x)$. The uniqueness of h is obvious; all that's left is to show continuity. Suppose $C \subseteq Y$ is closed; we show $h^{-1}[C]$ is closed in X. But $h^{-1}[C] = f^{-1}[C] \cup g^{-1}[C]$, by the fact that, for any $x \in X$, $h(x) \in C$ if and only if $x \in A$ and $f(x) \in C$ or $x \in B$ and $g(x) \in C$. Since f is continuous, $f^{-1}[C]$ is closed in A. Since A is closed in X, $f^{-1}[C]$ is closed in X. Likewise $g^{-1}[C]$ is closed in X. Thus $h^{-1}[C]$, a finite union of closed sets, is closed in X.

□

So now the following definition makes sense. Let α and β be two paths in X, with $\beta(0) = \alpha(1)$—i.e., β begins where α ends. Then the *concatenation* $\alpha * \beta$ is defined by

$$(\alpha * \beta)(t) = \begin{cases} \alpha(2t) & \text{if } t \in [0, \frac{1}{2}] \\ \beta(2t-1) & \text{if } t \in [\frac{1}{2}, 1]. \end{cases}$$

Thus if α is a path from a to b and β is a path from b to c, then $\alpha * \beta$ is a path from a to c. This is the starting point of what is called *homotopy theory*, one of the main areas of algebraic topology. It will also be of use below.

It is a natural question to ask whether connected spaces are path connected. The short answer is "not in general," and one need look no further than the euclidean plane for connected sets that are not path connected.

The Topologist's Sine Curve. This is an important example, which we will have occasion to revisit a number of times in the sequel.

We start by letting $A = \{0\} \times [-1, 1] \subseteq \mathbb{R}^2$ be the points on the y-axis with second coordinate between -1 and 1. Next, we let $B = \{\langle t, \sin\frac{1}{t}\rangle : 0 < t \leq 1\}$ be the graph of $\sin\frac{1}{t}$ for $t \in (0, 1]$. Then the space called the *topologist's sine curve* is $T = A \cup B$, taken as a subspace of $\mathbb{R}^2$. Here are some simple—but important—observations about T:

- A is an *arc*; i.e., a homeomorphic copy of the closed unit interval $\mathbb{I} = [0, 1]$.
- The projection map $p_1 : \mathbb{R}^2 \to \mathbb{R}$ onto the first coordinate, when restricted to B, is a homeomorphism onto $(0, 1]$. Thus both A and B are path connected.
- A is closed in $\mathbb{R}^2$, but B is not. Indeed, every point of A is an accumulation point of B; and, in fact, $T = \text{cl}(B)$. Since B is connected, so is T.
- Let $C \subseteq T$ be connected, and suppose C intersects both A and B. Let $a \in A \cap C$ and $b \in B \cap C$, where $b = \langle b_1, b_2\rangle$ (so $0 < b_1 \leq 1$ and $b_2 = \sin\frac{1}{b_1}$). Suppose that there is some $d = \langle d_1, d_2\rangle \in B \setminus C$ with $d_1 < b_1$. Then we have the disjoint open (in the plane) sets $U = \{\langle x, y\rangle \in \mathbb{R} : x < d_1\}$ and

$V = \{\langle x, y\rangle \in \mathbb{R} : x > d_1\}$; moreover $a \in U \cap C$ and $b \in V \cap C$. This gives a disconnection of C, a contradiction. So what we've just shown is that any connected subset of T intersecting both A and B must contain all points of B to the left of some positive x-coordinate, and must therefore contain points that are euclidean distance at least 2 apart. In particular, the diameter of C is ≥ 2 (see Exercise 7(5)).

- Suppose $a \in A$, $b \in B$, and α a path in T joining a to b. Let $W = \{t \in \mathbb{I} : \alpha[[0, t]] \subseteq A\}$. Then $0 \in W$ and 1 is an upper bound for W, so let w be the least upper bound of W. Then, as in the proof that the closed unit interval is connected, we know that $w \in \mathrm{cl}(W)$. Since α is continuous, this says that $\alpha(w) \in \alpha[\mathrm{cl}(W)] \subseteq \mathrm{cl}(\alpha[W]) \subseteq \mathrm{cl}(A) = A$. Now, α is continuous at w; hence there is some $\delta > 0$ such that whenever $t \in \mathbb{I}$ and $|t - w| < \delta$, then the euclidean distance between $\alpha(t)$ and $\alpha(w)$ is less than $1/2$. But we know $w < 1$ because $\alpha(1) \in B$. Hence we can find some $v \in \mathbb{I}$ with $w < v < w + \delta$. Then $\alpha[[w, v]]$ has diameter at most 1. On the other hand, since everything bigger than w is sent by α to B, we have $\alpha(v) \in B$. Thus, by the last bullet point made above, $\alpha[w, v]$ is a connected set intersecting both A and B; hence the diameter of $\alpha[v, w]$ is at least 2. This contradiction tells us that T is not path connected.

When Connected Implies Path Connected. The next result says that, under special circumstances, connected does imply path connected.

Proposition 21.3. *Every connected open subset of $\mathbb{R}^n$ is path connected.*

Proof. Given a nonempty connected open set $U \subseteq \mathbb{R}^n$ and $a \in U$, let $U_a = \{x \in U :$ there is a path in U from a to $x\}$. Then $a \in U_a$, so U_a is nonempty. And because we can concatenate paths, we know that U_a is also path connected. Thus it suffices to show $U_a = U$.

If $x \in U_a$, let $\epsilon > 0$ be chosen so that $B(x, \epsilon) \subseteq U$. Let α be a path in U joining a to x, and let $y \in B(x, \epsilon)$. Then $B(x, \epsilon)$ is convex (see Exercise 21(2)), so there is a straight-line path $\beta : \mathbb{I} \to B(x, \epsilon)$ joining x to y. Thus $\alpha * \beta$ is a path in U joining a to y. This argument tells us that U_a is open in $\mathbb{R}^n$, and hence open in U.

Now suppose $x \in U \setminus U_a$. Again we have some $\epsilon > 0$ so that $B(x, \epsilon) \subseteq U$. $B(x, \epsilon)$ cannot intersect U_a, though; otherwise we argue as above to obtain a path joining a to x and a contradiction. Thus $U \setminus U_a$ is also open in $\mathbb{R}^n$, hence open in U. But then U_a, the complement of $U \setminus U_a$ in U, is closed in U. Since $U_a \neq \emptyset$ and U is connected, we conclude $U_a = U$, by Proposition 19.2.

□

Path Components. Recall from Lecture 19 that the *component* $C(a)$ of a point a in space X is the largest–under set containment–connected subset of X containing a. $C(a)$ always contains a because singletons are connected; no two distinct components can overlap; and components are always closed subsets of X.

By analogy, we define the *path component* $C_p(a)$ of point $a \in X$ is the largest path connected subset of X containing a. The analogue of Proposition 19.4(1)—that if a family of path connected subsets of X has nonempty intersection, then the union of that family is path connected—still holds, so the definition of path component makes sense. Each point is contained in its own path component, no two path components intersect, and each component is a union of path components. Where the analogy breaks down is in the fact that path components needn't be closed. You're asked to prove these statements in the exercises.

Exercises 21. (1) Prove that every interval in $\mathbb{R}$ is path connected.

(2) A set $C \subseteq \mathbb{R}^n$ is *convex* if, for each two points of C, the line segment joining them is also in C. Give some examples of convex and nonconvex subsets of $\mathbb{R}^2$.

(3) Show that no two distinct path components can overlap.

(4) Show that each component is a union of path components.

(5) Exhibit an example of a path component of a space that isn't closed in the space.

(6) Prove that continuous images of path connected spaces are path connected.

(7) Prove that, in $\mathbb{R}^n$, path components of open sets are open.

(8) Prove that the union of a family of path connected subsets of a space is path connected, as long as there is a point in common with all the subsets.

(9) Prove the following version of Proposition 20.2: A finite product of spaces is path connected if and only if each factor space is path connected.

(10) Prove that if X is any topological space, the cone $\mathsf{C}(X)$ and the suspension $\mathsf{S}(X)$ are path connected.

Lecture 22: Local Connectedness

Abstract. A space is locally (path) connected if each point has arbitrarily small open neighborhoods that are (path) connected. We show: (i) a space is locally (path) connected if and only if each (path) component of each open set is open; (ii) connected, locally path connected spaces are path connected; and (iii) local (path) connectedness is preserved by identification maps (but not by continuous maps in general).

If P is a property of topological spaces—such as connectedness or path connectedness—then we can "localize" P by saying a space X *locally satisfies* P if each point of X has arbitrarily small open neighborhoods that, as subspaces, satisfy P. Thus a space X is *locally (path) connected* if for each $x \in X$ and open U containing x, there is a (path) connected open set V with $x \in V \subseteq U$. By Proposition 21.1, local path connectedness implies local connectedness.

Examples 22.1. (1) Every interval in $\mathbb{R}$ is locally path connected.
(2) Every open subset of $\mathbb{R}^n$ is locally path connected.
(3) The union of two disjoint closed intervals in $\mathbb{R}$ is locally path connected but not connected.
(4) The topologist's sine curve T (see Lecture 21) is connected, but not locally connected: any open neighborhood of $\langle 0, 0\rangle$, say, if of diameter less than 1, consists of infinitely many pairwise disjoint "open arcs;" i.e., homeomorphic copies of $\mathbb{R}$.

Proposition 22.2. *X is locally (path) connected if and only if each (path) component of each open set is open.*

Proof. We prove the version dealing with local connectedness, leaving the other version as an exercise.

Assume X is locally connected, with $U \subseteq X$ open. Let $C \subseteq U$ be a component of U. For each $x \in C$, there is a connected open V, with $x \in V \subseteq U$. Since C is the largest connected subset of U containing x, we have $V \subseteq C$. Thus C is a neighborhood of each of its points, hence open in X.

For the converse, suppose $x \in X$, with U an open neighborhood of x. Then the component C of x in U is open in X; hence x has arbitrarily small connected open neighborhoods. This tells us X is locally connected.

□

Proposition 22.3 . *Every connected locally path connected space is path connected.*

Proof. Assume X is both connected and locally path connected; for each $a \in X$, let $C_p(a)$ be the path component of a in X. Since X is open in itself, Proposition 22.2 tells us that $C_p(a)$ is open in X. If $C_p(a) = X$, we're done; otherwise if $x \in X \setminus C_p(a)$, then (see Exercise 21(3)) $C_p(x) \cap C_p(a) = \emptyset$. But then $\{C_p(a), \bigcup\{C_p(x) : x \notin C_p(a)\}\}$ forms a disconnection of X. Since this can't happen we conclude that $C_p(a) = X$ after all.

□

Since continuous images of (path) connected spaces are (path) connected—i.e., both connectedness and path connectedness are continuous invariants—it is natural to ask whether the same holds for their localizations. The general answer is in the negative: it is possible—indeed easy— to map the half-open interval $[0, 1)$, a locally path connected space, onto the *Warsaw circle*, the result of taking the topologist's sine curve T and joining the points $\langle 0, -1\rangle$ and $\langle 1, \sin 1\rangle$ with an arc that doesn't otherwise intersect T. This mapping can be made to be a continuous bijection (see also Exercise 13(9)), so local (path) connectedness is not even preserved by continuous bijections. But if the mapping is an identification, we have a positive result.

