MATH 171 NOTES

ARUN DEBRAY
JULY 19, 2014

These notes were taken in Stanford's Math 171 class in Spring 2014, taught by Professor Rick Schoen. I live-TEXed them using `vim`, and as such there may be typos; please send questions, comments, complaints, and corrections to `adebray@stanford.edu`. Thanks to Anshul Samar for finding and fixing a few errors.

CONTENTS

Part 1. Real numbers, sequences, limits, series, and functions

1. Sequences and Convergence: 4/1/14

"The hardest thing about [the textbook] is pronouncing the names of its authors."

The course website is `http://math.stanford.edu/~schoen/math171/`. This is a course in mathematical analysis, so in addition to the content of the course, the class also emphasizes the art of writing proofs, and thus we will be nickled-and-dimed in our proofs. Additionally, this is a Writing in the Major course, so the writing assignment will require exposition as well as mathematical content.

The first part of this class will be a quick treatment, ideally review, of properties of the real numbers: sequences, series, continuity, etc. In the textbook, this corresponds to chapters 1 to 6. A useful reference for this part of the class can be found at `http://math.stanford.edu/~schoen/math171/simon.pdf`, though it uses slightly different notation.

One thing we could spend quite some time on, but aren't going to, is the construction of the real numbers. It's a bit tedious and not worth the time, but then there is a uniqueness result given these axioms (which is actually not that hard to prove). Instead, we'll characterize the real numbers (denoted $\mathbb{R}$) axiomatically; there are three sets of axioms. First are the algebraic axioms, which state that the real numbers form a field.

F1. Both addition and multiplication are commutative: for all $a, b \in \mathbb{R}$, $a + b = b + a$ and $ab = ba$.

F2. Both addition and multiplication are associative: for all $a, b, c \in \mathbb{R}$, $(a+b)+c = a+(b+c)$ and $(ab)c = a(bc)$.
F3. There exist identity elements $0, 1 \in \mathbb{R}$ such that $0 \neq 1$ and for all $a \in \mathbb{R}$, $a + 0 = a$ and $a \cdot 1 = a$.
F4. Addition and multiplication distribute: $a(b+c) = ab + ac$ for all $a, b, c \in \mathbb{R}$.
F5. The existence of inverses: for all $a \in \mathbb{R}$, there exists a $-a$ such that $a + (-a) = 0$. If $a \neq 0$, then there also exists a multiplicative inverse a^{-1} such that $aa^{-1} = 1$.

Thus, the real numbers form a group under addition, and the nonzero real numbers form a group under multiplication. There are a lot of sets that satisfy these (all fields); the additional things needed to characterize $\mathbb{R}$ are order and completeness.

Next are the order axioms.

O1. There exists a $P \subset \mathbb{R}$ (the positive real numbers) such that for every $a \in \mathbb{R}$, exactly one of the following is true:
   i. $a \in P$,
   ii. $-a \in P$, or
   iii. $a = 0$.

   In particular, $0 \notin P$, and for any $a \neq 0$, either $a \in P$ or $-a \in P$.
O2. For any $a, b \in P$, $a + b \in P$ and $ab \in P$.

This allows us to define an ordering on the real numbers, in which $a > b$ iff $a - b \in P$. There are still other fields that satisfy all of the above axioms, e.g. the rational numbers.

Finally, the completeness axiom, which distinguishes the reals from the rationals.

S. For any nonempty subset $S \subset \mathbb{R}$ that is bounded above, there exists an $a \in \mathbb{R}$ that is the least upper bound for S.

To be a little more precise about what that means, with the order axioms one can define bounded sets, upper bounds, etc.

Definition.

- x is an upper bound for S if $s \leq x$ for all $s \in S$.
- A set $S \subseteq \mathbb{R}$ is bounded above if it has an upper bound.

One can define bounded below, lower bounds, bounded on both sides, etc. in the reasonable ways.

Definition. a is a least upper bound for S if $a \leq x$ for all upper bounds x of S. This is denoted $\sup S$ or $\operatorname{lub} S$.

It does not follow from the algebraic and order axioms that every bounded set has a least upper bound; for example, consider $\mathbb{Q}$ and the set $S \subset \mathbb{Q}$ given by $S = \{x : x^2 < 2\}$. This is clearly bounded above (e.g. by 2), but it doesn't have a least upper bound (since that would be $\sqrt{2}$, which isn't rational). The completeness axiom in some sense states that the real numbers don't have any holes.

It's not too hard to show that for any set that satisfies all of these axioms, it's equivalent to the real numbers in the sense that there's a one-to-one mapping that preserves all of these properties, so in some sense they're unique. It's a little more tedious, but perfectly possible, to rigorously show that there is a set satisfying these axioms.

Sequences. For some more notation, $\mathbb{N}$ denotes the natural numbers, $\mathbb{N} = \{1, 2, 3, \dots\}$.

Definition. A sequence $\{a_n\}_{n=1}^{\infty}$ (sometimes just denoted $\{a_n\}$) is a function $\mathbb{N} \to \mathbb{R}$, sending $n \mapsto a_n$. These a_n aren't assumed to be distinct (e.g. one could just have $a_n = 1$ for all n).

Definition.

- A sequence $\{a_n\}$ is increasing if $a_{n+1} \geq a_n$ for all $n \in \mathbb{N}$.
- $\{a_n\}$ is decreasing if $a_{n+1} \leq a_n$ for all $n \in \mathbb{N}$.
- $\{a_n\}$ is strictly increasing if $a_{n+1} > a_n$ for all $n \in \mathbb{N}$.
- $\{a_n\}$ is strictly decreasing if $a_{n+1} < a_n$ for all $n \in \mathbb{N}$.
- $\{a_n\}$ is monotone if it is either increasing or decreasing.
- $\{a_n\}$ is bounded above if the set $S = \{a_n : n \in \mathbb{N}\}$ is bounded above, and is bounded below if S is bounded below.
- $\{a_n\}$ is bounded if it's bounded both above and below, i.e. there is an $M \in \mathbb{R}$ such that $|a_n| \leq M$ for all n.

Notice that an increasing sequence is automatically bounded above, and a decreasing sequence is automatically bounded below.

Definition. A sequence $\{a_n\}$ converges if there exists a limit, i.e. an $\ell \in \mathbb{R}$ such that for all $\varepsilon > 0$ there exists an $N \in \mathbb{N}$ (which depends on ε) such that $|a_n - \ell| < \varepsilon$ for all $n \geq N$. In this case, $\{a_n\}$ is called a convergent sequence, and one writes $\lim\limits_{n \to \infty} a_n = \ell$.

One of the basic questions of analysis is whether sequences converge, and there are plenty of theorems which address this issue. The definition might not be obvious, but says that for any given tolerance ε, one should be able to show that the sequence is eventually that close to the limit.

Proposition 1.1. *A bounded monotone sequence converges.*

Proof. Since monotone can mean either increasing or decreasing, assume without loss of generality that $\{a_n\}$ is increasing (if not, then $\{a_n\}$ is decreasing, so $\{-a_n\}$ is increasing and bounded, so if it converges to ℓ, then $\{a_n\}$ converges to $-\ell$).

This theorem wouldn't be true in $\mathbb{Q}$; for example, one could cook up a sequence of rational numbers whose limit is $\sqrt{2}$. Thus, the completeness axiom has to be involved somehow in the proof.

Let $\ell = \sup\{a_n : n \in \mathbb{N}\}$, which exists by the completeness axiom.

Claim. Then, $\lim_{n\to\infty} a_n = \ell$.

Proof. Let $\varepsilon > 0$. Then, $\ell - \varepsilon$ isn't an upper bound for S, so there exists an a_N such that $a_N > \ell - \varepsilon$. But since the sequence is increasing, then for all $n \geq N$, $a_n \geq a_N > \ell - \varepsilon$. But since ℓ is an upper bound, then $a_n \leq \ell < \ell + \varepsilon$. Thus, for all $n \geq N$, $\ell - \varepsilon < a_n < \ell + \varepsilon$, i.e. $|\ell - a_n| < \varepsilon$. ⊠

Thus, by definition, this sequence converges. ⊠

Definition. A subsequence $\{a_{n_i}\}_{i=1}^{\infty}$ of $\{a_n\}$ is a choice of $n_i \in \mathbb{N}$ such that $n_{i+1} > n_i$, so that the subsequence is a choice of some of the elements of the sequence, but still in the original order. For example, one could have $a_2, a_4, a_6, \ldots$, or $a_2, a_3, a_5, a_7, a_{11}, \ldots$, or $a_5, a_{8376}, \ldots$.

A bounded sequence need not converge; for example, if $a_n = (-1)^n$, then it's bounded by -1 and 1, but it doesn't converge. However, the subsequence $a_2, a_4, a_6, \ldots$ converges to 1. This can be generalized:

Theorem 1.2 (Bolzano-Weierstrass)**.** *If $\{a_n\}$ is bounded, then it has a convergent subsequence.*

This motivates two more definitions.

Definition. Suppose $\{a_n\}$ is a bounded sequence.

- $\limsup_{n\to\infty} a_n = \lim_{n\to\infty} A_n$, where $A_n = \sup\{a_k : k \geq n\}$. Notice that A_n is decreasing, so the lim sup always exists. This is also denoted $\overline{\lim}_{n\to\infty} a_n$.
- Similarly, $\liminf_{n\to\infty} a_n = \lim_{n\to\infty} B_n$, where $B_n = \inf\{a_k : k \geq n\}$. By the same reasoning, B_n is increasing and bounded, so this limit also always exists. This is also denoted $\underline{\lim}_{n\to\infty} a_n$.

In general, these limits aren't the same; for example, with $a_n = (-1)^n$, then $\limsup a_n = 1$, but $\liminf a_n = -1$ (since all of the sets $\{a_k : k \geq n\}$ always contain both a -1 and a 1). It will always be true that $\limsup a_n \geq \liminf a_n$.

Proof of Theorem 1.2. The goal will be to show that there exists a subsequence $\{a_{n_i}\}$ that converges to $\ell = \overline{\lim}_{n\to\infty} a_n$. Specifically, the goal is to produce n_i such that $|a_{n_i} - \ell| < 1/i$ for some $i \in \mathbb{N}$. Then, the definition follows, because of the Archimedean property of the real numbers: for any $\varepsilon > 0$, there exists an $i \in \mathbb{N}$ such that $1/i < \varepsilon$. This requires a formal proof; see the textbook.

First, let's find n_1, by looking at $\ell + 1$. There must be some A_N such that $A_N < \ell + 1$ (because otherwise, ℓ isn't an upper bound for $\{a_n\}$); then, pick any element a of this set for a_{n_1} such that $a > \ell - 1$ (which must exist, or ℓ wouldn't be the least upper bound).

Now, the same thing can be done for 2, 3, etc. The only difference is that one requires $A_N < \ell + 1/(i+1)$, and picks an element a such that $a > \ell - 1/(i+1)$, so that it satisfies the bound; the only difference is that one chooses $N > n_i$ in the above, which is fine (since $A_N < \ell + 1/(i+1)$, but this is certainly still true if one increases N).

Then, for each i, $|a_{n_i} - \ell| < 1/i$, which shows that the sequence in question converges. ⊠

One interesting goal is to determine: could I have carried this out? One important goal of math classes is broadening the kinds of arguments you can make independently. In some sense, some kinds of proofs are just moving one's mouth in the right way. . .

Definition. A sequence $\{a_n\}$ is Cauchy (or is a Cauchy sequence) if for all $\varepsilon > 0$ there exists an N such that for all $m, n \geq N$, $|a_n - a_m| < \varepsilon$.

This is of great theoretical importance, and perhaps less important practically.

Theorem 1.3 (Cauchy criterion)**.** *$\{a_n\}$ is convergent iff it is Cauchy.*

Proof. This is a fairly easy consequence of the Bolzano-Weierstrass theorem; the easier direction will be shown today, and the other one next lecture. Today, assume that $\{a_n\}$ converges to ℓ, and the goal is to show that it's Cauchy.

The intuition is that if two things are very close to ℓ (and since it converges, we can make them as close to ℓ as we like), then they must be close to each other. In other words, this is going to be one of those $\varepsilon/2$ proofs.