Proposition 22.4 . *Let $f : X \to Y$ be an identification mapping (see Lecture 16). If X is locally (path) connected, then so is Y.*

Proof. We prove the version dealing with local connectedness, leaving the other version for the exercises.

Let $V \subseteq Y$ be open, with C a component of V. It suffices, in light of Proposition 22.2, to show that C is open in Y; and, for this—since f is an identification—it suffices to show that $f^{-1}[C]$ is open in X. So let $x \in f^{-1}[C]$ and let K be the component of x in $f^{-1}[V]$. Then, since $f^{-1}[V]$ is open in X and X is locally connected, we know that K is open in X. We're done if we can show $K \subseteq f^{-1}[C]$. But $f[K]$ is connected by Proposition 19.3, is a subset of V, and intersects C ($f(x) \in f[K] \cap C$). Since C is maximally connected in V, we know that $f[K] \subseteq C$; hence $K \subseteq f^{-1}[C]$.

□

Exercises 22. (1) According to the recipe given in the first paragraph above, what does *locally Hausdorff* mean? How does it relate to the property of being Hausdorff?

(2) Prove the version of Proposition 22.2 that addresses local path connectedness.
(3) Recall that a space is totally disconnected (see Exercise 19(6)) if each of its components is a singleton. Show that a space is totally disconnected and locally connected if and only if it is discrete.
(4) Provide the details to the claim that there is a continuous bijection from the half-open interval to the Warsaw circle.
(5) Prove the version of Proposition 22.4 that addresses local path connectedness.
(6) X is connected *im kleinen* at $x \in X$ if for each open neighborhood U of x there is an open V with $x \in V \subseteq U$ such that for each $y \in V$ there is a connected set C with $x, y \in C$ and $C \subseteq U$. (This is a mouthful!) Prove that X is locally connected if and only if X is connected im kleinen at each of its points.
(7) Prove the following versions of Proposition 20.2: A finite product of spaces is locally (path) connected if and only if each factor space is locally (path) connected.

Lecture 23: Introduction to Compactness

Abstract. A space is compact just in case whenever it is represented as the union of a family of open subsets, it may also be represented as the union of a finite subfamily of that family. In addition to proving some elementary general results about compacness—e.g., that it is preserved by continuous images—we show the Heine-Borel theorem, in dimension one, that characterizes the compact subsets of the real line as those which are both closed and bounded.

Closed Bounded Subsets of $\mathbb{R}$. The closed unit interval $\mathbb{I} = [0, 1]$ is a closed bounded subset of the real line with the usual metric. The intuition that being closed and bounded in euclidean n-space is somehow a topological property led to what we now know as the notion of compactness.

Proposition 23.1 (Nested Intervals Theorem). *Suppose* $\{[a_n, b_n] : n = 1, 2, \dots\}$ *is an indexed collection of nonempty closed bounded intervals in* $\mathbb{R}$*, which is* nested*, in the sense that* $[a_1, b_1] \supseteq [a_2, b_2] \supseteq \dots$ *. Then* $\bigcap\{[a_n, b_n] : n = 1, 2, \dots\} \neq \emptyset$*. If, in addition, the diameters of the sets* $[a_n, b_n]$ *converge to zero, the intersection consists of a single point.*

Proof. (This is similar to the Cantor Intersection Theorem (Exercise 14(2)), but does not explicitly involve Cauchy sequences.) Nestedness tells us that $a_1 \leq a_2 \leq \dots$, $b_1 \geq b_2 \geq \dots$, and $a_n \leq b_m$ for all $m, n \geq 1$. So the set $A = \{a_n : n = 1, 2, \dots\}$ has a least upper bound a and the set $B = \{b_n : n = 1, 2, \dots\}$ has a greatest lower bound b. If b were less than a, then b would not be an upper bound for A; hence we would have $b < a_n$ for some n. But a_n is a lower bound for B, greater than the greatest lower bound for B. This contradiction tells us that $a \leq b$; hence that $\emptyset \neq [a, b] \subseteq [a_n, b_n]$ for all $n \geq 1$. This shows the intersection of the nested collection to be nonempty. It is not difficult to show (Exercise 23(1)) that $[a, b]$ is *equal* to the intersection, not just a subset of it; hence if the diameters converge to zero; i.e., if $\lim_{n\to\infty}(b_n - a_n) = 0$, then $a = b$ and the intersection is precisely $\{a\}$. □

Let X be a space, with $A \subseteq X$. A family $\mathcal{B}$ of subsets of X is a *cover* of A if $A \subseteq \bigcup \mathcal{B}$. Set-theoretic adjectives, like *finite*, *countable*, etc., apply to the family $\mathcal{B}$ as a set; e.g., a finite cover has finitely many sets in it. On the other hand, adjectives that apply naturally to subsets of a topological space, like *open*, *closed*, etc., apply to the members of $\mathcal{B}$; e.g., an open cover is a cover whose members are open sets. While there is a nominal ambiguity in this terminology, it should be clear from context what is meant.

If $\mathcal{B}$ is a cover of A and $\mathcal{C} \subseteq \mathcal{B}$ also covers A, then $\mathcal{C}$ is commonly referred to as a *subcover* of $\mathcal{B}$. This is again confusing terminology, and properly means that $\mathcal{C}$ is a subfamily of $\mathcal{B}$ covering A. In the sequel we will use the word *subcover* only in the context where $A = X$.

Proposition 23.2 (Heine-Borel Theorem I). *For every open cover of a closed bounded interval* $[a, b] \subseteq \mathbb{R}$*, there is a finite subfamily also covering* $[a, b]$*.*

Proof. We begin with some terminology: First if A is a subset of a space X and $\mathcal{B}$ is a family of subsets of X, then A is *finitely coverable by* $\mathcal{B}$ if some finite subfamily of $\mathcal{B}$ covers A. Next, if $[c, d]$ is a closed bounded interval and $m = \frac{1}{2}(c + d)$ is the midpoint of c and d, then the subintervals $[c, m]$ and $[m, d]$ are called the *left half* and the *right half* of $[c, d]$, respectively.

Now suppose $\mathcal{U}$ is an open cover of $[a, b] \subseteq \mathbb{R}$, but that $[a, b]$ is not finitely coverable by $\mathcal{U}$. We work toward a contradiction.

If both left and right halves of $[a, b]$ were finitely coverable by $\mathcal{U}$, then the union of the two implied finite subfamilies would be a finite subfamily of $\mathcal{U}$ covering $[a, b]$. Hence at least one half of $[a, b]$ is not finitely coverable by $\mathcal{U}$; let it be $[a_1, b_1]$. We repeat the argument above for $[a_1, b_1]$, obtaining $[a_2, b_2]$, which is either the left half or the right half of $[a_1, b_1]$ whichever is first chosen as not being finitely coverable by $\mathcal{U}$. Then $[a, b] \supseteq [a_1, b_1] \supseteq [a_2, b_2] \supseteq \dots$ is a nested sequence of closed bounded intervals. Moreover, since the diameter of $[a_{n+1}, b_{n+1}]$ is exactly half of the diameter of $[a_n, b_n]$, we know the diameters converge to zero. Hence, by Proposition 23.1, the intersection of the nested sequence is a singleton set $\{c\}$. Since $c \in [a, b]$, there is some $U \in \mathcal{U}$ with $c \in U$. And since U is an open neighborhood of c, there is some $\epsilon > 0$ with $(c-\epsilon, c+\epsilon) \subseteq U$. But c is the least upper bound of the a_n, hence there is some $a_k \in (c-\epsilon, c+\epsilon)$. Similarly, since c is the greatest lower bound of the b_n, there is some $b_m \in (c - \epsilon, c + \epsilon)$. Assuming $m \geq k$, we have $[a_m, b_m] \subseteq (c - \epsilon, c + \epsilon) \subseteq U$. But then $\{U\}$ is a finite subfamily of $\mathcal{U}$ covering $[a_m, b_m]$, a contradiction.

□

Compact Spaces. The open cover property of closed bounded intervals in Proposition 23.2 easily translates into a topological condition, independent of metric notions. Define a space X to be *compact* if every open cover of X has a finite subcover; i.e., a finite subfamily also covering X. A subset A of X is *compact* if it is compact as a subspace, and this is easily seen (see Exercise 23(3)) to be equivalent to saying that every open cover of A has a finite subfamily covering A. Compactness is a topological property; indeed—like (path) connectedness—it is a continuous invariant.

Proposition 23.3. *Continuous images of compact spaces are compact.*

Proof. Suppose $f : X \to Y$ is a continuous surjection, and X is compact. Let $\mathcal{V}$ be an open cover of Y. Then $\mathcal{U} = \{f^{-1}[V] : V \in \mathcal{V}\}$ is an open cover of X. By compactness in X, there is a finite subfamily $\{V_1, \dots, V_n\} \subseteq \mathcal{V}$ such that $\{f^{-1}[V_1], \dots, f^{-1}[V_n]\}$ covers X. Hence $\{V_1, \dots, V_n\}$ covers Y.

□

Proposition 23.4. *Closed subsets of compact spaces are compact; compact subsets of Hausdorff spaces are closed.*

Proof. Suppose X is compact, with A a closed subset of X. Let $\mathcal{U}$ be an open cover of A. Then $\mathcal{U} \cup \{X \setminus A\}$ is an open cover of X; so there is a finite subcover $\mathcal{V}$. Hence $\mathcal{V} \setminus \{X \setminus A\}$ is a finite subfamily of $\mathcal{U}$ covering A. Thus A is compact.

Now suppose X is Hausdorff (X could be a metrizable space, but that's much too strong an assumption) and that $A \subseteq X$ is compact. Fix $x \in X \setminus A$; we find an open neighborhood V of x missing A. Indeed, for each $a \in A$, use the Hausdorff property of X to obtain disjoint open sets U_a and V_a with $a \in U_a$ and $x \in V_a$. Then $\{U_a : a \in A\}$ is an open cover of A; by compactness of A, there is a finite set $\{a_1, \dots, a_n\} \subseteq A$ such that $\{U_{a_1}, \dots, U_{a_n}\}$ covers A. Let $V = V_{a_1} \cap \dots \cap V_{a_n}$. Then V is an open neighborhood of x disjoint from $U_{a_1} \cup \dots \cup V_{a_n}$, hence disjoint from A.

□

There is a closed set criterion for compactness that essentially uses the DeMorgan laws, and is left as an exercise. First define a family of subsets of a set to satisfy the *finite intersection property* if each finite subfamily has nonempty intersection.

Proposition 23.5. *The following two conditions are equivalent for a space X.*

(a) *X is compact.*
(b) *If $\mathcal{F}$ is a family of closed subsets of X and $\mathcal{F}$ satisfies the finite intersection property, then $\bigcap \mathcal{F} \neq \emptyset$.*

Proposition 23.2 tells us that closed bounded intervals in $\mathbb{R}$ are compact sets; the following strengthens that assertion with very little extra effort.

Proposition 23.6 (Heine-Borel Theorem II). *A subset of the real line (usual metric) is compact if and only if it is closed and bounded.*

Proof. Suppose $A \subseteq \mathbb{R}$ is closed and bounded. Then A is a closed subset of some closed bounded interval $[a, b]$. Hence A is compact, by Propositions 23.2 and 23.4.