Specifically, let $\varepsilon > 0$ and choose an N such that $|a_n - \ell| < \varepsilon/2$ for all $n \geq N$. Then, for any $m, n \geq N$, one can use the triangle inequality to show that

$$|a_n - a_m| \leq |a_n - \ell| + |\ell - a_m| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

"What did the mathematician say as $\varepsilon \to 0$?" "There goes the neighborhood!"

Last lecture, we were in the middle of proving that every Cauchy sequence of real numbers converges, and every convergent sequence is Cauchy.

Continuation of the proof. We did the forward direction (that convergence implies Cauchy), so let's assume instead that $\{a_n\}$ is a Cauchy sequence, intending to show that it converges. We'll need to use completeness in some way, which will involve the Bolzano-Weierstrass Theorem; first, we'll show that every Cauchy sequence is bounded, and then use that to obtain a convergent subsequence. But then, this can be lifted to the whole sequence.

Let $\varepsilon = 1$; then, for all $m, n \geq N$, $|a_n - a_m| < 1$, and therefore for all $n \geq N$, $|a_n - a_N| < 1$, and thus that $|a_n| < 1 + |a_N|$ when $n \geq N$, which is because $||a| - |b|| \leq |a - b|$. In particular, for all n, $a_n \leq \max\{|a_1|, \dots, |a_N|, 1 + |a_N|\}$. Thus, $\{a_n\}$ is bounded.[1]

Thus, by Bolzano-Weierstrass, there exists a subsequence $\{a_{n_i}\}$ that converges to a limit ℓ. Now, let $\varepsilon > 0$, so there exists an N and an $n_i \geq N$ such that $|a_{n_i} - \ell| < \varepsilon/2$ and (since $\{a_n\}$ is Cauchy) $\{a_n - a_m\} < \varepsilon/2$ for all $m, n \geq N$. Now, use the Triangle Inequality,

$$|a_n - \ell| \leq |a_n - a_{n_i}| + |a_{n_i} - \ell| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Thus, $\lim_{n\to\infty} a_n = \ell$.[2] ⊠

Infinite Series. The study of infinite series tends to reduce to that of sequences, but there are subtleties.

Definition. Given an infinite sequence $\{a_n\}$, an infinite series is the sum $\sum_{n=0}^{\infty} a_n$. One can define the n^{th} partial sum $s_n = \sum_{i=1}^{n} a_i$; then, a series is said to converge if $\{s_n\}$ converges.

Example 2.1.
- $\sum_{n=0}^{\infty} a^n$ converges iff $|a| < 1$. This is called the geometric series.
- $\sum_{n=1}^{\infty} 1/n^p$ converges iff $p > 1$. When $p = 1$, this is called the harmonic series, and diverges — but very slowly.

Proofs of these are given in the notes or textbook, using the Integral test, etc.

Proposition 2.1. *If $\sum a_n$ converges, then $\lim_{n\to\infty} a_n = 0$.*

Proof. This is because $a_n = s_n - s_{n-1}$, but since s_n converges, then it's Cauchy, so $a_n \to 0$. ⊠

The converse is not true, because $\sum 1/n$ diverges.

Proposition 2.2. *If $\sum a_n = s$ and $\sum b_n = t$, then for any $\alpha, \beta \in \mathbb{R}$,*

$$\sum_{n=0}^{\infty} (\alpha a_n + \beta b_n) = \alpha s + \beta t.$$

The proof is not hard, and uses the corresponding fact from limits.

Proposition 2.3. *If $a_n \geq 0$, then $\sum_{n=1}^{\infty}$ converges iff $S = \{s_n : n = 1, 2, 3, \dots\}$ is bounded above, and in that case, $\sum_{n=1}^{\infty} a_n = \sup S$.*

This is also not too hard to prove, and uses the fact that $s_{n+1} \geq s_n$ and the result about bounded sequences.

Definition. $\sum_{n=1}^{\infty} a_n$ converges absolutely if $\sum_{n=1}^{\infty} |a_n|$ converges. If $\sum a_n$ converges but $\sum |a_n|$ diverges, then it's called conditionally convergent.

It happens that if $\sum a_n$ converges absolutely, then it converges (but not the other way around).

To analyze this, write $a = p_n - q_n$, where p_n is the positive part of a_n and q_n is the negative part:

$$p_n = \begin{cases} a_n, & a_n \geq 0 \\ 0, & a_n < 0 \end{cases} \qquad q_n = \begin{cases} -a_n, & a_n < 0 \\ 0 & a_n \geq 0. \end{cases}$$

Thus, $|a_n| = p_n + q_n$, and $\sum a_n = \sum (p_n - q_n)$, so if $\sum a_n$ converges absolutely, then both $\sum p_n$ and $\sum q_n$ converge, and if $\sum a_n$ is conditionally convergent, then both of them diverge. So in some sense, a conditionally convergent series is of the form "$\infty - \infty$," and rearranging the terms of the series could lead to the difference taking a different form (and even a different limit!). But let's talk about that more formally.

[1]This proof is very general, using only basic properties of the absolute value; it doesn't need the field axioms or completeness.

[2]It's another quite general property that if a Cauchy sequence has a convergent subsequence, then it converges.

Definition. Let $\{a_n\}$ be a sequence and $n \mapsto j_n$ be a one-to-one, onto map $\mathbb{N} \to \mathbb{N}$. Then, $\{a_{j_n}\}$ is called a rearrangement of $\{a_n\}$. Basically, this means the terms of $\{a_n\}$ are shuffled around.

Theorem 2.4. *If $\sum_{n=1}^{\infty} a_n$ is absolutely convergent and $\{a_{j_n}\}$ is a rearrangement of $\{a_n\}$, then $\sum_{n=1}^{\infty} a_{j_n}$ is also absolutely convergent.*

This is distinctly untrue of conditionally convergent series.

Proof of Theorem 2.4. It suffices to assume that $a_n \geq 0$ for all n. Then, we want to show that the sum is independent of any ordering.

Let $S = \sup\{\sum_{n \in F} a_n : F \subseteq \mathbb{N}$ is finite$\}$, the supremum of all finite sums. This supremum exists because if $N = \max F$, then $\sum_{n \in F} a_n \leq s_N \leq \sum_{n=0}^{\infty} a_n$, since the series converges. Then, S is invariant over all rearrangements, so $\sum a_{j_n}$ also converges to S. ⊠

Power Series. A power series is a series of the form $P(x) = \sum_{n=1}^{\infty} a_n x^n$, which may or may not converge.

Theorem 2.5. *For any power series $P(x)$, one of the following three possibilities holds:*

- *$P(x)$ only when $x = 0$.*
- *$P(x)$ converges absolutely for all $x \in \mathbb{R}$.*
- *There exists a $\rho \in \mathbb{R}$, called the radius of convergence, such that $P(x)$ converges absolutely when $|x| < \rho$ and diverges when $|x| > \rho$.*

The proof of this involves comparing a power series to the geometric series. The key insight is that for power series, if $P(x_0)$ converges and $|x| < |x_0|$, then $P(x)$ converges absolutely, because

$$|a_n x^n| = |a_n||x|^n = |a_n||x_0|^n \lambda^n \leq c\lambda^n,$$

where $\lambda = |x|/|x_0| < 1$, so we've bounded $P(x_0)$ by a geometric series which converges (the technical details involve the Cauchy criterion).

Proof of Theorem 2.5. Now, with this result, we can deal with the whole theorem. If $P(x)$ diverges for all $x \neq 0$, then the first case holds, and if $P(x)$ converges for all x, then we must be in case 2, because convergence at a given x implies absolute convergence at all x for which $|x| < |x_0|$.

If neither of these cases holds, then let $\rho = \sup\{|x_0| : P(x_0)$ converges$\}$; then, any x such that $|x| < \rho$ must converge absolutely (since there's an $x_0 > |x|$ such that $P(x_0)$ converges), and if $|x| > \rho$, then it must diverge. ⊠

Continuity of Real Functions.

Definition. $f : [a, b] \to \mathbb{R}$ is continuous at a $c \in [a, b]$ if for all $\varepsilon > 0$ there exists a $\delta > 0$ such that $|f(x) - f(c)| < \varepsilon$ whenever $|x - c| < \delta$.

Then, f is continuous on $[a, b]$ if it's continuous at all $c \in [a, b]$.

Theorem 2.6. *If f is continuous on $[a, b]$, then f is bounded above and below, and achieves its maximum and minimum on $[a, b]$, i.e. there exist m, M such that $m \leq f(x) \leq M$ for all $x \in [a, b]$, and there exist $x_*, x^* \in [a, b]$ such that $f(x_*) = m$ and $f(x^*) = M$.*

This is a basic result from calculus, though there it's usually assumed, rather than formally proven.

Lemma 2.7. *If f is continuous on $[a, b]$, then given $c \in [a, b]$, there exists a $\delta > 0$ such that $|f(x)| \leq |f(c)| + 1$ for all $x \in (c - \delta, c + \delta)$.*

Proof. Use the definition of continuity with $\varepsilon = 1$. ⊠

Notice that if there were a finite number of such intervals $(c - \delta, c + \delta)$, then it would be possible to obtain a single δ for all of $[a, b]$, rather than one that depends on c. There's a result called the Heine-Borel theorem that makes this actually work, and uses completeness (either directly or via Theorem 1.2; it's not true over $\mathbb{Q}$).

Theorem 2.8 (Heine-Borel)**.** *If $[a, b] \subseteq \bigcup_{I \in \mathscr{I}} I$ for a collection $\mathscr{I}$ of open intervals, then there exist $I_1, \ldots, I_n \in \mathscr{I}$ such that $[a, b] \subseteq \bigcup_{i=1}^{n} I_i$.*

This is the basis for an important definition in this course, which is compactness; this theorem states that closed intervals of the real line are compact.

Using this theorem (which will be proven next time), we can prove Theorem 2.6.

3. The Heine-Borel Theorem: 4/8/14

> *"I like to think of the size of a set as its amount of Stanford spirit — you know, its cardinality."* – Mehran Sahami

Last lecture, we stated the Heine-Borel theorem for the real numbers, which stated that if there exists a collection of open intervals $\mathscr{I}$ such that

$$[a,b] \subseteq \bigcup_{I\in\mathscr{I}} I,$$

then there exists a finite set $I_1,\dots,I_n$ such that $[a,b]\subseteq\bigcup_{i=1}^n I_i$. In other words, *every open covering has a finite subcovering*, a concept which we'll return to many times in the course in more general spaces.

This not just follows from the completeness axiom, but is equivalent to it: an ordered field for which this theorem holds is necessarily complete (similarly to how the theorem that every Cauchy sequence converges is equivalent to completeness). In particular, this is not true over $\mathbb{Q}$; for example, $[1,2]$ has a "hole" at $\sqrt{2}$, and one can take open intervals that get closer and closer to it on both sides, so any finite subcollection of them will be some distance away from it, and thus missing some rational numbers, but the infinite collection has everything. Thus, this theorem in some sense says that the real numbers don't have any holes.

Proof of Theorem 2.8. Let $S=\{c\in[a,b]:[a,c]\text{ can be finitely covered}\}$. We can cover $\{a\}$ in a single covering, and therefore can go some distance farther, so there exists a $c_0>a$ such that $c_0\in S$. Moreover, S is bounded above by b, so it has a supremum $c^*\le b$.

The theorem is shown if we can demonstrate that $c^*\in S$ and $c^*=b$; these together imply that $[a,b]$ can be finitely covered.

- Since $c^*\in[a,b]$, then there exists an $I\in\mathscr{I}$ containing c^*, so there exists a $c<c^*$ with $c\in I$. Thus, c is not an upper bound for S, and there exists a $c_0\in S$ such that $c_0>c$, so $[a,c_0]\subseteq I_1\cup\cdots\cup I_n\cup I$ is finitely covered, so $c^*\in S$. In particular, S contains its upper bound.
- If $c^*<b$, then $[a,c^*]\subseteq\bigcup_{i=1}^n I_i$, Thus, $c^*\in I_j$ for some j, and since the interval is open, there exists a $c>c^*$ such that $c\in I_j$, and therefore $c\in S$. This contradicts c^* being an upper bound, so $c^*=b$. ⊠

This is very useful for extracting facts in other contexts, as in the proof of Theorem 2.6.

Proof of Theorem 2.6. To show that f is bounded above and below and attains its minimum and maximum on $[a,b]$, it suffices to show that it's bounded above and attains its maximum, and then repeat the proof with $-f$ to show the remaining parts.

First, it will be necessary to show that f is bounded above. For any $x\in[a,b]$, there exists a $\delta_x>0$ such that $|f(y)-f(x)|<1$ if $|y-x|<\delta_x$. Thus, letting $I_x=(x-\delta_x,x+\delta_x)$, $f(y)\le f(x)+1$.

Let $\mathscr{X}=\{I_x:x\in[a,b]\}$, so that

$$[a,b]\subseteq\bigcup\mathscr{X}=\bigcup_{x\in[a,b]} I_x.$$

By the Heine-Borel theorem, this means there's a finite subcovering: there exists a finite set $x_1,\dots,x_n$ such that $[a,b]\subseteq\bigcup_{i=1}^n I_{x_i}$. But f is bounded on each interval, so for all $y\in[a,b]$, $f(y)\le\max_{1\le i\le n}(f(x_i)+1)$.

Now, we need to show that it achieves its maximum. Let $M=\sup\{f(x):x\le[a,b]\}$, which we now know is bounded above. Thus, it remains to show there is some $x_0\in[a,b]$ such that $f(x_0)=M$. Let $g(x)=1/(M-f(x))$, so that g is continuous on $[a,b]$ whenever $f(x)<M$. Thus, if f doesn't achieve its maximum, then g is continuous everywhere on $[a,b]$, but g isn't bounded, because M is the least upper bound, because the denominator can be made less than ε for any $\varepsilon>0$, but we saw that any function continuous on the whole interval must be bounded. Thus, there does exist an $x_0\in[a,b]$ such that $f(x_0)=M$. ⊠

These theorems are intimately related to the completeness axiom, e.g. if one takes $f:\mathbb{Q}\to\mathbb{Q}$ given by $f(x)=1/(2-x^2)$, then f is continuous at every rational point, but not bounded, because $\sqrt{2}\notin\mathbb{Q}$, so the denominator is as close to 0 as one wants, but never zero.

Cardinality of Sets. It turns out that the distinction between sets being countable and uncountable is very important for some theorems in analysis. Thus, we'll cover[3] just a little bit of set theory. In particular, we want to be able to say that two sets have the same cardinality.

If A is a finite set, then its cardinality $\#(A)$ is the number of elements it has.

Claim. Two finite sets A and B have the same cardinality if there is a map $f:A\to B$ that is a one-to-one correspondence (i.e. one-to-one and onto).

[3] No pun intended.

This is because such a function f preserves the number of elements: every $a \in A$ is mapped to exactly one element of B, and every $b \in B$ is the image of exactly one $a \in A$.

Then, one can expand this definition, saying that any two sets have the same cardinality if they have a one-to-one correspondence.

Definition.

- A is countably infinite if it has the same cardinality as $\mathbb{N} = \{1, 2, 3 \dots\}$.
- A is countable if it is finite or countably infinite.
- If A is not countable, it is said to be uncountable.

The elements of a countable set can in some sense be put into an index. This is formalized by the following claim:

Claim. A is countable iff there is an onto map $f : \mathbb{N} \to A$.

Proof. In the forward direction, suppose A is countable. If A is countably infinite, then this follows by definition, but if A is finite, then label its elements $A = \{a_1, \dots, a_n\}$, and define $f : \mathbb{N} \to A$ as

$$f(i) = \begin{cases} a_i, & \text{if } 1 \le i \le n \\ a_n, & \text{if } i > n. \end{cases}$$

This choice is arbitrary (as with many choices in set theory), but the point is that it works: f is onto.

The other direction is similarly straightforward: suppose $f : \mathbb{N} \to A$ is onto. If A is finite, then we're done, so assume that A is infinite. Then, define $g : \mathbb{N} \to A$ as follows:

- Let $g(1) = f(1)$.
- Let $g(2) = f(n_2)$, where $n_2 = \min\{i \in \mathbb{N} : f(i) \ne f(1)\}$.
- And so on: for each k, let $g(k) = f(n_k)$, where $n_k = \min\{i \in \mathbb{N} : f(i) \ne g(j) \text{ for all } j < k\}$.

In essence, this defined a subsequence of $\{f(n)\}$ that doesn't repeat itself.

The minimum exists in each case because f is onto the infinite set A, so its range cannot be any finite set of points; thus, there's a minimum value of the nonempty set of natural numbers which haven't been hit by f yet.

Thus, g is defined on all of $\mathbb{N}$, and it's onto (because f is) and one-to-one (because by construction, no two points take the same value). ⊠

This makes sense, because it says that the domain has at least as many points as the image. Of course, if A is a finite set, the sequence will repeat a lot.

Here are some examples of countable sets:

- $\mathbb{Z}$ is countable, given by the function

  $$f(n) = \begin{cases} -k, & \text{if } n = 2k \text{ for some } k \in \mathbb{N} \\ \ell, & \text{if } n = 2\ell - 1 \text{ for some } \ell \in \mathbb{N}, \end{cases}$$

  which is one-to-one and onto.
- $\mathbb{N} \times \mathbb{N} = \{(i, j) : i, j \in \mathbb{N}\}$. Geometrically, this is every point in the first quadrant with integer-valued coordinates, and thus they can be written in a bunch of diagonals (the lines $x + y = k$ for $k \in \mathbb{N}$), each with a finite number of elements. Then, count the elements in the diagonals in order of increasing k, as in Figure 1. In particular, we can explicitly write $f^{-1}(i, j) = (i + j + 2)(i + j - 1)/2$.
- Suppose that for $n \in \mathbb{N}$ there is a countable A_n. Then, $\bigcup_{n=1}^{\infty} A_n$ is countable. This follows from the above, because one can index each A_i as $\{a_{ij}\}$ for $j \in \mathbb{N}$, because it's countable, so we obtain an onto map $(i, j) \mapsto a_{ij}$ from $\mathbb{N} \times \mathbb{N} \to \bigcup_{n=1}^{\infty} A_n$; thus, since the former is countable, then so is the latter.
- $\mathbb{Q}$ is countable, because $\mathbb{Z} \times \mathbb{Z}$ is (following from the third point), and all rational numbers can be written as quotients of integers, and therefore can be represented as pairs of integers: $(i, j) \mapsto i/j$ is the general idea (some noodling around has to be done to take care of the case $j = 0$), which is onto $\mathbb{Q}$; thus, $\mathbb{Q}$ is countable.

Part 2. Metric Spaces

4. Countability and Metric Spaces: 4/10/14

> *"Not all things worth counting are countable and not all things that count are worth counting." – Albert Einstein*

Last time we defined a set A to be countable if A is finite or countably infinite (i.e. in bijection with $\mathbb{N}$), and provided a useful characterization: that A is countable iff there is an $f : \mathbb{N} \to A$ that is onto. Two useful corollaries follow:

Corollary 4.1. *If A is countable and $B \subset A$, then B is countable.*

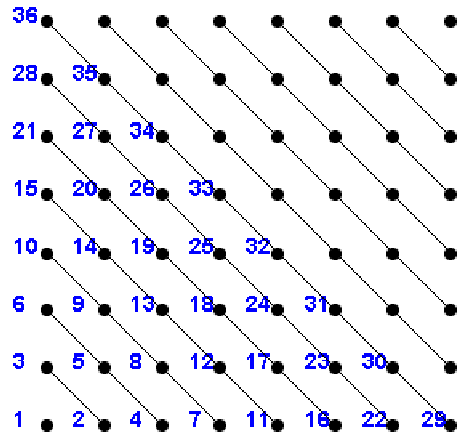

FIGURE 1. $\mathbb{N} \times \mathbb{N}$ is countable. Source: `http://ocw.mit.edu/ans7870/18/18.013a/textbook/HTML/index.html`.

Corollary 4.2. *If A and B are countable, then so is $A \times B = \{(a, b) : a \in A, b \in B\}$ (the Cartesian product).*

Proof of Corollary 4.1. Since A is countable, there exists an onto $f : \mathbb{N} \to A$, and we want to construct a $g : \mathbb{N} \to B$ that is onto. This can be given as follows: pick some $b_0 \in B$ (unless B is empty, in which case it's trivially countable), and let

$$g(n) = \begin{cases} f(n), & f(n) \in B \\ b_0, & f(n) \notin B. \end{cases}$$

This is onto, because any $b \in B$ is in A, so it's $f(n)$ for some $n \in \mathbb{N}$, but then $g(n) = f(n) = b$. ⊠

Proof of Corollary 4.2. This will end up using something like what we did last time. Since A and B are countable, then there exist onto maps $f : \mathbb{N} \to A$ and $g : \mathbb{N} \to B$, so there is a function $(f \times g) : \mathbb{N} \times \mathbb{N} \to A \times B$ given by $(f \times g)(a, b) = (f(a), g(b))$. $f \times g$ is onto, because for any $a \in A$ and $b \in B$, $a = f(m)$ and $b = g(n)$ for some $m, n \in \mathbb{N}$, so $(f \times g)(m, n) = (a, b)$.

Since $\mathbb{N} \times \mathbb{N}$ is countable, as shown last time, then there exists some onto map $\mathbb{N} \to \mathbb{N} \times \mathbb{N}$, so taking the composition of this map with $f \times g$ provides an onto function $\mathbb{N} \to A \times B$, showing it's countable. ⊠

We also claimed last lecture that the countable union of countable sets is countable, i.e. if $A_1, A_2, \ldots$ is indexed by $\mathbb{N}$ and each A_n is countable, then so is $A = \bigcup_{n=1}^{\infty} A_n$. Since A_n is countable, then its elements can be put into a sequence $a_n^{(i)}$, where each $a \in A$ is an $a_n^{(i)}$ for some i. Then, define $F : \mathbb{N} \times \mathbb{N} \to A$ by $F(n, i) = a_n^{(i)}$, so F is onto (since every element of the union is in A_n for some n, and every element of A_n is an $a_n^{(i)}$ for some i), and $\mathbb{N} \times \mathbb{N}$ is countable, so there's a $G : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ that's onto, so $F \circ G : \mathbb{N} \to A$ is onto, and thus A is countable.

Notice that this involved taking an infinite union of sets. This is perfectly reasonable: if one has any index set $\mathscr{I}$ (of any cardinality) such that for each $i \in \mathscr{I}$ there's some set A_i, we can construct the union $\bigcup_{i \in \mathscr{I}} A_i$. This is simply the set of things that lies within at least one of the sets we're considering.

The result about Cartesian products generalizes to all finite products of countable sets, but it doesn't generalize to the countable product of $\mathbb{N}$ with itself: this has the cardinality of $\mathbb{R}$.

Theorem 4.3. *If $a, b \in \mathbb{R}$ and $a < b$, then $[a, b]$ is uncountable.*

Notice that $\mathbb{Q}$ is countable, yet it's dense in $\mathbb{R}$. Thus, there are in some sense many, many irrational numbers for every rational number.

Proof of Theorem 4.3. Let $x_1, x_2 \ldots$ be a sequence of elements of $[a, b]$. Now, divide $[a, b]$ into three equal parts I_1, I_2, and I_3, all of which are closed intervals (so that, for example, $I_1 \cap I_2$ is just their common endpoint). Thus, $x_1 \notin J_1$, where J_1 is one of I_1, I_2, or I_3. Now, repeat the process with J_1: divide it into three intervals. Then, one of them, which will be called J_2, is such that $x_2 \notin J_2$. Thus, $x_1, x_2 \notin J_2$, since $J_2 \subset J_1$.

Then, split J_2 into thirds... inductively construct a sequence of closed intervals $J_1 \supseteq J_2 \supseteq J_3 \supseteq \cdots$, such that the length of J_n is $\ell(J_n) = (b - a)/3^n$, and $x_1, \ldots, x_n \notin J_n$. (The explicit inductive procedure is as in the case for J_2, but isn't written here because it's basically the same). Thus, by the Nested Interval theorem on the homework, $\bigcap_{n=1}^{\infty} J_n$ is nonempty, so pick an x in this intersection.[4] Then, $x \neq x_m$ for all m, so $[a, b]$ isn't countable. ⊠

[4]Since the lengths of the J_n shrink to 0, then there is exactly one such x, in fact.