Conversely, suppose A is compact. Let $\mathcal{U}$ consist of the open intervals $(-n, n)$, $n = 1, 2, \dots$. By the archimedean property of the real line every real number lies in some $(-n, n)$; hence $\mathcal{U}$ is an open cover of A because it is an open cover of the larger set $\mathbb{R}$. By compactness, there is a finite subfamily of $\mathcal{U}$ covering A. Since the sets in $\mathcal{U}$ get bigger as n increases, we know that $A \subseteq (-n, n)$ for sufficiantly large $n \geq 1$. Thus A is bounded. Since metric spaces are Hausdorff, we know A is closed, by Proposition 23.4.

□

Exercises 23. (1) Finish the proof of Proposition 23.1.
(2) Give an example of a nested family of nonempty open unbounded intervals with empty intersection.
(3) Prove that $A \subseteq X$ is compact in its subspace topology if and only if every cover of A by open subsets of X has a finite subfamily that also covers A.
(4) Prove that if X is compact, Y is Hausdorff, and $f : X \to Y$ is continuous, then f is a closed mapping.
(5) Prove Proposition 23.5.
(6) Prove that every finite space is compact.
(7) Prove the following covering characterization of connectedness: A space X is connected if and only if, for every open cover $\mathcal{U}$ and every pair $a, b \in X$ of points, there exists a finite sequence $\langle U_1, \dots, U_n \rangle$ of members of $\mathcal{U}$ such that $a \in U_1$, $b \in U_n$, and, for $1 \leq i \leq n - 1$, $U_n \cap U_{n+1} \neq \emptyset$.

Lecture 24: More on Compactness

Abstract. We prove that the product of a finite number of compact spaces is compact. Armed with this, we then show the full Heine-Borel theorem, which characterizes the compact subsets of higher-dimensional euclidean space as those that are both closed and bounded. Finally, we prove that compact Hausdorff spaces are normal.

A major objective of this lecture is to extend Proposition 23.6 (Heine-Borel II) to euclidean n-space, and for this we need a theorem about preservation of compactness in products.

The Product Theorem. Recall that if X and Y are spaces, with topologies $\mathcal{T}_X$ and $\mathcal{T}_Y$ respectively, then the product topology on $X \times Y$ consists of all unions of "open boxes" $U \times V$, where $U \in \mathcal{T}_X$ and $V \in \mathcal{T}_Y$. The collection of such open boxes forms what is called an *open base* for the product topology. As another familiar example, consider a metric space $\langle X, d\rangle$, and let our open base consist of all d-balls $B_d(x, \epsilon)$, for $x \in X$ and $\epsilon > 0$. In general, if $\mathcal{T}$ is a topology on X, a subfamily $\mathcal{B} \subseteq \mathcal{T}$ is an *open base* for $\mathcal{T}$ if every set in $\mathcal{T}$ is a union of sets in $\mathcal{B}$. Alternatively, if for each $U \in \mathcal{T}$ and each $x \in U$, there is a $B \in \mathcal{B}$ such that $x \in B \subseteq U$.

If X is a space and $\mathcal{B}$ is an open base for the topology on X, we say X is $\mathcal{B}$-*compact* if every open cover of X by members of $\mathcal{B}$ has a finite subcover. This is ostensibly a weaker form of compactness than full compactness, as we are allowing only a restricted variety of open covers to begin with. In fact, it is no weaker.

Proposition 24.1 . *Let X be a space, with $\mathcal{B}$ an open base for X. The following are equivalent.*

(a) *X is $\mathcal{B}$-compact.*
(b) *X is compact.*

Proof. We need only prove the $(a \implies b)$-direction, so suppose $\mathcal{U}$ is an open cover of X. Since $\mathcal{B}$ is an open base, we have, for each $U \in \mathcal{U}$ a family $\mathcal{B}_U \subseteq \mathcal{B}$ such that $U = \bigcup \mathcal{B}_U$. Thus $\mathcal{V} = \bigcup\{\mathcal{B}_U : U \in \mathcal{U}\}$ is an open cover of X by members of $\mathcal{B}$. And since X is $\mathcal{B}$-compact, there is a finite subcover. That is, there are $U_1, \ldots, U_n \in \mathcal{U}$ and sets $B_1, \ldots, B_n \in \mathcal{B}$ such that $B_i \in \mathcal{B}_{U_i}$, $1 \leq i \leq n$, and $B_1 \cup \cdots \cup B_n = X$. But $B_i \subseteq U_i$ for each i. Thus $\{U_1, \ldots, U_n\}$ is a finite subcover of $\mathcal{U}$.

□

The following analogue of Proposition 20.2 has a very different proof.

Proposition 24.2 (Product Theorem). *Let $X_1, \ldots, X_n$ be spaces. Then $X_1 \times \cdots \times X_n$ is compact if and only if each X_i is compact.*

Proof. Let the product space be Π, and let $p_i : \Pi \to X_i$ be projection onto the ith coordinate. If Π is compact, we infer compactness in X_i by Proposition 23.3 and the fact that p_i is a continuous surjection.

For the converse, we prove the assertion in the case $n = 2$. For then, a very easy induction gets the result for arbitrary finite n.

So assume X and Y are both compact, and let $\mathcal{B}$ consist of all open boxes $U \times V$, where U and V are open in X and Y, respectively. Let $\mathcal{W}$ be a cover of $X \times Y$ by members of $\mathcal{B}$. By Proposition 24.1, it suffices to find a finite subcover of $\mathcal{W}$.

For each $x \in X$, the set $\{x\} \times Y \subseteq X \times Y$ is homeomorphic to Y, and is therefore compact. Let $\mathcal{W}_x$ be a finite subfamily of $\mathcal{W}$ covering $\{x\} \times Y$. Let's write $\mathcal{W}_x = \{U_1 \times V_1, \dots, U_n \times V_n\}$. Then $U_x = U_1 \cap \dots \cap U_n$ is an open neighborhood of x, and $U_x \times Y \subseteq \bigcup \mathcal{W}_x$. Now $\{U_x : x \in X\}$ is an open cover of the compact space X, so let $x_1, \dots, x_m$ be chosen so that X is covered by $\{U_{x_1}, \dots, U_{x_m}\}$. Then $\{U_{x_1} \times Y, \dots, U_{x_m} \times Y\}$ is a finite cover of $X \times Y$, and so $\mathcal{W}_{x_1} \cup \dots \cup \mathcal{W}_{x_m}$ is a finite subcover of $\mathcal{W}$.

□

The Full Heine-Borel Theorem. We are now in a position to push Proposition 23.6 into higher dimensions.

Proposition 24.3 (Heine-Borel Theorem III). *A subset of $\mathbb{R}^n$ (usual metric) is compact if and only if it is closed and bounded.*

Proof. If $A \subseteq \mathbb{R}^n$ is closed and bounded, then $A \subseteq [a_1, b_1] \times \dots \times [a_n, b_n]$, a product of closed bounded intervals. By Propositions 23.2 and 24.2, this product is compact. A is now a closed subset of a compact space, and is therefore compact by Proposition 23.4.

The rest of the proof is left as an exercise.

□

Compactness and the Separation Axioms. When the Hausdorff axiom—distinct points have disjoint neighborhoods—is coupled with compactness, there is a lot more that can be said regarding separating sets. The following is an analogue of Proposition 18.4, which showed normality for metrizable spaces.

Proposition 24.4 (Compact Hausdorff Normality Theorem). *All compact Hausdorff spaces are normal.*

Proof. Let X be compact Hausdorff, with A, B disjoint closed subsets of X. We wish to find disjoint open sets U, V, with $A \subseteq U$ and $B \subseteq V$.

Fix $a \in A$. In the proof of Proposition 23.4, since X is Hausdorff, we found an open neighborhood U_a of a missing B. And how we did this was to show that U_a is actually disjoint from an open set V_a containing B. So cover A with these open sets U_a for $a \in A$, keeping track of the corresponding open sets $V_a \supseteq B$, with $U_a \cap V_a = \emptyset$. Then there is a finite subfamily $\{U_{a_1}, \dots, U_{a_n}\}$ covering A since A is compact. Set $U = U_{a_1} \cup \dots \cup U_{a_n}$ and $V = V_{a_1} \cap \dots \cap V_{a_n}$.

□

Exercises 24. (1) Finish the proof of Proposition 24.3

(2) Prove that a continuous bijection from a compact space to a Hausdorff space is a homeomorphism.

(3) Furnish the details of the proof of Proposition 24.4.
(4) Let $\langle X, d\rangle$ be a metric space. Show that every compact subset is both closed and bounded. Exhibit a metric space for which not every closed bounded subset is compact.
(5) If $\langle X, d\rangle$ is a compact metric space, show that d is *totally bounded*, in the sense that if $\epsilon > 0$, there is a finite set $A = A_\epsilon \subseteq X$ such that $d(x, A) < \epsilon$ for every $x \in X$. (Think of antenna masts in a cell phone network.)
(6) Prove the Extreme Value Theorem of calculus: If $f : [a, b] \to \mathbb{R}$ is continuous, then f achieves both a maximum value and a minimum value.

Lecture 25: Other Forms of Compactness

Abstract. A space is countably compact if every countable open cover has a finite subcover. The space satisfies the Bolzano-Weierstrass property if each of its infinite subsets has an accumulation point. Compactness implies countable compactness, which, in turn, implies satisfying the Bolzano-Weierstrass property. In metric spaces, all three conditions are equivalent. Also we prove Lebesgue's theorem, that if a metric space satisfies the Bolzano-Weierstrass property, then for each open cover of the space, there is a positive radius such that each open ball of that radius lies in a member of the open cover.

Compactness has a rich history; there are several informative articles on the internet that give the details. The open cover version is due to Pavel Aleksandroff and Pavel Urysohn in 1929. They called the property *bicompact* to avoid terminology clashes; but over time their definition became recognized as the most basic and far reaching, and the *bi-* prefix fell out of use. In this lecture we concentrate on two other compactness-like properties. Both of them are consequences of compactness in general, and are equivalent to compactness for metrizable spaces.

One intuition regarding compactness is that if you take an infinite number of distinct "steps" in the space, some of the steps must get arbitrarily close to some particular point of the space. The first compactness-like property we consider is a mathematical realization of this intuition. Define a space X to satisfy the *Bolzano-Weierstrass property* if each infinite subset of X has an accumulation point in X.

Suppose that $A \subseteq X$ has *no* accumulation point in X. Then every point in $X \setminus A$ has a neighborhood missing A; hence A is closed. Furthermore, every point of A has a neighborhood that misses every other point of A; this says that A is a *discrete* subset of X: as a subspace of X, A inherits the discrete topology; every subset of A is relatively open in A.

Clearly closed discrete subsets of a space have no accumulation points in the space, so the Bolzano-Weierstrass property in a space amounts to saying that every closed discrete subset of the space is finite.

Our second compactness-like property is a trivial consequence of—and bears a close surface resemblance to— compactness itself. A space is *countably compact* if each of its countable open covers has a finite subcover.

Proposition 25.1 . *Every countably compact space satisfies the Bolzano-Weierstrass property.*

Proof. Suppose X fails to satisfy the Bolzano-Weierstrass property, and let A be an infinite closed discrete subset. Then A, being infinite, contains a countably infinite set C. Since A has no accumulation points in X, so does any subset of A; so we infer that C also is closed and discrete in X. let $\mathcal{U} = \{X \setminus C\} \cup \{U_c : c \in C\}$, where U_c is an open neighborhood of c that misses the rest of C; i.e., $U_c \cap (C \setminus \{c\}) = \emptyset$. Then $\mathcal{U}$ is an open cover of X, which is countable because C is. For each $c \in C$ the only member of $\mathcal{U}$ containing c is U_c; hence any finite subfamily of $\mathcal{U}$ would fail to cover C. Thus $\mathcal{U}$ is a countable open cover with no finite subcover, witnessing the fact that X is not countably compact.