Notice how heavily this proof (particularly the invocation of the Nested Interval theorem) relies on the completeness properties of $\mathbb{R}$. There are other proof of this, including a more famous one about decimal expansions.

At this point, the "review" portion of the class is completed.

Metric Spaces. So far, everything we've done has dealt with the real line, and now, we can extend that to more complicated spaces. We'll start with metric spaces, which have an abstract definition but many more concrete examples.

Definition. A metric[5] for a set M is a function $d : M \times M \to [0, \infty)$ such that:

- $d(x, y) \geq 0$ for all $x, y \in M$, and $d(x, y) = 0$ iff $x = y$. In particular, $d(x, x) = 0$, and if $x \neq y$, then $d(x, y) > 0$. This is known as the positivity property.
- d is symmetric: $d(x, y) = d(y, x)$.
- The crucial property for analysis is the triangle inequality: that for any $x, y, z \in M$, $d(x, z) \leq d(x, y) + d(y, z)$.

Definition. A metric space is a pair (M, d), where d is a distance function on a set M.

This is an extremely abstract notion: there are metrics on many kinds of objects.

Example 4.1.

- $(\mathbb{R}, d)$, where $d(x, y) = |x - y|$. The triangle inequality is easily verified.
- $(\mathbb{R}^n, d)$, with the Euclidean distance
$$d(x, y) = \left(\sum_{i=1}^{n} (x_i - y_i)^2\right)^{1/2}.$$
This is also called the ℓ^2 distance on $\mathbb{R}^n$.
- The space ℓ^1 of absolutely summable sequences, $\ell^1 = \{a = \{a_n\}_{n=1}^{\infty} : \{a_n\}$ is a real sequence such that $\sum_{n=1}^{\infty} |a_n|$ is finite$\}$, with the ℓ^1 distance
$$d(a, b) = \sum_{n=1}^{\infty} |a_n - b_n|.$$
Notice that this sum is finite, because those for a and b are.

There will be more examples later. All of these are real vector spaces (e.g. one can add sequences in ℓ^1, and their sum converge, and same with scalar multiplication), which allows them to be generalized to something called a normed vector space.

Definition. If V is a real vector space, a function $v \mapsto \|v\|$ is a norm if it satisfies the following axioms:

(1) $\|v\| \geq 0$ for all $v \in V$ and $\|v\| = 0$ iff $v = 0$.
(2) $\|\lambda v\| = |\lambda| \|v\|$ for all $\lambda \in \mathbb{R}$ and $v \in V$.
(3) $\|v + w\| \leq \|v\| + \|w\|$ for all $v, w \in V$.

Notice that all of the examples in Example 4.1 are normed spaces: $\|x\| = |x|$ is a norm on $\mathbb{R}$, and $\|x\| = \left(\sum x_i^2\right)^{1/2}$ is a norm on $\mathbb{R}^n$. Finally, on ℓ^1, the ℓ^1 norm is given by
$$\|a\| = \sum_{n=1}^{\infty} |a_n|.$$

Claim. If V is a vector space and $\|\cdot\|$ is a norm on V, then $d(v, w) = \|v - w\|$ turns V into a metric space.

Proof. Positivity is satisfied because $\|v\| = 0$ iff $v = 0$, so $d(x, y) = 0$ iff $x - y = 0$, i.e. $x = y$, and symmetry of d follows from the second property (with $\lambda = -1$). Finally, for the triangle inequality, if $v_1, v_2, v_3 \in V$, then
$$\begin{aligned} d(v_1, v_3) &= \|v_3 - v_1\| = \|(v_3 - v_2) + (v_2 - v_1)\| \\ &\leq \|v_3 - v_2\| + \|v_2 - v_1\| \\ &= d(v_1, v_2) + d(v_2, v_3). \end{aligned}$$
⊠

Normed vector spaces are very special among metric spaces.

Example 4.2. For an example of a metric space that's not a normed space, let $\Gamma \subseteq \mathbb{R}^2$ be a curve (that is, the set of points on the curve), and d be the distance function on $\mathbb{R}^2$. Then, for any $p, q \in \Gamma$, let the distance in Γ is just $d(p, q)$ (the chord distance), and (Γ, d) is a metric space because $(\mathbb{R}^2, d)$ is. But it's not a vector space.

---

[5] This is technically an ambiguous term: in geometry, there's a related, but distinct, function also called a metric.

This suggests a much more general construction: if (M, d) is a metric space and $N \subseteq M$, then (N, d) is also a metric space. This suggests a few deep questions, such as whether every metric space can be embedded in a Euclidean space.

Example 4.3. If M is any nonempty set, the discrete metric on M is the distance function

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

This isn't a very interesting metric, but it is in fact a metric.

Example 4.4. If p is a prime number, then there's a norm on $\mathbb{Q}$ called the p-adic norm measuring how divisible things are by p: for any $x \in \mathbb{Q}$, write $x = p^n(a/b)$ for $a, b \in \mathbb{Z}$ that aren't divisible by p (so that $n \in \mathbb{Z}$); then, define the p-adic norm to be $\|x\|_p = p^{-n}$, and define $\|0\|_p = 0$, so that things that are highly divisible by p are near 0.

This defines a norm on $\mathbb{Q}$, though since $\mathbb{Q}$ isn't a real vector space, then the definition of a norm has to be taken slightly differently. Specifically, for the second property (for how the norm interacts with scalar multiplication), one needs to take $\|\lambda x\|_p = \|\lambda\|_p \|x\|_p$.

But the triangle inequality is even more fun, because the p-adic norm induces something called an ultrametric, defined by a norm satisfying something called the ultra-triangle inequality:[6]

$$\|x + y\|_p \leq \max\{\|x\|_p, \|y\|_p\}.$$

The p-adics aren't really the point of the class, but the point is to show that metrics show up across mathematics, e.g. within number theory for this example. This begins the field called p-adic analysis.

Example 4.5. Generalizing $\mathbb{R}^n$ with its standard distance to infinite dimensions, one obtains the set of square-summable sequences

$$\ell^2 = \left\{ a = \{a_n\}_{n=1}^{\infty} : \sum_{n=1}^{\infty} a_n^2 \text{ is finite} \right\}.$$

This has the Euclidean norm $\|a\| = \left(\sum_{n=1}^{\infty} a_n^2\right)^{1/2}$, just like $\mathbb{R}^n$.

This is actually an example of an inner product space, along with $\mathbb{R}$ and $\mathbb{R}^n$, meaning that the norm is given by an inner product in the following sense.

Definition. If V is a real vector space, a function $\langle v, w \rangle \in \mathbb{R}$ for $v, w \in V$ is an inner product if:

- $\langle v, v \rangle \geq 0$ for all $v \in V$ and is equal to 0 iff $v = 0$.
- $\langle v, w \rangle = \langle w, v \rangle$.
- It's bilinear: $\langle c_1 v_1 + c_2 v_2, w \rangle = c_1 \langle v_1, w \rangle + c_2 \langle v_2, w \rangle$ (and thus this is also true in the w slot).

It turns out that inner products define norms (and therefore metric spaces); the instructive example is $\mathbb{R}^n$.

Definition. If V is a vector space and $\langle \cdot, \cdot \rangle : V \to \mathbb{R}$ is an inner product on V, then $(V, \langle \cdot, \cdot \rangle)$ is called an inner product space.

Then, just like in $\mathbb{R}^n$, one obtains a norm $\|v\| = \sqrt{\langle v, v \rangle}$ on any inner product space. But to show this, we need a result called the Cauchy-Schwarz inequality (which has a simple, but beautiful proof in any inner product space), which allows one to define angles between vectors (since it looks very similar to the formula for angles in $\mathbb{R}^n$). The inequality states that

$$|\langle v, w \rangle| \leq \|v\| \|w\|,$$

with equality iff v and w are linearly dependent. The proof will be deferred to next lecture, but this inequality implies the triangle inequality (which is probably the hardest part of showing that inner products induce metrics): in an inner product space,

$$\begin{aligned} \|v + w\|^2 &= \langle v + w, v + w \rangle \\ &= \|v\|^2 + 2\langle v, w \rangle + \|w\|^2 \\ &\leq \|v\|^2 + 2\|v\|\|w\| + \|w\|^2 = (\|v\| + \|w\|)^2. \end{aligned}$$

[6] The "ultra-triangle inequality" sounds like a supervillain's doomsday device, doesn't it?

"But what units are we measuring the distance in?"
"Metric, of course."

Recall that last time, we defined a metric space as a pair (M, d), where d is a distance function on M, i.e. $d : M \times M \to [0, \infty)$ with the three properties of positivity (two points are distinct iff they have a strictly positive distance), symmetry ($d(x, y) = d(y, x)$), and the triangle inequality.

We also introduced the concept of a norm on a vector space (§69 of the textbook, where it's called instead a normed linear space), a function $\|\cdot\| : V \to \mathbb{R}$ that has a similar positivity axiom as well as a triangle inequality, but is also homogeneous (i.e. is compatible with multiplication with scalars). Normed spaces induce metric spaces: $d(x, y) = \|x - y\|$.

Example 5.1.

- The standard Euclidean norm, also called the L^2 norm, on $\mathbb{R}^n$ is given by
$$\|x\|_2 = \left(\sum_{i=1}^n x_i^2\right)^{1/2}.$$
- On $\mathbb{R}^n$ one also has the L^1 norm
$$\|x\|_1 = \sum_{i=1}^n |x_i|.$$
- Finally, the following is a norm, which is a little harder to show: the L^∞ norm is $\|x\|_\infty = \max\{|x_i| : 1 \le i \le n\}$.

We also saw the definition of an inner product space (§71 of the textbook): a vector space V with an inner product $(\cdot, \cdot)$ that has the same positivity and symmetry conditions, and is bilinear (i.e. linear in both arguments). In general, an inner product space is a normed space, with norm $\|v\| = \sqrt{\langle v, v \rangle}$. The proof we gave assumed the Cauchy-Schwarz inequality, which we'll get to in a moment. Here are some examples of normed spaces.

Example 5.2.

- On $\mathbb{R}^n$, let $(x, y) = x \cdot y$. This induces the L^2 norm above.
- Let ℓ^2 be the space of sequences $\{a_n\}$ such that $\sum a_n^2$ (called square-summable sequences). Then, it has an inner product (which once again requires the Cauchy-Schwarz inequality to check well-definedness)
$$(x, y) = \sum_{n=1}^\infty x_n y_n.$$
- If f and g are continuous functions on $[a, b]$, the L^2 norm on the space of continuous functions is
$$(f, g) = \int_a^b f(x) g(x)\,\mathrm{d}x.$$

Theorem 5.1 (Cauchy-Schwarz Inequality). *Assume $(V, (\cdot, \cdot))$ is an inner product space. Then, $|(v, w)| \le \|v\|\|w\|$ and equality holds iff v and w are linearly dependent (i.e. one is a multiple of the other).*

Proof. The proof follows directly from the properties of the inner product; first note that if $v = 0$ then the entire thing is obvious, so we can assume that $v \neq 0$. Now, define $f(t) = \|tv + w\|^2$, and expand out as
$$\begin{aligned} f(t) &= \|tv + w\|^2 = (tv + w, tv + w) \\ &= t(v, tv + w) + w(tv + w) \\ &= t^2\|v\|^2 + 2(v, w) + \|w\|^2. \end{aligned}$$
Since $v \neq 0$, then this is a nondegenerate quadratic polynomial, so by the positivity of the inner product and the norm, then it's nonnegative, so it has at most one root. Thus, the discriminant is negative:
$$(2(v, w))^2 - 4\|v\|^2\|w\|^2 \le 0,$$
and it's equal to 0 iff there is a real zero t_0, i.e. such that $f(t_0) = 0$. Thus, $(v, w)^2 \le \|v\|^2\|w\|^2$, so $|(v, w)| \le \|v\|\|w\|$. If there is a zero, then $\|t_0 v + w\|^2 = 0$, so by the positivity of the inner product, $t_0 v + w = 0$, so v and w must be linearly dependent. ⊠

Armed with this result, we can prove some things about the examples.

Claim. If $x, y \in \ell^2$, then $\sum x_n y_n$ is absolutely convergent.

Proof. for any $k \in \mathbb{N}$, let $\mathbf{x} = (|x_1|, \cdots, |x_k|)$ and $\mathbf{y} = (|y_1|, \dots, |y_k|)$, so that

$$\mathbf{x} \cdot \mathbf{y} = \sum_{n=1}^{k} |x_n||y_n| \leq \left(\sum_{n=1}^{k} x_n^2\right)^{1/2} \left(\sum_{n=1}^{k} y_n^2\right)^{1/2},$$

but since the norm is positive, this is less than $\|x\|\|y\|$, and thus the partial sums are bounded, so the series converges. ⊠

Sequences in Metric Spaces. Many of the ideas previously seen before in analysis over the real numbers can be extended to metric spaces. So for the next few weeks, let M be a metric space, and let $\{x_n\}_{n=1}^{\infty}$, with $x_n \in M$, be a sequence from M.

Definition. $\{x_n\}$ converges to a limit $L \in M$ if for all $\varepsilon > 0$ there exists an N such that $d(x_n, L) < \varepsilon$ for all $n \geq N$.

The only difference is the replacement of the absolute value with the distance function, but because this is so much more general, different choices of metric mean different things for convergence.

There are a bunch of basic properties of sequences, which are as easy to verify here as they are over $\mathbb{R}$.

Proposition 5.2. *Limits are unique: if* $\lim_{n\to\infty} x_n = L$ *and* $\lim_{n\to\infty} = L'$*, then* $L = L'$.

This will use the positivity of the metric, along with the triangle inequality.

Definition. If $x_0 \in M$, then the open ball of radius a at x_0 is $B_a(x_0) = \{x \in M : d(x, x_0) < a\}$.

These are typically thought of as spheres, which is reasonable enough even if some metrics look quite different!

Proposition 5.3. *Suppose* $x, y \in M$ *and* $x \neq y$*. Then, there exist* $r_1, r_2 \in \mathbb{R}$ *such that* $B_{r_1}(x) \cap B_{r_2}(y) = \emptyset$.

That is, any two distinct points can be separated by open balls.

Proof of Proposition 5.3. Let $r = d(x, y)$ and $r_1 = r_2 = r/2$, and suppose there is some $z \in B_{r_1}(x) \cap B_{r_2}(y)$. Then, $d(x,z) < r/2$ and $d(y,z) < r/2$, so $d(x,z) + d(z,y) < r = d(x,y)$, which contradicts the triangle inequality. ⊠

This will be very useful in proving the uniqueness of limits.

Proof of Proposition 5.2. Let $\varepsilon_0 = d(L, L')$ and if $\varepsilon_0 > 0$, then let $\varepsilon = \varepsilon_0/2$. Then, there exists an N such that $d(x_n, L) < \varepsilon$ if $n \geq N$, and $d(x_n, L') < \varepsilon$ if $n \geq N$, so $x_n \in B_{\varepsilon/2}(L) \cap B_{\varepsilon/2}(L')$, so L and L' can't be separated by open balls and thus must be the same. ⊠

Subsequences are defined in the same way as in the real numbers.

Claim. If $\{x_n\}$ converges to L, then any subsequence $\{x_{n_i}\}_{i=1}^{\infty}$ also converges to L.

Theorem 5.4. *If one considers* $\mathbb{R}^n$ *with the Euclidean distance,*[7] *let* $\{x^{(i)}\}_{i=1}^{\infty}$ *(i.e.* $x^{(i)} \in \mathbb{R}^n$*) be a sequence, where the* j^{th} *component of* $x^{(i)}$ *is* $x_j^{(i)}$*. Then,* $\lim_{i\to\infty} x^{(i)} = x$ *iff* $\lim_{j\to\infty} x_j^{(i)} = x_j$.

Proof. In the forward direction, assume that $\lim_{i\to\infty} x^{(i)} = x$. Then, let $\varepsilon > 0$, so there exists an N such that $\|x^{(i)} - x\| < \varepsilon$ for all $i \geq N$. But

$$|x_j^{(i)} - x_j| \leq \left(\sum_{k=1}^{n} |x_k^{(i)} - x_k|^2\right)^{1/2} = \|x^{(i)} - x\| < \varepsilon,$$

so $\lim_{i\to\infty} x_j^{(i)} = x_j$.

The reverse direction isn't hard to prove, but use the fact that n is finite more subtly (it's completely untrue in ℓ^2, for example), so assume that $\lim_{i\to\infty} x_j^{(i)} = x_j$ for $j = 1, \dots, n$. Lots of things in analysis reduce to some basic inequality, and here the one is (from the definition of the norm)

$$\|x^{(i)} - x\| \leq \sqrt{n} \max\{|x_j^{(i)} - x_j| : j = 1, \dots, n\},$$

and there exists an N such that $|x_j^{(i)} - x_j| < \varepsilon/\sqrt{n}$ for all $i \geq N$ and $j = 1, \dots, n$ (take the maximum of the ones that work for each component), so $\|x^{(i)} - x\| < \varepsilon$. ⊠

This says that convergence on $\mathbb{R}^n$ is easy to understand in terms of convergence of real sequences.

---

[7] A slight modification of this argument works for the L^1 and L^∞ norms.

Example 5.3. To see how this fails on infinite-dimensional spaces, in ℓ^1 (the set of absolutely convergent sequences) let $x^{(1)} = (1,0,0,\dots)$, $x^{(2)} = (0,1,0,0,\dots)$, and so on. More compactly,[8] the i^{th} component of $x^{(n)}$ is $x_i^{(n)} = \delta_{in}$ (the Kronecker delta, equal to 1 if $i = n$ and 0 otherwise).

Thus, $\lim_{n\to\infty} x_i^{(n)} = 0$ for all i, because $x_i^{(n)} = 0$ for all sufficiently large i (i.e. $i > n$). But

$$d(x^{(n)}, 0) = \sum_{i=1}^{\infty} |x_i^{(n)}| = 1,$$

so the sequence cannot converge to 0 (and in fact doesn't converge). However, the first part of the theorem didn't depend on n being finite; thus, in ℓ^1 as well as in $\mathbb{R}^n$, if $\lim_{n\to\infty} x^{(n)} = x$, then $\lim x_i^{(n)} = x_i$, a fact which relies on the basic inequality $|x_i^{(n)} - x_i| \le \|x^{(n)} - x\|$ (since the sum of a bunch of absolute values is greater than any individual term).

Note that infinite series don't make sense in general metric spaces, since we have no notion of addition. (They do work in normed spaces, but we won't really worry about that here.)

Closed and Open Sets. Once again, (M,d) is a metric space, and suppose $X \subseteq M$.

Definition. $x \in M$ is a limit point of X if there exists a sequence $\{x_n\} \subseteq X$ such that $\lim_{n\to\infty} x_n = x$.

A set doesn't necessarily contain its limit points; for example, a and b are limit points of (a,b) in $\mathbb{R}$.

Definition. The set $\overline{X}$ of the limit points of X is called the closure of X.[9] If $X = \overline{X}$, then X is said to be closed.

Notice that for all sets, $X \subseteq \overline{X}$.

For example, in $\mathbb{R}$, closed intervals are closed sets, and in $\mathbb{R}^2$, a set is closed if it contains its boundary sets. This can help with intuition.

Theorem 5.5. *M, $\emptyset$, and $\{x\}$ for any $x \in X$ are all closed sets.*

Proof. For M this is obvious, for all convergent sequences on M can't converge to anywhere else. For $\emptyset$, it's a logical exercise: there are no sequences on $\emptyset$. And for the singleton set, the only sequence is $x_n = x$, which converges to x, so $\{x\}$ is closed. ⊠

Theorem 5.6.

(1) A finite union of closed sets is closed.
(2) Any intersection of closed sets is closed.

Note that the finiteness condition for part (1) is necessary, because $[1/n, 1-1/n]$ is closed for all n, but

$$\bigcup_{n=3}^{\infty} \left[\frac{1}{n}, 1 - \frac{1}{n}\right] = (0,1),$$

which isn't closed (it doesn't contain its limit points 0 and 1).

Proof of Theorem 5.6. For part (1), let $F_1, \dots, F_n$ be closed sets, and let $F = \bigcup_{i=1}^n F_i$, so the goal is to show that F is closed. Take a sequence $\{x_j\} \subseteq F$ such that $\lim_{j\to\infty} x_j = x$. Then, since there are a finite number of F_i, then there is some F_i that contain an infinite number of x_j, so there is a subsequence $\{x_{j_k}\}_{k=1}^{\infty} \subseteq F_i$, but this converges to the same value x. Since F_i is closed, then $x \in F_i$ and therefore in F as well.

(2) is easy from this definition of closed: if one takes a sequence that lies in an intersection, then it lies in all of the sets, so the limit must be in each set as well (since they're closed), and therefore in the intersection! ⊠

6. Open Sets and Closed Sets: 4/17/14

"Sets, like doors, can be open, closed, or neither. We can call a set that is neither open nor closed a jar."

Last time, we defined closed sets in a metric space, and saw that if $\mathscr{F}$ is a collection of closed subsets of M, then $\mathscr{X} = \bigcap_{F\in\mathscr{F}} F$ is also closed, because any convergent sequence $\{x_n\}$ in all of the F must have its limit x in each of the F (since each F is closed), so $x \in \mathscr{X}$.

This implies the following result.

Corollary 6.1. *The closure of any set A satisfies $\overline{A} = \bigcap \mathscr{F}$, where $\mathscr{F} = \{F : F \supseteq A \text{ is closed}\}$.*

Proof. This is because if one takes any limit point from A, then it lies in all $F \in \mathscr{F}$ (since any limit point of $A \subseteq F$ is a limit point of F), and conversely, $\overline{A}$ is a closed set, so $\overline{A} \in \mathscr{F}$, and thus $\bigcap \mathscr{F} \subseteq \overline{A}$. ⊠

[8] Pun intended?

[9] "So a limit point walks into X-bar… and the bartender says, 'We're closed.' "

This characterization as the smallest closed set that contains A is occasionally also useful.

Definition. $U \subseteq M$ is open if for all $x \in U$, there exists an $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq U$.[10]

Note that M and $\emptyset$ are both open, the former for any ε and the latter for logical reasons (if some statement includes "for all elements of a set," then the empty set trivially satisfies it). They're also closed, as we saw yesterday, so a set can be both open and closed.

Example 6.1. For a more interesting example, consider the discrete metric space on some set M, given by

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

Then, $B_{1/2}(x) = \{x\}$, so every $U \subseteq X$ is open.

On $\mathbb{R}$, open intervals are open sets, and closed intervals are closed sets.

There's a similar characterization of open sets as Theorem 5.6, but with the important words switched!

Theorem 6.2.

(1) A finite intersection of open sets is open.
(2) An arbitrary union of open sets is open.

Example 6.2. A counterexample to extending (1) to arbitrary intersections is that

$$[0,1] = \bigcap_{n=1}^{\infty} \left(-\frac{1}{n}, 1 + \frac{1}{n} \right).$$

This is good to keep in mind.

Proof of Theorem 6.2.

- (2) is easier, so we'll do it first. Suppose $\mathscr{U}$ is a collection of open sets, and let $U = \bigcup \mathscr{U}$. Then, for any $x \in U$, there exists a $U' \in \mathscr{U}$ such that $x \in U'$, so there exists an $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq U' \subseteq U$, so U is open by definition.
- For (1): let $U = \bigcap_{i=1}^n U_i$ for some open sets $U_1, \dots, U_n$. Thus, if $x \in U$, then $x \in U_i$ for each i, so there exist $\varepsilon_1, \dots, \varepsilon_n$ such that $B_{\varepsilon_i}(x) \subseteq U_i$. If $\varepsilon = \min\{\varepsilon_1, \dots, \varepsilon_n\}$, then for each i, $B_\varepsilon(x) \subseteq B_{\varepsilon_i}(x) \subseteq U_i$, so $B_\varepsilon(x) \subseteq U$, so U is open by definition. (Notice that if there were infinitely many such sets, then the set of ε_i might not have had a positive lower bound.) ⊠

Proposition 6.3. *It's also true that open balls $B_r(x)$ are open sets.*

Proof. For any $x \in M$ and $r > 0$, look at $B_r(x)$. For any $y \in B_r(x)$, let $\varepsilon = r - d(x,y)$, so that $\varepsilon > 0$ (because $d(x,y) < r$). Thus, by the triangle inequality, for any $z \in B_\varepsilon(y)$, $d(x,z) \le d(y,z) + d(x,y) < r - d(x,y) + d(x,y) = r$, so $z \in B_r(x)$, and thus $B_r(x)$ is open. ⊠

This is good, because the names are similar, so it would have been bad if there were confusion.