□

Our aim in this lecture is to show that, for metrizable spaces, all three compacness-like properties are equivalent; and as a step towards this, we introduct the following distinctly metric notion.

A metric space $\langle X, d\rangle$ satisfies the *Lebesgue property* if whenever $\mathcal{U}$ is an open cover of X, there is a real $\lambda > 0$—called a *Lebesgue number* for $\mathcal{U}$—such that each d-ball of radius λ lies in some member of $\mathcal{U}$.

Examples 25.2. (1) Let $\mathbb{N}^+$ be the positive integers, and consider the open covering $\mathcal{U} = \{(-n, n) : n \in \mathbb{N}^+\}$ of $\mathbb{R}$. Then every $\lambda > 0$ is a Lebesgue number for $\mathcal{U}$.

(2) For each $n \in \mathbb{N}^+$, let $U_n = (n - \frac{1}{n}, n + \frac{1}{n})$. Then $\mathcal{U} = \{\mathbb{R} \setminus \mathbb{N}^+\} \cup \{U_n : n \in \mathbb{N}^+\}$ is an open cover of $\mathbb{R}$ with no Lebesgue number: if $\lambda > 0$ is given, choose n large enough so that $1/n < \lambda$. Then $(n - \lambda, n + \lambda)$ lies in no member of $\mathcal{U}$.

Proposition 25.3 (Lebesgue's Theorem). *Every metric space satisfying the Bolzano-Weierstrass property also satisfies the Lebesgue property.*

Proof. In the interests of deriving a contradiction, assume $\langle X, d\rangle$ satisfies the Bolzano-Weierstrass property, but also there is an open cover $\mathcal{U}$ of X for which there is no Lebesgue number. For $n = 1, 2, \ldots$, $1/n$ fails to be a Lebesgue number, so there is a point $x_n \in X$ such that $B(x_n, \frac{1}{n})$ does not lie in any member of $\mathcal{U}$. If the set $A = \{x_n : n \in \mathbb{N}^+\}$ were finite, then there would be some $x \in X$ such that $x_n = x$ for infinitely many indices n. But pick any $x \in X$. Then there is some $U_x \in \mathcal{U}$ such that $x \in U_x$; hence there is an $\epsilon_x > 0$ such that $B(x, \epsilon_x) \subseteq U_x$. If $1/n < \epsilon_x$, then $x \neq x_n$ because $B(x_n, \frac{1}{n}) \not\subseteq U_x$. Hence, for each $x \in X$, $\{n : x = x_n\}$ is finite; i.e., A is infinite.

By the Bolzano-Weierstrass property, there is some accumulation point $a \in X$ for A. Then we have $\epsilon_a > 0$ such that $B(a, \epsilon_a) \subseteq U_a \in \mathcal{U}$. Because a is an accumulation point of A, each neighborhood of a intersects A in an infinite set (see Exercise 9(4)). Hence $B(a, \frac{1}{2}\epsilon_a) \cap A$ has infinitely many points. In particular, $\{n : x_n \in B(a, \frac{1}{2}\epsilon_a)\}$ is infinite. So pick n large enough so that $\frac{1}{n} < \frac{1}{2}\epsilon_a$, as well as $x_n \in B(a, \frac{1}{2}\epsilon_a)$. If $y \in B(x_n, \frac{1}{n})$, then $d(y, a) \leq d(y, x_n) + d(x_n, a) < \frac{1}{n} + \frac{1}{2}\epsilon_a < \frac{1}{2}\epsilon_a + \frac{1}{2}\epsilon_a = \epsilon_a$. Hence $B(x_n, \frac{1}{n}) \subseteq B(a, \epsilon) \subseteq U_a \in \mathcal{U}$, a contradiction.

□

In Exercise 24(5) you're asked to prove that compact metric spaces are totally bounded; i.e., that for each $\epsilon > 0$ there is a finite subset of the space such that each point of the space is within ϵ of some point of the subset. Here we show an ostensibly stronger result.

Proposition 25.4. *Every metric space satisfying the Bolzano-Weierstrass property is totally bounded.*

Proof. We prove the contrapositive. Suppose $\langle X, d\rangle$ is not totally bounded. Then there is some $\epsilon > 0$ such that no family of ϵ-balls covers X. Start with $a_1 \in X$.

Then $B(a_1, \epsilon) \neq X$, so there is some $a_2 \in X$ with $d(a_1, a_2) \geq \epsilon$. Suppose that we have a finite family $\{a_1, \ldots, a_n\}$ such that $d(a_i, a_j) \geq \epsilon$ for $i \neq j$, $1 \leq i, j \leq n$. Then $\{B(a_1, \epsilon), \ldots, B(a_n, \epsilon)\}$ does not cover X, so there is some a_{n+1} such that $d(a_{n+1}, a_i) \geq \epsilon$ for $i \leq i \leq n$.

Using induction, we have defined an infinite set $A = \{a_1, a_2, \ldots\}$ so that $d(a_i, a_j) \geq \epsilon$ for all $i, j \geq 1$, $i \neq j$. If $a \in A$, then the open ϵ ball centered at a misses the rest of A; hence A is discrete. If $x \in X \setminus A$, then $B(x, \frac{1}{2}\epsilon)$ can contain at most two points of A. (Use the triangle inequality.) Thus we may find a $\delta > 0$ small enough so that $B(x, \delta) \cap A = \emptyset$. This shows that we have an infinite closed discrete subset of X, so X cannot satisfy the Bolzano-Weierstrass property.

□

We are now ready to prove the main result of this lecture.

Proposition 25.5. *In metrizable spaces, the properties of compactness, countable compactness, and satisfying the Bolzano-Weierstrass property are all equivalent.*

Proof. We already know that compactness $\implies$ countable compactness $\implies$ satisfying the Bolzano-Weierstrass property in general, so we complete the circle of implications by showing that compactness follows from the Bolzano-Weierstrass property in metrizable spaces.

Suppose $\mathcal{U}$ is an open cover of X; say d is a metric compatible with the topology on X. By Proposition 25.3, there is some $\lambda > 0$ such that each open λ-ball is contained in some member of $\mathcal{U}$. And by Proposition 25.4, there are finitely many points $a_1, \ldots, a_n$ such that $\{B(a_1, \lambda), \ldots, B(a_n, \lambda)\}$ covers X. Since $B(a_i, \lambda) \subseteq U_i \in \mathcal{U}$ for some $1 \leq i \leq n$, the family $\{U_1, \ldots, U_n\}$ is a finite subcover of $\mathcal{U}$. Hence X is compact.

□

We did not show that the Bolzano-Weierstrass property implies compactness for general topological spaces, and for good reason: it's false. However, a proof is beyond the scope of this course.

Exercises 25.

(1) Show that a compact metric space is complete (see Lecture 14).
(2) Show that closed subspaces of countably compact spaces are countably compact. Do the same for the Bolzano-Weierstrass property.
(3) Prove that a subset A of $\mathbb{R}^n$ is compact if and only if every nested sequence $A_1 \supseteq A_2 \supseteq \ldots$ of relatively closed nonempty subsets of A has nonempty intersection.
(4) Prove that every totally bounded metric space is bounded.
(5) Prove that every closed discrete subset of the real line is countable. [Hint: the union of a countable family of finite sets is countable.]
(6) Show by example that satisfying the Bolzano-Weierstrass property is not inherited by subspaces.

Lecture 26: Countability Properties of Metric Spaces

Abstract. A space is second countable if it has an open base consisting of countably many sets; it is first countable if each point has a countable neighborhood base. A metric space is always first countable; and it is second countable just in case either of the following two conditions holds: (i) it is separable; i.e., it has a countable dense subset; or (ii) it satisfies the Lindelöf property; i.e., every open cover has a countable subcover.

First and Second Countability. Let $\langle X, \mathcal{T} \rangle$ be a topological space, with $x \in X$. As done previously, we denote by $\mathcal{T}_x$ those open sets that contain x. A family $\mathcal{B}_x \subseteq \mathcal{T}_x$ is an *open neighborhood base* at x if each set in $\mathcal{T}_x$ contains a set in $\mathcal{B}_x$. X is *first countable* if every one of its points has a countable open neighborhood base.

Recall that a family $\mathcal{B} \subseteq \mathcal{T}$ is an *open base* for $\mathcal{T}$ if for each $x \in X$ and each $U \in \mathcal{T}_x$, there is some $B \in \mathcal{B}$ such that $x \in B \subseteq U$. X is *second countable* if its topology has a countable open base.

So first countability is a "local" property of a space, while second countability is a "global" one.

Proposition 26.1 . (1) *Every second countabe space is first countable.*
(2) *Every metrizable space is first countabile.*
(3) *Not every metrizable space is second countable.*

Proof. Ad (1): If $\mathcal{B}$ is a countable open base for $\langle X, \mathcal{T} \rangle$ and $x \in X$, let $\mathcal{B}_x = \{B \in \mathcal{B} : x \in B\}$. Then $\mathcal{B}_x$ is countable because its superset $\mathcal{B}$ is. If $U \in \mathcal{T}_x$, there is some $B \in \mathcal{B}$ with $x \in B \subseteq U$. This B is automatically in $\mathcal{B}_x$, so $\mathcal{B}_x$ is a countable open neighborhood base at x.

Ad (2): Suppose $\langle X, \mathcal{T} \rangle$ is metrizable, and let d be a compatible metric. For $x \in X$, let $\mathcal{B}_x$ consist of all open balls $B(x, \frac{1}{n})$, $n \in \mathbb{N}^+$. Then $\mathcal{B}_x$ is countable; if $U \in \mathcal{T}_x$, there is some $\epsilon > 0$ such that $B(x, \epsilon) \subseteq U$. Pick $n \in \mathbb{N}^+$ such that $\frac{1}{n} < \epsilon$ (the archimedean property again). Then $B(x, \frac{1}{n}) \subseteq U$ also, proving that $\mathcal{B}_x$ is a countable open neighborhood base at x.

Ad (3): Let X be an uncountable set, with d the $\{0,1\}$-valued discrete metric. Then $\{x\}$ is an open set for each $x \in X$. Suppose $\mathcal{B}$ is an open base for the discrete topology. Then $\{x\}$ is a union of members of $\mathcal{B}$; hence it must be the case that $\{x\} \in \mathcal{B}$. Thus $\mathcal{B}$ has at least as many sets as there are points of X; hence $\mathcal{B}$ is uncountable. This is true for any open base, so X cannot be second countable.

□

Separability and the Lindelöf Property. Recall that a subset A of a space X is *dense* if $\mathrm{cl}(A) = X$; equivalently if every nonempty open subset of X intersects A. X is defined to be *separable* if it has a countable dense subset. The real line is separable because it has the set of rational numbers as a countable dense subset. An uncountable discrete space cannot be separable because the only dense subset is the set itself.

The next property is defined using a variation on the open cover definitions of compactness and countable compactness. A space X satisfies the *Lindelöf property* (or is a *Lindelöf space*) if every open cover of X has a countable subcover. So clearly, being a countably compact space that satisfies the Lindelöf property means being compact.