Definition. Within some set M, the complement of an $A \subseteq M$ is $A' = \{x \in M : x \notin A\}$.

Theorem 6.4. *U is open iff U' is closed.*

This is sometimes used to define closed sets: the definition of open sets is given as above, and then closed sets are defined as the complements of the open ones. The same properties hold, and you get the same closed sets.

Proof of Theorem 6.4. In the forward direction, assume that U is open, and we want to prove that U' is closed. Let $\{x_n\}$ be a convergent sequence from U', and $x = \lim_{n\to\infty} x_n$. Then, if $x \in U$, then there exists an $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq U$, and thus since $\{x_n\}$ converges, then there is an N such that when $n \ge N$, $|x_n - x| < \varepsilon$, i.e. $x_n \in B_\varepsilon(x)$, so they're in U' but not U, which is a contradiction. Thus, $x \in U'$, so U' is closed.

Suppose conversely that U' is closed; then, the goal is to show that U is open. Let $x \in U$; then, there must exist some $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq U$, because if not, then there exists an $x_n \in B_{1/n}(x) \cap U'$ for all $n \in \mathbb{N}$. Thus, $\{x_n\}$ converges to x, but all of the $x_n \in U'$, so x is a limit point of U' and thus in U', since U' is closed. But this is a contradiction, since we assumed $x \in U$. ⊠

The study of open and closed sets really belongs to the domain of point-set topology, which is considerably more general. One can do some, but not all, of the analysis done here on topological spaces (but the notions of open and closed sets are similar).

[10] The professor used O as the symbol for an open set; I prefer U, as it's easier to disambiguate in my handwriting.

Definition. If T is a set, a collection $\mathscr{U}$ of subsets of T is called a topology (and T is a topological space) if the following axioms are true:

- $T, \emptyset \in \mathscr{U}$.
- $\mathscr{U}$ is closed under finite intersections: if $U_1, \dots, U_n \in \mathscr{U}$, then $\bigcup_{i=1}^n U_i \in \mathscr{U}$.
- $\mathscr{U}$ is closed under any unions: if $\mathscr{U}_1 \subseteq \mathscr{U}$, then $\bigcup \mathscr{U}_1 \in \mathscr{U}$.

Then, the elements of $\mathscr{U}$ are considered the open sets in the topology. The discussion above implies that metric spaces induce a topology. Not all topologies induce metric spaces, which isn't as obvious (though we showed that in the topology induced by a metric space, any two points have disjoint open neighborhoods, which isn't true of all topological spaces). There aren't complete criteria for whether a space is metrizable (comes from a metric), but plenty of partial criteria. Also, lots of notions such as continuity, convergence, etc. still hold in general topological spaces.

Functions and Maps on Metric Spaces. Let (M_1, d_1) and (M_2, d_2) be metric spaces.

Definition. If $f : M_1 \to M_2$ and $x \in M_1$, then f is continuous at x if for all $\varepsilon > 0$, there exists a $\delta > 0$ such that whenever $d_1(x, y) < \delta$, then $d_2(f(x), f(y)) < \varepsilon$. If f is continuous at all $x \in M_1$, then it is said to be continuous on M_1.

This is very similar to the notion of continuity for real-valued functions.

Example 6.3.

- If (M_1, d_1) is discrete, then every $f : M_1 \to M_2$ is continuous! This is because if $\delta = 1/2$ for any ε, then $B_\delta(x) = \{x\}$ for all $x \in M_1$, so $d_2(f(x), f(x)) = 0 < \varepsilon$, so the definition is satisfied.
- Let $M = \{a = \{a_n\}_{n=1}^\infty : a_n = 0 \text{ for large enough } n\}$, with $d(x, y) = \sum_{n=1}^\infty |a_n - b_n|$. This is clearly a finite sum, since there are only a finite number of terms. Then, $M \subset \ell^1$, with the same metric.

  Take $f : M \to \mathbb{R}$ by $f(a) = \sum_{n=1}^\infty n a_n$. Each sum is finite, yet this function isn't continuous in the L^1 norm (though it may be in others; the choice of norm is important for infinite-dimensional spaces). For example, at the zero sequence, construct a sequence $\{a^{(i)}\}$ on M where $a_n^{(i)} = (1/n)\delta_{in}$. Then, $d(a^{(i)}, 0) = 1/i$, so $a^{(i)} \to 0$, but $f(a^{(i)}) = 1$ and $f(0) = 0$.

This latter example proved something is discontinuous at x by finding a sequence that converges to x, but such that f applied to that sequence doesn't converge to $f(x)$. Here's why.

Theorem 6.5. *Suppose $a \in M_1$ and $f : M_1 \to M_2$; then, f is continuous at a iff for all sequences $\{x_n\}$ such that $\lim_{n\to\infty} x_n = a$, it's true that $\lim_{n\to\infty} f(x) = f(a)$.*

Proof. In the forward direction, assume f is continuous at a, and let $\{x_n\}$ be a sequence such that $x_n \to a$.[11] Then, for any $\varepsilon > 0$, the continuity of f implies there is a δ such that $d_2(f(x), f(a)) < \varepsilon$ if $d_1(x, a) < \delta$.

Since $\{x_n\}$ converges, then for every $\delta > 0$ there's an N such that if $n \geq N$, then $d_1(x_n, a) < \delta$, and thus $d_2(f(x), f(a)) < \varepsilon$, which means $f(x_n) \to f(a)$ as $n \to \infty$.

Suppose that conversely, whenever $x_n \to a$, then $f(x_n) \to f(a)$, and assume that f is not continuous at a, i.e. there exists an $\varepsilon_0 > 0$ such that for all $\delta > 0$, there exists an x such that $d_1(x, a) < \delta$, but $d_2(f(x), f(a)) \geq \varepsilon_0$. We want to get out of this a sequence $x_n \to a$ such that $f(x_n) \not\to f(a)$, so pick your favorite sequence going to 0, i.e. $\delta = 1/n$: thus, there exists an x_n for each n such that $d_1(x_n, a) < 1/n$ but $d_2(f(x_n), f(a)) \geq \varepsilon_0$, so $x_n \to a$, but $f(x_n) \not\to f(a)$, so there's a contradiction. ⊠

One way to think of this is that one can interchange the taking of limits with continuous functions; they commute, in some sense.

Here are a few relatively easy facts about continuous functions, which are left as exercises:

- If $f, g : M \to \mathbb{R}$ are continuous at an $a \in M$ and $c \in \mathbb{R}$, then $f + g$, $f - g$, and fg are continuous at a, and f/g is as well if $g(x) \neq 0$.
- If $a \in M$, then $f(x) = d(x, a)$ is continuous, because as seen in the homework,
$$|f(x) - f(y)| = |d(x, a) - d(y, a)| \leq d(x, y).$$
  This is known as the Lipschitz condition for continuity of a function ($d(f(x), f(a))$ is bounded by a constant multiple of $d(x, a)$).

Theorem 6.6. *If $f : M_1 \to M_2$, then the following are equivalent:*

(1) f is continuous on M_1.
(2) For all closed $C \subseteq M_2$, the set $f^{-1}(C) = \{x : f(x) \in C\}$ is closed.[12]

[11] This is shorthand for $\lim_{n\to\infty} x_n = a$, as it's easier to write, not to mention TeX up.

[12] This notation f^{-1} means the preimage, and places no constraints on whether f is actually invertible.

(3) *For all open sets $U \subseteq M_2$, the set $f^{-1}(U)$ is open.*

This is less obvious from the definition, but is simple and elegant. It's also important, connecting the theory of continuous functions to the topology. It's also very convenient for some proofs.

Corollary 6.7. *If M_1, M_2, and M_3 are metric spaces, and $g : M_1 \to M_2$ and $f : M_2 \to M_3$ are continuous, then $f \circ g$ is continuous.*

Proof. Let $U \subseteq M_3$ be open; then, $(f \circ g)^{-1}(U) = g^{-1}(f^{-1}(U))$, but U is open and f is continuous, so $f^{-1}(U)$ is open and since g is continuous, so is $g^{-1}(f^{-1}(U))$. ⊠

7. Subspaces and Compactness: 4/22/14

> *"When I wrote [the ε-δ definition of continuity], only God and I understood what I was doing. Now, God only knows." – Karl Weierstrass*

Last time, we stated a criterion for a function $f : M_1 \to M_2$ of metric spaces to be continuous: continuity is equivalent to the preimage of closed sets being closed. That is, if $C \subseteq M_2$ is closed and f is continuous, then $f^{-1}(C) = \{x \in M_1 : f(x) \in C\}$ is closed, or equivalently one can require that $f^{-1}(U)$ is open for all open subsets of M_2.

Proof of Theorem 6.6.

- (1) $\implies$ (2): Assume f is continuous, and let $C \subseteq M_2$ be a closed set; then, we want to show that $f^{-1}(C)$ is closed. Let $\{x_n\}$ be a sequence in $f^{-1}(C)$ with $x_n \to x$. But then, since f is continuous, then $f(x_n) \to f(x)$. Since C is closed and $f(x)$ is a limit point of C, then $f(x) \in C$, so $x \in f^{-1}(C)$.
- (2) $\implies$ (3): Assume $f^{-1}(C)$ is closed when C is closed. If $U \subseteq M_2$ is open, then its complement within M_2, U', is closed, so $f^{-1}(U')$ is closed in M_1. Then, one can check that $f^{-1}(U') = (f^{-1}(U))'$, so $f^{-1}(U)$ is open (since the complement of a closed set is open, and vice versa).
- (3) $\implies$ (1): Let $x \in M_1$ and $\varepsilon > 0$. Then, $B_\varepsilon(f(x)) \subseteq M_2$ is open, so $f^{-1}(B_\varepsilon(M_2))$ is open in M_1 by the assumption. But clearly $x \in M$, so there is some $\delta > 0$ such that $B_\delta(X) \subseteq f^{-1}(B_\varepsilon(M_2))$, so for all y such that $d_1(x, y) < \delta$, $d_2(f(x), f(y)) < \varepsilon$, so by definition f is continuous. ⊠

Subspaces. Lots of interesting metric spaces arise as subspaces of other spaces, so it's worth describing how open and closed sets behave in this induced relative topology.

Definition. If (M, d) is a metric space and $X \subseteq M$, then $(X, d|_X)$ is also a metric space (i.e. d restricted to X), and d is called the relative metric (sometimes induced metric in geometry).

Example 7.1. Consider $[a, b] \subseteq \mathbb{R}$. Then, open balls in $[a, b]$ include (x, y) for $x, y \in (a, b)$, but there are also half-open intervals: for any $c \in (a, b)$, $(c, d]$ is open in $[a, b]$, since $B_\varepsilon(b)$ is also a half-open interval for sufficiently small ε. Thus, relatively open sets need not be open in the space the metric was induced from.

For another interesting example, if one induces the usual metric on $(a, b) \subset \mathbb{R}$, then for any $c \in (a, b)$, $(a, c]$ is relatively closed; it contains all of its limit points, since a isn't a limit point (because it's not in (a, b)).

Theorem 7.1. *Let M be a metric space and $X \subseteq M$ be given the induced metric.*

(1) *$C \subseteq X$ is closed iff $C = C_1 \cap X$ for some C_1 closed in M.*
(2) *$U \subseteq X$ is open iff $U = U \cap X$ for some U_1 open in M.*

Corollary 7.2.

(3) *If X is closed in M and $C \subseteq X$ is closed in X, then C is closed in M.*
(4) *If X is open in M and $U \subseteq X$ is open in X, then U is open in M.*

These follow because the intersection of two closed sets is closed, and the intersection of two open sets is open.