Proposition 26.2. *Every second countable space is separable and satisfies the Lindelöf property.*

Proof. Suppose X has a countable open base $\mathcal{B}$, consisting of nonempty sets. For each $B \in \mathcal{B}$, let $x_B \in B$ be chosen. Then $A = \{x_B : B \in \mathcal{B}\}$ is a countable subset of X because $\mathcal{B}$ is countable. To see that A is dense, it suffices to show that every nonempty open subset of X intersects A. But every nonempty open U contains some $B \in \mathcal{B}$, and this B contains $x_B \in A$. Thus X is separable.

To prove the Lindelöf property, suppose $\mathcal{U}$ is an open cover of X. For each $x \in X$, there is a member of $\mathcal{U}$ containing x, so there is a member B_x of the countable base $\mathcal{B}$ contining x and contained in a member of $\mathcal{U}$. Since there are only countably many such sets B_x—even though there may be uncountably many points x—we have a countable cover $\{B_1, B_2, \dots\}$ of X consisting of sets from $\mathcal{B}$, each one contained in a member of $\mathcal{U}$. Suppose $B_n \subseteq U_n \in \mathcal{U}$ for $n = 1, 2, \dots$. Then $\{U_1, U_2, \dots\}$ is a countable subcover of $\mathcal{U}$.

□

Proposition 26.3. *For metrizable spaces, second countability, separability, and satisfying the Lindelöf property are all equivalent.*

Proof. By Proposition 26.2, second countability implies satisfying the Lindelöf property; we complete the "round robin" by showing that—in the presence of metrizablity—Lindelöf $\implies$ separable $\implies$ second countable.

Suppose $\langle X, d\rangle$ is a metric space satisfying the Lindelöf property. Just as in the proof that compact metric spaces are totally bounded, we may obtain, for each $n \in \mathbb{N}^+$, a countable set A_n such that every $x \in X$ is within distance $1/n$ of some member of A_n. (Just cover X with open $1/n$-balls and extract a countable subcover, letting A_n consist of the centers of the balls in that subcover.) Then $A = \bigcup_{n\in\mathbb{N}^+} A_n$ is a countable union of countable sets, hence countable. If $B(x, \epsilon)$ is any open ball, pick n large enough so that $1/n < \epsilon$. Then $B(x, \epsilon) \cap A_n \neq \emptyset$, so $B(x, \epsilon) \cap A \neq \emptyset$. Thus A is a countable dense subset of X, making X separable.

Now suppose $\langle X, d\rangle$ is separable, with $A \subseteq X$ countable and dense. As in the proof of Proposition 26.1(2), for each $a \in A$, let $\mathcal{B}_a = \{B(a, \frac{1}{n}) : n \in \mathbb{N}^+\}$ be the standard countable open neighborhood base at a. Let $\mathcal{B} = \bigcup_{a\in A} \mathcal{B}_a$. Then $\mathcal{B}$ is a countable family of open sets—being a countable union of countable families of open sets—and it remains to show it to be an open base. So let U be open, with $x \in U$. We find $\epsilon > 0$ such that $B(x, \epsilon) \subseteq U$. Next find $n \in \mathbb{N}^+$ large enough so that $\frac{1}{n} < \frac{\epsilon}{2}$. Finally, since A is dense, we find $a \in A$ with $d(a, x) < \frac{1}{n}$. Then $x \in B(a, \frac{1}{n}) \in \mathcal{B}$. Moreover, if $y \in B(a, \frac{1}{n})$, then $d(y, x) \leq d(y, a) + d(a, x) < \frac{1}{n} + \frac{1}{n} < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$. Hence $x \in B(a, \frac{1}{n}) \subseteq B(x, \epsilon) \subseteq U$, showing that $\mathcal{B}$ is a countable open base for X.

□

We end this section with a beautiful result that infers metrizability from purely topological conditions. Recall from Lecture 18 that we defined a space to be Housdorff (or T_2) if distinct points may be separated by disjoint open sets, and to be normal (or T_4) if the same can be said for separating disjoint closed sets. The important separation axiom in the middle is *regularity* (or being a T_3 space) and stipulates that a point and closed set not containing the point may be so separated. It was skipped over in Lecture 18 for time considerations; we consider it here only for the purpose of stating—but not proving—Urysohn's famous metrization theorem.

Proposition 26.4 (Urysohn's Metrization Theorem). *Any regular second countable space is metrizable.*

Remark 26.5. By way of a historical aside: Proposition 26.4 was actually proved by Tychonoff in 1926. What Urysohn proved a year earlier was metrizability for *normal* second countable spaces.

Preservation Theorems (Optional). Any result telling us that a topological property pertaining to a space extends to certain kinds of subspaces, or certain kinds of continuous images, or products of spaces, is called a "preservation theorem." Examples we have already seen include:

- Connectedness and compactness are preserved by continuous surjections; local connectedness is preserved by identification maps.
- All subspaces of metrizable spaces are metrizable; closed subspaces of compact spaces are compact.
- The product of two compact (resp., connected) spaces is compact (resp., connected).

In the exercises below you are asked to prove some fairly straightforward preservation theorems regarding the countability properties discussed in this lecture. For example, separability is preserved by all continuous surjections, but second countability seems to require extra conditions on the mapping—e.g., openness. (Examples telling us conclusively that not all continuous images of second countable spaces are second countable involve constructions of non-metrizable spaces, a subject for a later course.) On the other hand, separability is inherited by open—but not all—subspaces, while there is no such restriction on second countability.

Here we present a somewhat less straightforward result concerning the preservation of second countability by continuous maps. It is most applicable if the domain is also compact (see Exercise 26(6)), but this is not a necessary condition. We first define a mapping to be *perfect* if the inverse image of points in the range space are compact subsets of the domain.

Proposition 26.6. *Second countability is preserved by perfect continuous closed surjections.*

Proof. Let $f : X \to Y$ be a perfect continuus closed surjection, and suppose $\mathcal{B}$ is a countable open base for X. For each $B \subseteq \mathcal{B}$, $X \setminus B$ is closed in X; and since f is a closed map, $f[X \setminus B]$ is closed in Y. Thus $Y \setminus f[X \setminus B]$ is open in Y; hence $\mathcal{C} = \{Y \setminus f[X \setminus B] : B \in \mathcal{B}\}$ is a countable family of open subsets of Y. We will be done once we show it to be an open base.

So pick an open $V \subseteq Y$, with $y \in V$. We need to find some $C \in \mathcal{C}$ with $y \in C \subseteq V$.

Now, we lose no generality in assuming that the union of a finite number of sets in $\mathcal{B}$ is also in $\mathcal{B}$. (If $\mathcal{B}$ doesn't satisfy this condition, we replace it with the smallest superfamily of open sets that does. The new family is still an open base, and it is still countable.)

Back to $y \in V$, we let $x \in f^{-1}(y)$. $f^{-1}[V]$ is open in X, so there is a $B_x \in \mathcal{B}$ with $x \in B_x \subseteq f^{-1}[V]$. These sets B_x, for $x \in f^{-1}(y)$, form an open cover of the compact set $f^{-1}(y)$, so a finite number of them suffice. By our assumption in the last paragraph, we know that there is a single $B \in \mathcal{B}$ such that $f^{-1}(y) \subseteq B \subseteq f^{-1}[V]$.

It remains to show that $y \in Y \setminus f[X \setminus B] \subseteq V$. Since $f^{-1}(y) \cap (X \setminus B) = \emptyset$, y cannot be in $f[X \setminus B]$. Thus $y \in Y \setminus f[X \setminus B]$.

Finally, suppose $z \in Y \setminus f[X \setminus B]$. Then $z \notin f[X \setminus B]$, so $f^{-1}(z) \cap (X \setminus B) = \emptyset$. Hence $f^{-1}(z) \subseteq B \subseteq f^{-1}[V]$, so $z \in V$. This proves $y \in Y \setminus f[X \setminus B] \subseteq V$, as desired.

□

Exercises 26. (1) Prove that every subpace of a second countable space is second countable.

(2) Prove that every open subspace of a separable space is separable.
(3) Prove that every closed subspace of a Lindelöf space is Lindelöf.
(4) Prove that every subspace of a separable (resp., Lindelöf) metrizable space is separable (resp., Lindelöf).
(5) Prove that separability and satisfying the Lindelöf property are continuous invariants, and that second countability is a continuous open invariant.
(6) Prove that if X is compact and second countable, Y is Hausdorff, and $f : X \to Y$ is a continuous surjection, then Y is second countable.
(7) Prove that, in a Lindelöf space, every uncountable subset has an accumulation point in the space.
(8) Show that if X is second countable then any family of nonempty pairwise disjoint open subsets of X must be countable.
(9) Prove that the product of two spaces is separable (resp., second countable) if and only if each of the spaces is separable (resp., second countable).
(10) The product of two Lindelöf spaces needn't be Lindelöf, but an example is beyond the scope of this course. Show that if the product is Lindelöf then so is each of the factors. Use an argument analogous to that given in the proof of Proposition 24.2 to show that the product of two Lindelöf spaces is Lindelöf if one (or more) of the factors is compact.

Lecture 27: Introduction to Continua

Abstract. A continuum is a compact connected space. Any downwardly-directed intersection of subcontinua of a Hausdorff space is also a subcontinuum; this provides an important source of Hausdorff continua. In fact, using this process, we show how to construct a nondegenerate (i.e., more than one point) metric continuum that is indecomposable; i.e., not the union of two proper subcontinua.

A *continuum*—the plural is *continua*—is a nonempty topological space that is also compact and connected. Since both compactness and connectedness tend to favor "few" open sets, and the separation axioms tend to favor "many," the addition of a separation axiom to the mix lends a "dynamic tension" that is very popular in mathematics. So addition of the T_1 axiom—finite sets are closed—gives us T_1 *continua*; and, more importantly, the addition of the T_2 axiom—distinct points can be separated by disjoint open sets—gives us *Hausdorff continua.* T_1 continua need not be Hausdorff, but Hausdorff continua are normal (Proposition 24.4). So the next most reasonabe way of restricting the class of topological continua is obtained not by stipulating normality, but by upping the ante to metrizability. Since this course is essentially about metric topology, all of our examples of continua are metrizable. However, for many of our results, the metric assumption is not necessary.

A note about terminology: many practitioners in the study of continua use the term *continuum/ua* to refer to what we call *metric continuum/ua.* It makes sense in their case to drop the modifier, since all they're talking about is metric spaces.

By a *subcontinuum* of a topological space, we mean a nonempty subspace that is also compact and connected. Since subspaces of T_1 (resp., Hausdorff, metrizable) spaces are T_1 (resp., Hausdorff, metrizable), a subcontinuum of a topological space will inherit any of the separation properties mentioned above that pertain to the space.

The term *nondegenarate* applied to a topological space means that the space has two or more points. A nondegenerate topological continuum can have exactly two points, but a nondegenerate T_1 continuum must be infinite. And while there are countable nondegenarate T_1 continua, any nondegenerate Hausdorff continuum must contain at least as many points as there are real numbers—with this cardinal number being not only possible, but ubiquitous: all nondegenerate continua have exactly the cardinality of the real line. (The proofs of these statements must be deferred to a later course.)

Some Basic Examples. As shown in Lecture 20, the only subcontinua of the real line are the closed bounded intervals, and they're all homeomorphic to $\mathbb{I}$. As before, we refer to any homeomorphic copy of $\mathbb{I}$ as an *arc*. Arcs have exactly two kinds of points from a topological point of view: cut points, those points whose removal leaves a disconnected set—i.e., any $t \in \mathbb{I}$ with $0 < t < 1$—and noncut points, in this case the end points $0, 1 \in \mathbb{I}$.

In contrast to the one-dimensional situation, the subcontinua of $\mathbb{R}^2$ are so varied and difficult to characterize that many modern topologists study them exclusively. (So that is one topological difference between 1 and 2.)

We have already met some of the most basic examples of continua in earlier lectures. For example we have the n-cube $\mathbb{I}^n = [0,1]^n$ (a.k.a. the n-cell) and the

n-sphere $\mathbb{S}^n = \{x \in \mathbb{R}^n : \|x\| = 1\}$, the vectors in $\mathbb{R}^n$ of unit norm, *not*, as the notation would suggest, a cartesian power of anything. We also have the "exotic" examples of the topologist's sine curve (Lecture 21) and its relative, the Warsaw circle (Lecture 22). Both the latter continua fail to be locally connected; what makes them different from each other is that only the Warsaw circle is path connected.

Directed Intersections. In the setting of Hausdorff continua, the compact subsets—in particular the subcontinua—are all closed. This is a very useful feature.

Recall that a *nested sequence* in a set X is a family $\{A_1, A_2, \dots\}$ of subsets of X with $A_n \supseteq A_{n+1}$ for all $n \in \mathbb{N}^+$. Nested sequences are special cases of *nested families*, families $\mathcal{C}$ of subsets of X such that if $A, B \in \mathcal{C}$, then either $A \subseteq B$ or $B \subseteq A$. Still more generally, $\mathcal{C}$ is *downward* (resp., *upward*) *directed* if, for any $A, B \in \mathcal{C}$ there is a $C \in \mathcal{C}$ such that $C \subseteq A \cap B$ (resp., $C \supseteq A \cup B$). Clearly nested families are both downward and upward directed. The following affords us a very powerful method for producing continua.

Proposition 27.1 (Directed Intersections Theorem) . *In a compact Hausdorff space, the intersection of a nonempty downward directed family of subcontinua is a subcontinuum.*

Proof. Let $\mathcal{C}$ be a downward directed family of subcontinua of compact Hausdorff space X. Then each $A \in \mathcal{C}$ is closed, so $C = \bigcap \mathcal{C}$ is closed. Since X is compact, C is compact too. It remains to show C is nonempty and connected.

Since $\mathcal{C}$ is a downward directed family, it satisfies the finite intersection property: an easy induction shows that the intersection of any finite subfamily of $\mathcal{C}$ contains a member of $\mathcal{C}$, and is hence nonempty. And since $\mathcal{C}$ consists of closed subsets of a compact Hausdorff space, its intersection must be nonempty, by Proposition 23.5.

Suppose, for the sake of contradiction as always, that C is disconnected; say $\{H, K\}$ is a disconnection of C. Then both H and K are nonempty closed sets. And, since they're disjoint, and since compact Hausdorff spaces are normal, there are disjoint open sets $U, V \subseteq X$ such that $H \subseteq U$ and $K \subseteq V$.

Now both U and V intersect every $A \in \mathcal{C}$. Because each such A is connected, it cannot be the case that $A \subseteq U \cup V$. This tells us that the family $\{X \setminus (U \cup V)\} \cup \mathcal{C}$ of closed subsets of X satisfies the finite intersection property, and hence has nonempty intersection. But this intersection is $(X \setminus (U \cup V)) \cap C$, empty since $C \subseteq U \cup V$. This contradiction tells us C is a subcontinuum after all.

□

Indecomposable Continua. One of the lessons of continuum theory is: "Don't take anything for granted." Many of the intuitions about the physical world that we extend to continua are either very difficult to prove or false outright. One of the false ones is always being able to take a nondegenerate continuum and decompose it as the union of two proper—but necessarily overlapping—subcontinua. All the continua considered so far have this property, but nested intersections give us ones that do not.

A continuum X is *decomposable* if there are proper subcontinua A, B of X such that $X = A \cup B$. X is *indecomposable* otherwise. While you can "break a decomposable continuum in two," a nondegenerate indecomposable metric continuum will only "shatter into an uncountable number of pieces, each with empty interior." Here we construct a nondegenerate indecomposable continuum in the plane using nested intersections. (While many examples of indecomposable continua have attributions and even pet names—e.g., Knaster's Buckethandle, Moise's Pseudo-arc, the Lakes of Wada—this one does not. However its first appearance seems to be in the 1961 text *Topology*, by J. G. Hocking and G. S. Young, and those authers are at pains to pin a maker's mark on all the other nontrivial examples they describe. So we must conclude that this example originated with them.)

For $n \geq 2$, a finite indexed family $\mathcal{C} = \{C_1, \dots, C_n\}$ of subsets of a set X is called a *simple chain between x and z through y* if:

- $x \in C_1 \setminus (C_2 \cup \cdots \cup C_n)$;
- $z \in C_n \setminus (C_1 \cup \cdots \cup C_{n-1})$;
- $y \in C_1 \cup \cdots \cup C_n$; and
- $C_i \cap C_j \neq \emptyset$ if and only if $|i - j| \leq 1$.

The elements of a simple chain are called *links*; each link intersects only its nearest neighbors.

To make our construction, we begin by fixing three points a, b, c in $\mathbb{R}^2$, the vertices of an equilateral triangle, say. We create a nested sequence K_n of subcontinua, each of which is the union of the members of a simple chain $\mathcal{C}_n$ whose links are closed disks, each of diameter $\leq 2^{-n}$, $n = 0, 1, 2, \dots$. Every disk in $\mathcal{C}_{n+1}$ is contained in some disk in $\mathcal{C}_n$; furthermore, each $\mathcal{C}_{3i}$ is between b and c through a, each $\mathcal{C}_{3i+1}$ is between c and a through b, and each $\mathcal{C}_{3i+2}$ is between a and b through c.

Each K_n is a subcontinuum of $\mathbb{R}^2$ containing a, b, c. Hence, by Proposition 27.1, the same goes for $K = \bigcap_{n=0}^{\infty} K_n$. We show that K is indecomposable.

Suppose M is a proper subcontinuum of K, and—for the sake of obtaining a contradiction—assume both b and c are in M. Since M is proper, we may fix a point $p \in K \setminus M$. Then the euclidean distance from p to M is positive, so it is possible to find $k \in \mathbb{N}$ large enough so that no disk in $\mathcal{C}_{3k}$ containing p—of which there are either one or two—can intersect M. Let $\mathcal{C}_{3k} = \{C_1, \dots, C_m\}$. Since $\mathcal{C}_{3k}$ is a chain between b and c, neither C_1 nor C_m can contain p. Let $\mathcal{C}'_{3k}$ be the result of removing all the (one or two) members of $\mathcal{C}_{3k}$ containing p. Then $\bigcup \mathcal{C}'_{3i}$ has two components, A and B, one containing b, the other containing c. But $M \subseteq A \cup B$; hence we have obtained a disconnection of M.

We infer that no proper subcontinuum of K can contain both b and c; by symmetry, no proper subcontinuum of K can contain any two of the three points a, b, c. So if we could decompose K into proper subcontinua L and M, then the pigeonhole principle would be roundly violated and we would have a contradiction. Therefore K is a nondegenerate—because it contains at least three points—indecomposable continuum.

Exercises 27. (1) A continuum X is *irreducible about* two of its points if no proper subcontinuum contains both of the points. Show that an arc is irreducible about two of its points. Show that the indecomposable continuum K defined above is irreducible about any two of the three points a, b, c.

(2) If a continuum has three points such that it is irreducible about any two of them, then prove that the continuum is indecomposable. (For nondegenerate metrizable continua, the converse is also true.)
(3) A continuum X is *unicoherent* if, whenever X is decomposed into the union $A \cup B$ of two subcontinua, the intersection $A \cap B$ is connected. Prove that every every arc and every indecomposable continuum is unicoherent. Is a circle unicoherent? How about the figure eight curve (i.e., two circles intersecting at a single point)?
(4) Give an example of a nested sequence of connected subsets of the euclidean plane, whose intersection is disconnected. Show that there can be no such example if we consider connected subsets of the euclidean line.
(5) A continuum is *hereditarily indecomposable* if each of its subcontinua is indecomposable. (Believe it or not, such continua exist in abundance.) Show that if A and B are any two subcontinua of a hereditarily indecomposable continuum, then either $A \subseteq B$ or $B \subseteq A$ or $A \cap B = \emptyset$.
(6) If X is a continuum and $a \in X$, then the *composant* of X containing a is $\{b \in X : X \text{ is not irreducible between } a \text{ and } b\}$. Show that composants are always connected. What are the composants of an arc? A simple closed curve?
(7) Let X be a connected topological space, and let $\mathcal{A}$ be a finite cover of X by closed connected subsets. Prove that there is some $A \in \mathcal{A}$ such that $\bigcup\{B \in \mathcal{A} : B \neq A\}$ is connected.
(8) Prove that a continuum—indeed, a compact space—is metrizable if and only if it is Hausdorff and second countable.

Lecture 28: Irreducibility

Abstract. We introduce Zorn's lemma, as an important tool in proving continuum-theoretic facts. One fact is that if a subset of a Hausdorff continuum is given, there is a smallest subcontinuum containing that subset. The subcontinuum is called irreducible about the subset; a continuum is called irreducible if it is irreducible about a two-point set. For a given point in a continuum, the composant of that point is the set of all other points such that the continuum is not irreducible about the given point and the other point. A metric continuum has exactly one, three, or uncountably many composants.

In the exercises of Lecture 27 we introduced the notion of irreducibility about two points. Somewhat more generally, if X is a continuum and $A \subseteq X$ is nonempty, then we say that X is *irreducible about* A if no proper subcontinuum of X contains A. One goal of this lecture is to show that any nonempty subset of a Hausdorff continuum is contained in a subcontinuum that is irreducible about that subset. But first we need some new notions.

Partial Orderings and Zorn's Lemma. A *partially ordered set* is a pair $\langle P, \leq \rangle$, where P is a set and $\leq$ is a binary relation on—i.e., a set of ordered pairs of—P. We write $x \leq y$ to indicate that x is $\leq$-related to y (see Lecture 4). For example, if $P = \wp(X)$, the set of subsets of X, we could write $A \leq B$ to mean $A \subseteq B$ (or $A \supseteq B$).

For a binary relation $\leq$ on P to be a fully fledged partial order, it needs to satisfy: (1) the *reflexivity* condition, $x \leq x$; (2) the *antisymmetry* condition, $x = y$ if $x \leq y$ and $y \leq x$; and (3) the *transitivity* condition, $x \leq z$ if both $x \leq y$ and $y \leq z$. If it so happens that $\langle P, \leq \rangle$ also satisfies the condition that each two of its elements are $\leq$-comparable—for any $x, y \in P$, either $x \leq y$ or $y \leq x$—then the partial order is called *total*. For example, the real line with $x \leq y$ meaning that $x = y$ or x lies to the left of y, is a total order. The power set of any set with more than one element (with $\leq$ either $\subseteq$ or $\supseteq$) is a partial order that is not total.

To say that a subset $A \subseteq P$ of a partially ordered set is *totally ordered*, we simply mean that any two elements of A are $\leq$-comparable. If $A \subseteq P$ and $x \in P$ is such that $a \leq x$ for all $a \in A$, we say x is an *upper bound* for A. An element $m \in P$ is *maximal* if whenever $x \in P$ and $m \leq x$, we have $m = x$: m is not strictly less than anything else in P. [Note: What maximality does *not* say is that everything in P is $\leq m$.]