Proof of Theorem 7.1.

- For (1), in the forward direction, suppose $C \subseteq X$ is closed and consider its closure $\overline{C} = \{x : x \text{ is a limit point of } M\}$, which is closed in M. Then, $C = \overline{C} \cap X$ (since if x is a limit point of C within M, then if it is in C, then it must also be in X, since C is closed in X).

  In the reverse direction, suppose $C = C_1 \cap X$ for some C_1 closed in M. Let $x_n \to x$ be a sequence in C; then, x is a limit point, so it must be in C_1, so if $x \in X$, then $x \in C$, so C is closed.
- For (2), assume $U \subseteq X$ is open, so for all $x \in U$ there exists an $\varepsilon_x > 0$ such that $B^X_{\varepsilon_x}(x) \subseteq U$ (here meaning the open ball taken in X). Thus,

$$U = \bigcup_{x \in U} B^X_{\varepsilon_x}(x) \subseteq \bigcup_{x \in U} B^M_{\varepsilon_x}(x) = U_1,$$

since $B^X_\varepsilon(x) = B^M_\varepsilon(x) \cap X$, and hence $U = U_1 \cap X$.

Conversely, if $U = U_1 \cap X$ for some U_1 open in M, then there exists some $\varepsilon > 0$ such that $B^M_\varepsilon(x) \subseteq U_1$, and thus $B^X_\varepsilon(x) \subseteq U_1 \cap X = U$, so U is open. ⊠

It's important to remember that closedness and openness in a subset depend only on the subset in question!

Compactness. This is another important basic notion that generalizes from $\mathbb{R}$. The idea is that compact spaces should satisfy a property similar to the Heine-Borel property; in fact, the closed interval $[a, b] \subseteq \mathbb{R}$ is the archetypal example of a compact set.

Definition.

- Let $\mathscr{O}$ be a collection of open sets of M. $\mathscr{O}$ is called an open covering of M if
$$M = \bigcup \mathscr{O} = \bigcup_{U \in \mathscr{O}} U.$$
- M is compact if every open covering has a finite subcovering.

Example 7.2.

- Suppose (M, d) has the discrete metric, i.e. $d(x, y) = 1$ if $x \neq y$, and 0 if $x = y$. Then, each one-point set $\{x\}$ is open, so M is compact iff it is finite.
- $(0, 1)$ is not compact; as we saw in class, there is no finite subcovering of
$$(0, 1) = \bigcup_{n=3}^{\infty} \left(\frac{1}{n}, 1 - \frac{1}{n} \right).$$
- However, by the Heine-Borel theorem, $[0, 1]$ is compact. Note that the theorem is about open intervals, as opposed to open sets in general, which requires thinking about the relative topology. Let $\mathscr{O}$ be an open covering of $[a, b]$, and suppose $x \in [a, b]$. Then, there is some $U_x \in \mathscr{O}$ such that $x \in U_x \subseteq [a, b]$ (i.e. U_x is relatively open). Thus, there is some $U^{(1)}_x$ such that $U_x = [a, b] \cap U^{(1)}_x$ by Theorem 7.1, and thus there exists an open interval I_x such that $x \in I_x \subseteq U^{(1)}_x$, so by the Heine-Borel theorem,
$$[a, b] \subseteq \bigcup_{x \in [a,b]} I_x, \qquad \text{and thus} \qquad [a, b] \subseteq \bigcup_{i=1}^{n} I_{x_i}$$
for some $x_1, \dots, x_n$. Then, $I_{x_i} \cap [a, b] \subseteq U^{(1)}_{x_i} \cap [a, b] = U_{x_i}$, so
$$[a, b] = \bigcup_{i=1}^{n} I_{x_i} \cap [a, b] = \bigcup_{i=1}^{n} U_{x_i}.$$

Notice that even though we require equality, rather than inclusion, in the finite subcovering, $[a, b]$ is a metric space in its own right, so it doesn't actually make a difference. The perspective is is important because compactness is intrinsic, so doesn't depend on some relative topology.

Theorem 7.3. *Assume M is compact and $f : M \to \mathbb{R}$ is continuous. Then:*

(1) f is bounded.
(2) f achieves its maximum and minimum.

Proof. The proof will be very similar to that for closed intervals.

- For (1), let $x \in M$ and $\varepsilon = 1$; then, there exists a $\delta_x > 0$ such that $|f(y) - f(x)| < 1$ if $d(x, y) < \delta_x$, and thus $|f(y)| < |f(x)| + 1$ for $y \in B_{\delta_x}(x)$. Therefore,
$$M = \bigcup_{x \in M} B_{\delta_x}(x),$$
so since it's compact, then there exist $x_1, \dots, x_n$ such that
$$M = \bigcup_{i=1}^{n} B_{\delta_i}(x_i)$$
(where $\delta_i = \delta_{x_i}$ to simplify notation), and thus $|f(y)| \leq \max_{1 \leq i \leq n} \left(|f(x_i)| + 1 \right)$, which is an explicit upper bound for $f(x)$.

- For (2), let $\overline{f} = \sup\{f(x) : x \in M\}$; then, let $F(x) = 1/(\overline{f} - f(x))$; then, this function is unbounded, since for any ε, there's an $x \in M$ such that $f(x) > 1/\varepsilon$, so it can't be continuous (since M is compact); thus, there must be some $\overline{x}$ such that $f(\overline{x}) = \overline{f}$. The minimum is identical, except with $-f$ instead of f. ⊠

For any $x \in M$, $f(y) = d(x, y)$ is continuous, so if M is compact, then there exists some $R > 0$ such that $M = \overline{B_R(x)}$. Thus, M is bounded in this sense, which is a very useful criterion for showing some spaces are not compact, e.g. $\mathbb{R}$, $\mathbb{R}^n$, or ℓ^1. In some sense, compactness implies boundedness.

That said, boundedness doesn't imply compactness: let $M = \{a \in \ell^1 : \|a\| \le 1\} = \overline{B_1(0)}$, which is clearly a closed, bounded subspace of ℓ^1, but is noncompact; this is easiest to show with a criterion which we'll prove next time:

Theorem 7.4. *A metric space M is compact iff every sequence in M has a convergent subsequence.*

Then, we already know that $\overline{B_1(0)} \subseteq \ell^1$ lacks the Bolzano-Weierstrass property, so it can't be compact.

8. Sequential Compactness and Connectedness: 4/24/14

"The indexing I used here is that the j^{th} component of $x_j^{(i)}\ldots$, no wait, jk."

Definition.

- A metric space M is sequentially compact if every sequence has a convergent subsequence.
- M is called totally bounded if for all $\varepsilon > 0$, there exist $x_1, \ldots, x_n \in M$ such that
$$M = \bigcup_{i=1}^{n} B_\varepsilon(x_i).$$

We saw that compactness in $\mathbb{R}$ implies both the Heine-Borel property (which was used as the generalization of compactness to general metric spaces) and the Bolzano-Weierstrass property which extends to sequential compactness. Fortunately, there's no difference.

Theorem 8.1. *(M, d) is compact iff it is sequentially compact.*

This statement isn't actually true for general topological spaces, so we'll need to use the metric somehow. However, a compact topological space is sequentially compact.

Proof of Theorem 8.1. In the forward direction, let (M, d) be a compact metric space, and let $\{x_n\}$ be a sequence from M. Suppose that there is no convergent subsequence of $\{x_n\}$; then, for any $x \in M$, there exists an $\varepsilon_x > 0$ such that $B_\varepsilon(x)$ contains at most finitely many x_n. Then, let $\mathscr{U} = \{B_{\varepsilon_x}(x) : x \in M\}$. Since M is compact, then this open covering has a finite subcovering $B_{\varepsilon_1}(x_1), \ldots, B_{\varepsilon_n}(x_n)$ (where $\varepsilon_i = \varepsilon_{x_i}$), but each one contains only finitely many x_n, and thus $\{x_n\}$ is a finite sequence, which is a contradiction.

The above proof works in any topological space (though it has to use a slightly different presentation than open balls). The other direction, which does depend on the metric, is harder.

Assume (M, d) is sequentially compact. Then, we can show that it is totally bounded, for if not, then let $x_1 \in M$, and then there exists an $x_2 \notin B_\varepsilon(x_1)$, and since $B_\varepsilon(x_1) \cup B_\varepsilon(x_2) \ne M$, then there's an $x_3 \notin B_\varepsilon(x_1) \cup B_\varepsilon(x_2)$, and so on, producing a sequence such that $d(x_n, x_m) \ge \varepsilon$ when $n \ne m$, since
$$x_n \notin \bigcup_{i=1}^{n-1} B_\varepsilon(x_i).$$
This implies that it lacks a convergent subsequence, which contradicts sequential compactness; thus, M is totally bounded.

Next, we will show that for any open covering $\mathscr{U}$, there is an $\varepsilon > 0$ such that for all $x \in M$, $B_\varepsilon(x) \subseteq U$ for some $U \in \mathscr{U}$. This is also by contradiction; suppose not, so that there is no such ε. Then, for any $n \in \mathbb{N}$, there's an $x_n \in M$ such that $B_{1/n}(x_n)$ isn't contained in $U \in \mathscr{U}$ for all U. Thus, there's a sequence $\{x_{n_i}\}_{i=1}^\infty$ which converges to a limit x. Then, $x \in U$ for some $U \in \mathscr{U}$, so there exists an $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq U$, but this means there's an n_i such that $x_{n_i} \in B_{\varepsilon/2}(x)$, and this implies that $B_{\varepsilon/2}(x_{n_i}) \subseteq B_\varepsilon(x) \subseteq U$, which is a contradiction.

Finally, we can prove compactness. If $\mathscr{U}$ is any open covering, then there's an $\varepsilon > 0$ such that for every $x \in M$, there's a $U_x \in \mathscr{U}$ with $B_\varepsilon(x) \subseteq U_x$. But since M is totally bounded, then there exist $x_1, \ldots, x_n$ such that
$$M = \bigcup_{i=1}^{n} B_\varepsilon(x_i) \subseteq \bigcup_{i=1}^{n} U_{x_i},$$
which is a finite subcover of $\mathscr{U}$. ⊠

Theorem 8.2. *Any closed subset of a compact metric space is compact (i.e. in the induced metric).*

Proof. If $F \subseteq M$ is closed and $\{x_n\}$ is a subsequence from F, then since M is compact and therefore sequentially compact, then $\{x_n\}$ has a convergent subsequence $\{x_{n_i}\}$, converging to some $x \in M$. Thus, x is a limit point of F, so since F is closed, then $x \in F$, and therefore F is sequentially compact and thus compact. ⊠

Definition. A subset F of a metric space M is bounded if there is some $C \in \mathbb{R}$ such that $d(x, y) \leq C$ for all $x, y \in M$. Equivalently, $F \subseteq B_{R_x}(x)$ for some sufficiently large R_x for all $x \in F$.

Theorem 8.3. *If $F \subseteq M$ is compact, then it is closed (in M) and bounded.*

Proof. Since the distance function is continuous, then $f(x) = d(x, y)$ for some $y \in F$ is bounded (on $\mathbb{R}$), so F must be bounded as well.

For closedness: let $\{x_n\}$ be a sequence from F such that $x_n \to x$. Since F is compact, then there's a convergent subsequence x_{n_i} that converges to something in F. But this has to be x, so $x \in F$, and thus F contains all of its limit points, and is closed. ⊠

The converse is not true; there exist closed, bounded subsets of metric spaces that aren't compact (e.g. $\overline{B_1(0)} \subset \ell^1$). However, this is a fair characterization in some important spaces.

Theorem 8.4. *A subset $F \subseteq \mathbb{R}^n$ is compact iff F is closed and bounded.*

Proof. The forward direction is the content of Theorem 8.3, applied to $\mathbb{R}^n$, so conversely assume that F is closed and bounded. Then, any $x \in F$ has coordinates $(x_1, \ldots, x_n)$, which must also all be bounded, so if $\{x^{(i)}\}$ is a sequence from F, then $\{x_j^{(i)}\}$ is also bounded in $\mathbb{R}$ for $j = 1, \ldots, n$.