The following "maximality principle" is not provable from the usual set-theoretic axioms; it is one of many logically equivalent forms of the Axiom of Choice. It is due to Max Zorn and Kazimierz Kuratowski, and is commonly referred to as "Zorn's lemma." Its main attribute is that it asserts the existence of something without actually providing a construction; when we invoke it, we are providing a "nonconstructive" proof. We state it here as an axiom of set theory, which we believe and take on faith.

Zorn's Lemma. *Let $\langle P, \leq \rangle$ be a nonempty partially ordered set, such that every nonempty totally ordered subset has an upper bound in P. Then P has a maximal element.*

Irreducibility Existence. If X is a simple closed curve—i.e., homeomorph of a circle—and A consists of two points on the curve, there are exactly two subcontinua of X that are irreducible about A, and neither one has anything special over the other. In general, given $\emptyset \neq A \subseteq X$, there is no constructive procedure for producing K_A, "the" subcontinuum of X irreducible about A. In particular, this is not an analogue of the closure operation. While the intersection of all closed subsets containing A is *the* smallest closed set containing A, the intersection of all subcontinua containing A is a compact superset of A—at least if the ambient space is Hausdorff—that may well fail to be connected.

Proposition 28.1 (Irreducibility Existence Theorem). *Let X be a Hausdorff continuum, with $\emptyset \neq A \subseteq X$. Then there is a subcontinuum of X that is irreducible about A.*

Proof. We let $\langle \mathcal{P}, \leq \rangle$ be the set of all subcontinua of X that contain A—nonempty because $X \in \mathcal{P}$–where $K \leq M$ in $\mathcal{P}$ means $K \supseteq M$. Then a totally ordered subset of $\mathcal{P}$ is a nested family, *à la* Lecture 27. By Proposition 27.1, the intersection of a nested family of subcontinua containing A also is a subcontinuum containing A; hence every nonempty totally ordered subset of $\mathcal{P}$ has a $\leq$-upper bound in $\mathcal{P}$. By Zorn's lemma, $\langle \mathcal{P}, \leq \rangle$ has a maximal element, meaning that there is a subcontinuum K of X such that $A \subseteq K$ and such that no proper subcontinuum of K can contain A.

□

Composants. A continuum is *irreducible* if it is irreducible about a two-point subset. For each $x \in X$, the *composant* of X containing x is denoted $\kappa(x)$, and is the union of all proper subcontinua of X that contain x. Alternatively, it is the set $\{y \in X :$ is not irreducible about $\{x, y\}\}$. If $\kappa(x) = X$, then for each $y \in X$, there is a proper subcontinuum of X containing both x and y. So X is irreducible just in case some composant of X is a proper subset of X, and $x \in X$ is a *point of irreducibility* if $\kappa(x) \neq X$. The connection between irreducibility and indecomposability is a deep one, and we can only scratch the surface here.

Proposition 28.2. *Suppose x is a point of irreducibility for continuum X. Then there is no decomposition of X into two proper subcontinua, both of which contain x.*

Proof. Suppose $X = K \cup M$, where K and M are subcontinua and $x \in K \cap M$. Then X is irreducible about $\{x, y\}$ for some $y \in X$. If $y \in K$, then $K = X$; if $y \in M$, then $M = X$

□

Proposition 28.3. *If X is a decomposable continuum, then X is a composant. If, in addition, X is irreducible about $\{x, y\}$, then X, $\kappa(x)$, and $\kappa(y)$ are three distinct composants.*

Proof. Let $X = K \cup M$, where K and M are proper subcontinua. Then—since X is connected—there is some $z \in K \cap M$. For any $y \in X$; if $y \in K$ or if $y \in M$, X is not irreducible about $\{y, z\}$. Thus $\kappa(z) = X$.

Now suppose X is irreducible about $\{x, y\}$. Then $x \notin \kappa(y)$ and $y \notin \kappa(x)$; hence $\kappa(x)$ and $\kappa(y)$ are two new composants.

□

We state the following important result, concerning composants in *metrizable* continua, without proof. We include some of its corollaries in the exercises.

Proposition 28.4 (Composants Theorem). *Let X be a nondegenerate metrizable continuum. Then exactly one of the following conditions holds:*

(i) *X is decomposable and not irreducible, in which case X is the only composant.*
(ii) *X is decomposable and irreducible, in which case X has exactly three composants.*
(iii) *X is indecomposable, in which X has uncountably many pairwise disjoint composants.*

Remark 28.5. David Bellamy has produced examples of (necessarily nonmetrizable) indecomposable Hausdorff continua with exactly n composants, where $n = 1, 2$.

Exercises 28. (1) Prove that a Hausdorff continuum is irreducible about any dense subset.
(2) Prove that if a simple closed curve is irreducible about a subset, then the subset is dense.
(3) Prove that a nondegenerate indecomposable metrizable continuum is irreducible.
(4) Prove that a nondegenerate metrizable continuum is indecomposable if and only if it contains three points such that the continuum is irreducible about any two of those points.
(5) Give an example of a nondegenerate metrizable continuum that is irreducible about some subset with three points, but about no subset with just two points.
(6) Show that in any indecomposable continuum (no separation conditions necessary), the family of composants forms a partition of the continuum.

Lecture 29: Cut Points

Abstract. Cut points of connected spaces are points whose complements are disconnected subsets of the space. We prove that a Hausdorff continuum is decomposable just in case it contains a proper subcontinuum with nonempty interior; also we show—using Zorn's lemma—that every nondegenerate Hausdorff continuum contains at least two non-cut points.

A point x of a connected space X is a *cut point* if $X \setminus \{x\}$ is disconnected. A *cutting* of X is a triple $\langle x, U, V\rangle$, where $\{U, V\}$ is a disconnection of $X \setminus \{x\}$. Recall that we introduced the notion of cut point in Lecture 20, in order to prove that the real line—in which every point is a cut point—is not homeomorphic to euclidean n-space for $n > 1$: the latter has non-cut points; indeed no point is a cut point. Our main result in this lecture is that every nondegenerate Hausdorff continuum has at least two non-cut points.

A Preliminary Result. The following is a basic result about connected topological spaces.

Proposition 29.1. *Let X be a connected topological space, with A a connected subset. If H is clopen in $X \setminus A$, then $H \cup A$ is connected.*

Proof. In order to prove $H \cup A$ connected, suppose—by way of obtaining a contradiction—that $\{M, N\}$ is a disconnection of $H \cup A$. Since A is connected, either $A \subseteq M$ or $A \subseteq N$; say the first holds. Then $N \subseteq H$. Since N is clopen in $H \cup A$, N is clopen in the smaller set H. H is clopen in $X \setminus A$; hence N is clopen in $X \setminus A$.

Lemma: If N is closed (resp., open) in each of two subsets S and T of a space X, then N is closed (resp., open) in $S \cup T$.

Proof. (of lemma) Suppose $N = S \cap P = T \cap Q$, where P and Q are closed (resp., open) in X. Then $N = (S \cup T) \cap (P \cap Q)$. Indeed, the right side becomes $((S \cap P) \cap Q) \cup ((T \cap Q) \cap P) = (N \cap Q) \cup (N \cap P) = N$, since $N \subseteq P$ and $N \subseteq Q$. Since P and Q are both closed (resp., open) in X, so is their intersection. Hence N is closed (resp., open) in $S \cup T$.

□

N is clopen in $H \cup A$ as well as in $X \setminus A$; by the lemma, N is clopen in $(H \cup A) \cup (X \setminus A) = X$. Since N is not all of X—it misses M—we contradict the connectedness of X.

□

As an application of Proposition 29.1, we can give a nice charactization of decomposability in Hausdorff continua.

Proposition 29.2. *A Hausdorff continuum is decomposable if and only if it has a proper subcontinuum with nonempty interior.*

Proof. First assume X is decomposable, say $X = K \cup M$, where both K and M are proper subcontinua. Then $U = X \setminus M$ is a nonempty open subset of K; hence K is a proper subcontinuum of X with nonempty interior.

For the converse, suppose K is a proper subcontinuum of X, having nonempty interior. Let $P = X \setminus K$. If P is connected, then so is its closure $M = \text{cl}(P)$. And since K has interior points, M is a proper subcontinuum of X. Thus K and M give a decomposition of X into proper subcontinua.

On the other hand, P may be disconnected, say with disconnection $\{U, V\}$. By Proposition 29.1, both $U \cup K$ and $V \cup K$ are connected. Since P is open, so are U and V. $U = X \setminus (V \cup K)$ and $V = X \setminus (U \cup K)$, so $U \cup K$ and $V \cup K$ give a decomposition of X into proper subcontinua.

□

Existence of Non-cut Points. In contrast with the connected-but-not-compact real line, no Hausdorff continuum can consist of all cut points.

Proposition 29.3 (Non-cut Point Existence Theorem). *Every nondegenerate Hausdorff continuum has at least two non-cut points. Moreover, such a continuum is irreducible about its set of non-cut points.*

Proof. Let X be a nondegenerate Hausdorff continuum. If X has no cut points, we're done; so suppose $\langle c, U, V\rangle$ is a cutting of X. We show that U (and hence also V) contains a non-cut point of X. Assume, by way of obtaining a contradiction, that every point of U is a cut point of X. Then, for each $x \in U$, we have a cutting $\langle x, U_x, V_x\rangle$. If both U_x and V_x meet $V \cup \{c\}$, then they disconnect it, contradicting Proposition 29.1: $V \cup \{c\} \subseteq U_x \cup V_x$ because $x \notin V \cup \{c\}$. (This argument is used repeatedly in the rest of the proof.) So let's uniformly choose our labeling so that $U_x \subseteq U$ every time.

Let $\mathcal{K} = \{U_x \cup \{x\} : x \in U\}$, a family of subcontinua of X, by Proposition 29.1. Suppose x and y are distinct points of U. Can it be the case that both $x \in U_y$ and $y \in U_x$? If so, then we have $V_y \cup \{y\}$ intersecting U_x. $x \notin V_y \cup \{y\}$, so $V_y \cup \{y\} \subseteq U_x \cup V_x$. Since $V_y \cup \{y\}$ is connected, it cannot intersect V_x; hence $V_y \cup \{y\} \subseteq U_x$. But $c \in V_y \setminus U_x$, a contradiction.

So if $x \in U$ and $y \in U_x$, then $x \notin U_y$. $U_y \cup \{y\}$ is connected, intersects U_x, and is contained in $U_x \cup V_x$; hence $U_y \cup \{y\} \subseteq U_x$.

At this point, we create a partially ordered set $\mathcal{P}$, whose members are nested families in $\mathcal{K}$, and whose ordering is containment of one nested family in another. It is easy to check that the union of a nested family of nested families from $\mathcal{K}$ is itself a nested family from $\mathcal{K}$; hence, by Zorn's lemma, there is a maximal nested family $\mathcal{M} \subseteq \mathcal{K}$. This means that if $K \in \mathcal{K} \setminus \mathcal{M}$, then K is incomparable with some member of $\mathcal{M}$.

Let $M = \bigcap \mathcal{M}$. By Proposition 27.1, M is a subcontinuum of X. Let $m \in M$. Then m is also in U, so there is a cutting $\langle m, U_m, V_m\rangle$ of X, again where $U_m \subseteq U$. For each $U_x \cup \{x\} \in \mathcal{M}$, either $m \in U_x$ or $m = x$. In the first case, $U_m \cup \{m\} \subseteq U_x$ (see two paragraphs above); in the second, $U_m \cup \{m\} = U_x \cup \{x\}$. Therefore $U_m \cup \{m\} \subseteq M$.

Now let $r \in U_m$. Then $U_r \cup \{r\} \subseteq U_m \subseteq U_m \cup \{m\}$. Also $m \notin U_r$, by the argument four paragraphs above. Since $r \neq m$, we conclude that $U_r \cup \{r\}$ is a proper subset of $U_m \cup \{m\}$. This tells us that $U_r \cup \{r\} \in \mathcal{K} \setminus \mathcal{M}$, but also that $\mathcal{M} \cup \{U_r \cup \{r\}\}$ is a nested family from $\mathcal{K}$ that properly contains $\mathcal{M}$. Hence we contradict the maximality of $\mathcal{M}$, and there must be a non-cut point of X, located in U, afterall. This concludes the proof that every nondegenerate Hausdorff continuum contains at least two non-cut points.

Finally, let N be the set of non-cut points of X, with K a proper subcontinuum of X containing N. Let $x \in X \setminus K$. Then x is a cut point of X; hence there is a cutting $\langle x, U_x, V_x \rangle$ of X. Since $K \subseteq U_x \cup V_x$, we may assume that $K \subseteq V_x$. But then $N \subseteq V_x$, and hence there are no non-cut points of X in U_x. This contradicts the argument above.

□

Characterizing Arcs and Simple Closed Curves. We end this section by stating–but not proving–topological characterizations of arcs and simple closed curves.

Proposition 29.4 (Arc Characterization Theorem). *Let X be a topological space. The following two conditions are equivalent:*

(i) *X is an arc; i.e., a homeomorphic copy of the closed unit interval $\mathbb{I}$.*
(ii) *X is a metrizable continuum with exactly two non-cut points.*

A subset A of a connected space X is a *disconnecting* (or *separating*) *set* if $X \setminus A$ is disconnected. (So $x \in X$ is a cut point just in case $\{x\}$ is a disconnecting set.)

Proposition 29.5 (Simple Closed Curve Characterization Theorem). *Let X be a topological space. The following two conditions are equivalent:*

(i) *X is a simple closed curve; i.e., a homeomorphic copy of the unit circle.*
(ii) *X is a metrizable continuum with no cut points, such that every two-point subset is a disconnecting set.*

Remark 29.6. In light of Urysohn'e metrization theorem—Proposition 26.4—the word *metrizable* in Propositions 29.4 and 29.5 may be replaced with the much more topological-sounding *second countable Hausdorff.*

Exercises 29.

(1) Use Proposition 29.1 to show that if X is connected, A is a connected subset of X, and C is a component of $X \setminus A$, then $X \setminus C$ is connected.
(2) If $\langle x, U, V \rangle$ is a cutting of a Hausdorff continuum, we know that $U \cup \{x\}$ is connected. Is U necessarily connected?
(3) Where does the proof of Proposition 29.2 break down if we weaken the Hausdorff condition to T_1?
(4) Where does the proof of Proposition 29.3 break down if we weaken the Hausdorff condition to T_1?
(5) Prove the "easy" directions of Propositions 29.4 and 29.5.
(6) Prove that indecomposable Hausdorff continua have no disconnecting subcontinua; in particular they can have no cut points.

Lecture 30: Proper Subcontinua

Abstract. The existence of nondegenerate proper subcontinua of nondegenerate continua is not a foregone conclusion, but needs a careful argument to prove. The main result—with the assist of Zorn's lemma—is that any point of a nondegenerate Hausdorff continuum is contained in arbitrarily small nondegenerate subcontinua.

Does a nondegenerate continuum always have a proper nondegenerate subcontinuum? The answer is yes—by definition—if the continuum is decomposable; and lots of continua, even indecomposable ones, contain arcs, so the answer should be an unqualified affirmative. The good news is that our intuitions haven't led us astray; the answer is indeed yes for Hausdorff continua. The bad news is that it is not particularly easy to prove this; indeed there seems to be no way to do it without bringing in Zorn's lemma.

If A and B are subsets of a space X, we say A is *disconnected from* B if there is a disconnection $\{H, K\}$ of X such that $A \subseteq H$ and $B \subseteq K$.

Proposition 30.1 (Cut Wire Theorem). *Let X be a compact Hausdorff space, with A and B two nonempty closed subsets. The following conditions are equivalent:*

(i) *A is disconnected from B.*
(ii) *No connected subset of X intersects both A and B.*

Proof. Any disconnection of A from B will also disconnect any set intersecting both A and B; hence (ii) follows trivially from (i).

For the converse, we prove the special case where $A = \{a\}$ and $B = \{b\}$ are singletons; the general case is left as an exercise.

Suppose that a is not disconnected from b, and let $\mathcal{H}$ consist of all closed subsets H of X such that $\{a, b\} \subseteq H$, but a is not disconnected from b in H. $\mathcal{H}$ is nonempty because $X \in \mathcal{H}$. $\mathcal{H}$ is partially ordered by inclusion; so, as in the proof of Proposition 29.3, we find a maximal nested family $\mathcal{M} \subseteq \mathcal{H}$. By mimicking the argument in the proof of Proposition 27.1, we may infer that $M = \bigcap \mathcal{M}$ is a closed subset of X in which a is not disconnected from b. Hence $M \in \mathcal{H}$ (Exercise 30(2)).

We prove that M is connected, and hence a subcontinuum of X containing both a and b. Suppose not, say $\{P, Q\}$ is a disconnection of M. Since M is closed, both P and Q are closed too. a is not disconnected from b in M; hence both points are in one of the pieces, say $a, b \in P$. P is a proper subset of M, so $P \notin \mathcal{M}$. Since $\mathcal{M} \cup \{P\}$ is nested, the maximality of $\mathcal{M}$ implies that $P \notin \mathcal{H}$. But then there is a disconnection $\{R, S\}$ of P, such that $a \in R$ and $b \in S$. Since P is closed, so are R and S. Thus $\{R, S \cup Q\}$ is a disconnection of a from b in M, a contradiction. This completes the proof.

□

The next theorem tells about components of proper open sets hitting the frontier of those sets.

Proposition 30.2 (Hitting the Frontier Theorem). *Let X be a Hausdorff continuum, with U a proper nonempty open subset of X. Then any component of cl(U) must intersect fr(U) = cl(U) $\cap (X \setminus U)$.*

Proof. Let K be a component of $\mathrm{cl}(U)$, and assume $K \cap \mathrm{fr}(U) = \emptyset$. Since K is a maximally connected subset of $\mathrm{cl}(U)$, there is no subcontinuum of $\mathrm{cl}(U)$ that hits both K and $\mathrm{fr}(U)$. By Proposition 30.1, there is a disconnection $\{M, N\}$ of $\mathrm{cl}(U)$ such that $K \subseteq M$ and $\mathrm{fr}(U) \subseteq N$.

We now let $P = N \cup (X \setminus U)$. Since $\mathrm{cl}(U) = M \cup N$, we have $X = M \cup P$. M contains K and P contains the complement of a proper open set; hence both M and P are nonempty closed subsets of X. To avoid disconnecting the continuum X, though, it must be the case that $M \cap P \neq \emptyset$. But M is disjoint from N, so it must be the case that M intersects $X \setminus U$. But $M \cap (X \setminus U) \subseteq \mathrm{cl}(U) \cap (X \setminus U) = \mathrm{fr}(U) \subseteq N$. This contradicts the fact that M and N are disjoint, and we conclude that K must intersect the frontier of U.

□

As an easy corollary of Proposition 30.2, we have the existence of proper nondegenerate subcontinua of nondegenerate Hausdorff continua—and much more.

Proposition 30.2 (Subcontinuum Existence Theorem). *Let X be a Hausdorff continuum, with C a proper subcontinuum and $U \subseteq X$ an open set containing C. Then there is a subcontinuum K of X, with $C \subseteq K \subseteq U$, and $C \neq K$. In particular, any point of a nondegenerate Hausdorff continuum is contained in arbitrarily small nondegenerate subcontinua.*

Proof. X is connected and C is a proper closed subset, so C is properly contained in U. Hence we may pick $x \in U \setminus C$, with $U' = U \setminus \{x\}$, an open set containing C. Then, using normality in compact Hausdorf spaces, we fix an open set V with $C \subseteq V \subseteq \mathrm{cl}(V) \subseteq U'$. This ensures that $V \neq X$, and we may apply Proposition 30.2: Let K be the component of $\mathrm{cl}(V)$ that contains C. Then we have $C \subseteq K \subseteq U$ and $K \cap (X \setminus V) \neq \emptyset$. But C is contained in V, hence we know that C is properly contained in K.

The second part of the statement follows immediately from the first because singletons are subcontinua.

□

Exercises 30.

(1) Suppose A and B are nonempty closed subsets of a compact Hausdorff space, and that each $a \in A$ is disconnected from each $b \in B$. Show that A is disconnected from B. (Hint: mimic the proof that compact Hausdorff spaces are normal.)

(2) Adapt the proof of Proposition 27.1 to infer that $M \in \mathcal{H}$ in the proof of Proposition 30.1.

(3) A space X is *n-homogeneous*, $n = 1, 2, \ldots$, if whenever A and B are equinumerous subsets of cardinality $\leq n$, there is a homeomorphism $h : X \to X$ such that $h[A] = B$. Justify the following:
   (a) An arc is not 1-homogeneous.

   (b) An irreducible continuum is not 2-homogeneous. (Proposition 30.3 helps. An irreducible continuum can, however, be 1-homogeneous: look up, for example, the dyadic solenoid or the pseudo-arc on Wikipedia.)
   (c) A simple closed curve is n-homogeneous for any $n = 1, 2, \ldots$.
(4) Stronger than merely disconnecting a space X, a subset A of X is *dispersing* if $X \setminus A$ is disconnected, as well as totally disconnected: $X \setminus A$ has more than one component, and each component is a singleton. $a \in X$ is a *dispersion point* if $\{a\}$ is dispersing.
   (a) Show that a connected topological space can have at most one dispersion point. (Hint: Exercise 29(1) may be of use. See Wikipedia for the Knaster-Kuratowski fan, a connected subset of the plane with a dispersion point.)
   (b) Use Proposition 30.3 to show that no subcontinuum of a Hausdorff continuum is dispersing.
(5) Use Proposition 30.3, plus Exercise 29(1), to show that any nondegenerate Hausdorff continuum may be written as the union of two disjoint connected subsets, each containing a nondegenerate subcontinuum.

(6) Use Proposition 30.3, plus Exercise 27(6), to show that the composants of a Housdorff continuum are always dense in the continuum.

Further Reading

Topology is a vast subject, with many branches; we were barely able to scratch the surface in this course. For the reader wishing to delve deeper, here are some texts that will provide ample help.

Fred H. Croom: *Principles of Topology*, 1989.

John G. Hocking & Gail S. Young: *Topology*, 1961.

James R. Munkres: *Topology*, 2000.

Sam B. Nadler, Jr.: *Continuum Theory, an Introduction*, 1992.

George F. Simmons: *Introduction to Topology and Modern Analysis*, 1963.

Lynn A. Steen & J. Arthur Seebach: *Counterexamples in Topology*, 1978.