Now, we can use Bolzano-Weierstrass: $\{x_1^{(i)}\}$ is bounded, so it has a convergent subsequence $\{x_1^{(i_j^1)}\}_{j=1}^{\infty}$, with limit $y_1 \in \mathbb{R}$. Then, we want these indices to be the sequence a convergent subsequence is chosen from for the second coordinate, and so on, which is an adventure in notation: $\{x_2^{(i_j^1)}\}$ is a bounded real sequence, so by the Bolzano-Weierstrass theorem, there's a convergent subsequence $\{x_2^{(i_j^2)}\}_{j=1}^{\infty}$, whose limit exists and can be called y_2. But since this is a subsequence of the previous one, then it's still true that $\lim_{j\to\infty} x_1^{(i_j^2)} = y_1$.

Repeat this n times, yielding an $\{x^{(i_j^n)}\}$ in $\mathbb{R}^n$ such that all of its convergent functions converge: $x_p^{(i_j^n)} \to y_p$ for $p = 1, \ldots, n$. Then, by Theorem 5.4, this implies that $x^{(i_j^n)} \to y = (y_1, \ldots, y_n)$. ⊠

Definition. A function $f : M_1 \to M_2$ is uniformly continuous if for all $\varepsilon > 0$ there exists a δ such that whenever $d_1(x, y) < \delta$, then $d_2(f(x), f(y)) < \varepsilon$.

The idea is that in conventional continuity, the choice of δ depends on x, which is suboptimal for some uses.

Theorem 8.5. *If M_1 is compact and $f : M_1 \to M_2$ is continuous, then f is uniformly continuous.*

Proof. This will look very much like a past homework problem: let $\varepsilon > 0$. Then, for all $x \in M_1$, there exists a $\delta_x > 0$ such that $d_2(f(x), f(y)) < \varepsilon/2$ whenever $d_1(x, y) < \delta$. Let $\mathscr{U} = \{B_{\delta_x/2}(x) : x \in M\}$, which is an open covering of M_1, so since M_1 is compact, there exist $x_1, \ldots, x_n$ such that

$$M = \bigcup_{i=1}^{n} B_{\delta_i/2}(x_i),$$

where $\delta_i = \delta_{x_i}$ to conserve indices.

Let $\delta = \min(\delta_1/2, \ldots, \delta_n/2)$, so that if $x, y \in M_1$ are such that $d(x, y) < \delta$, then there is an i such that $x \in B_{\delta_i/2}(x_i)$, so that $d(y, x_i) \leq d(y, x_i) + d(x, x_i) < \delta$, and thus $d_2(f(x), f(y)) < \varepsilon$. ⊠

Moving to connectedness, one might initally define two points in a metric space to be connected if there's a continuous path between them. This is known as path-connectedness, and is often the same notion, but differs in some less-than-nice spaces. Thus, we'll want some other characterization that relies on the fact that connectedness should be an intrinsic property.

Definition. A metric space M is connected if the only subsets that are both open and closed are $\emptyset$ and M.

For example, if I_1 and I_2 are disjoint closed intervals in $\mathbb{R}$, then $I_1 \cap I_2$ is disconnected, because I_1 and I_2 are both contained in open and closed intervals of $\mathbb{R}$.

"One of the first things I learned [when giving math talks] is to never write on a board that doesn't slide."

Recall the definition of connectedness: that a metric space (M,d) is connected if its only subsets that are both open and closed are $\emptyset$ and M.

Theorem 9.1. *The following are equivalent:*

(1) *M is* not *connected.*
(2) *There exist nonempty open $U_1, U_2 \subset M$ such that $U_1 \cap U_2 = \emptyset$ and $U_1 \cup U_2 = M$.*
(3) *There exist nonempty closed sets $F_1, F_2 \subset M$ such that $F_1 \cap F_2 = \emptyset$ and $F_1 \cup F_2 = M$.*

Proof.

- For (1) $\implies$ (2), there exists some nonempty $A \subsetneq M$ such that A is open and closed, so let $U_1 = A$ and $U_2 = A'$ (i.e. the complement of A in M). Then, U_1 is open, and U_1 is closed, so its complement U_2 is open. And since they're complements, then $U_1 \cap U_2 = \emptyset$ and $U_1 \cup U_2 = M$.
- For (2) $\implies$ (3), take the given U_1 and U_2; then, since $U_1 = U_2'$ is the complement of an open set, it's closed, and similarly for $U_2 = U_1'$, so U_1 and U_2 are closed sets satisfying the statement.
- For (3) $\implies$ (1), suppose F_1 and F_2 are the sets given; since $F_1 = F_2'$, then F_1 is also open, but since F_1 and F_2 are both nonempty, then F_1 is a closed and open set that is nonempty and not all of M. Thus, M is disconnected. ⊠

In order to prove the Intermediate Value theorem, we will want to characterize connected subsets of the real line. Connectedness is an intrinsic property; though we define connected subsets in terms of the relative metric, it's an intrinsic property, just like compactness.

Many arguments about connectedness are done by contradiction, since it's easier to work with disconnected spaces sometimes.

Theorem 9.2. *If M_1 is connected and $f : M_1 \to M_2$ is continuous, then $f(M_1)$ is connected.*

Proof. Suppose that $f(M_1) = U_1 \cap U_2$ with U_1, U_2 disjoint and open in M_2. Thus, there exist U_1', U_2' open in M such that $U_1 = U_1' \cap f(M_1)$ and $U_2 = U_2' \cap f(M_2)$. But then, $f^{-1}(U_1')$ and $f^{-1}(U_2')$ are open in M_1, but $M = f^{-1}(U_1') \cup f^{-1}(U_2')$ and $f^{-1}(U_1') \cap f^{-1}(U_2') = \emptyset$. Thus, one of $f^{-1}(U_1')$ and $f^{-1}(U_2')$ is empty, so without loss of generality suppose it's $f^{-1}(U_1') = \emptyset$; then, $U_1' \cap f(U_1) = \emptyset$ and thus $U_1 = \emptyset$, so $f(M_1)$ must be connected. ⊠

Of course, M_2 might not be connected; we only know about the image of M_1.

Theorem 9.3. *A subset of $\mathbb{R}$ is connected iff it is a point or an interval. In particular, $\mathbb{R}$ is connected.*

Proof. Suppose $A \subset \mathbb{R}$ is connected, and assume there are at least two points in A (otherwise, we're done), and order them: assume $a, b \in A$ with $a < b$. Then, $(a,b) \subseteq A$, because if $c \in (a,b)$ but $c \notin A$, then write $A = (A \cap (-\infty, c)) \cup (A \cap (c, \infty))$, both of which are open intervals (as the intersection of two open intervals) that are each other's complements, so by Theorem 9.1, we're good.

Now, A must be the interval between inf(A) and sup(A) (could contain either, both, or neither), where either of these being infinite is understood.

Conversely, suppose I is either a point or an interval; points are obviously connected, so suppose I is a nontrivial interval, and suppose that I is disconnected, so that $I = F_1 \cup F_2$ for disjoint nonempty closed $F_1, F_2 \subset I$. Let $a \in F_1$ and $b \in F_2$, and without loss of generality, assume $a < b$; thus, there exists an $\varepsilon > 0$ such that $(a-\varepsilon, a+\varepsilon) \subseteq F_1$ and $(b-\varepsilon, b+\varepsilon) \subseteq F_2$.

Let $c = \sup\{x \in F_1 : x < b\}$ (the supremum exists because it contains a, so is nonempty, and is bounded above by b). Then, $a + \varepsilon \le c \le b - \varepsilon$, so $c \in \overline{F_1} = F_1$, but $c \in \overline{F_2}$ as well (as I is an interval, so c is the infimum of the complement), and thus $F_1 \cap F_2 \neq \emptyset$. ⊠

Putting together the last two statements gives us the following.

Theorem 9.4 (Intermediate Value Theorem)**.** *If M is connected and $f : M \to \mathbb{R}$ is continuous, then $f(M)$ is an interval. That is, if $y_1 < y_2$ and $y_1 = f(x_1)$ and $y_2 = f(x_2)$, then for any $y \in (y_1, y_2)$ there exists an $x \in M$ with $f(x) = y$.*

In metric spaces, there's also a notion of path-connectedness. In "nice" spaces, this is equivalent (a continuous path exists between any two points), but this is stronger than connectedness; there exist spaces that are connected, but not path-connected, though they're usually messy.

Background to the Arzelà-Ascoli Theorem.

Definition. If M is a compact metric space, then define $C(M) = \{f : M \to \mathbb{R} : f \text{ is continuous}\}$. This is a vector space, and can be turned into a normed space in a few ways; we will take the L^∞ norm $\|f\|_\infty = \sup\{|f(x)| : x \in M\}$. Then, $(C(M), \|\cdot\|_\infty)$ is a normed linear space, and thus a compact metric space.

Then, some properties of this space will be given.

Proposition 9.5. *Cauchy sequences converge in $C(M)$; that is, $C(M)$ is complete.*

Proof. Suppose $\{f_n\} \subseteq C(M)$ is Cauchy, so that for all $\varepsilon > 0$, there exists an N such that $\|f_n - f_m\|_\infty < \varepsilon$ when $m, n \geq N$, so for any $x \in M$, $\{f_n(x)\}$ is also Cauchy, because by the norm, $|f_n(x) - f_m(x)| \leq \|f_n - f_m\|_\infty < \varepsilon$. But the reals are complete, so this Cauchy sequence converges to some $f(x)$. This is the notion of pointwise convergence; the sequence of functions converges at every point.

But at every point we also have that $|f_n(x) - f_m(x)| < \varepsilon$ for all $x \in M$ and $n, m \geq N$. Letting $m \to \infty$, $|f_(x) - f(x)| \leq \varepsilon$. This means that the pointwise limit is actually a uniform limit (uniform convergence): $\|f_n - f\|_\infty \leq \varepsilon$ for all $n \geq N$. Then, that the limit is in $C(M)$ follows from Theorem 9.6, which comes next, but won't depend on this result. ⊠

Theorem 9.6. *The uniform limit of a sequence of continuous functions is continuous. To be precise, suppose $\{f_n\}$ is a convergent sequence (converging to f) of continuous functions on a metric space M such that for all $\varepsilon > 0$, there exists an N such that $|f_n(x) - f(x)| < \varepsilon$ for all $n \geq N$ and $x \in M$. Then, f is continuous.*

Proof. Let $c \in M$ and $\varepsilon > 0$. Then, there exists an N such that $|f_N(y) - f(y)| < \varepsilon/3$ for all $y \in M$. But f_N is continuous at x, so there exists a $\delta > 0$ such that $|f_N(x) - f_N(y)| < \varepsilon/3$ if $d(x, y) < \delta$, so

$$\begin{aligned} |f(x) - f(y)| &\leq |f(x) - f_N(x)| + |f_N(x) - f_N(y)| + |f_N(y) - f(y)| \\ &< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon. \end{aligned}$$ ⊠

Thus, $C(M)$ is complete.

Recall that a metric space M is totally bounded if for all $\varepsilon > 0$, M can be covered by a finite number of open balls of radius ε, i.e. there exist $x_1, \dots, x_n \in M$ such that

$$M = \bigcup_{i=1}^{n} B_\varepsilon(x_i).$$

All compact spaces are totally bounded, as are some noncompact ones (e.g. $(0,1)$).

Proposition 9.7. *M is totally bounded iff every sequence has a Cauchy subsequence.*

Proof. In the reverse direction, assume every sequence has a Cauchy subsequence, but that M isn't totally bounded, so that M cannot be covered by a finite number of balls of radius ε_0 for some $\varepsilon_0 > 0$.

Now, let $x_1 \in M$, so that there exists an $x_2 \notin B_{\varepsilon_0}(x_1)$, and then an $x_3 \notin B_{\varepsilon_0}(x_1) \cup B_{\varepsilon_0}(x_1)$, and so on, since no finite number can cover M; thus, $d(x_n, x_m) \geq \varepsilon_0$ for all $n \neq m$, so $\{x_n\}$ contains no Cauchy subsequence.

The converse is also a familiar argument: assume M is totally bounded, and let $\{x_n\}$ be a Cauchy subsequence. We'll look at balls of radius $1/k$; for any given k, there are a finite number that cover M, so at least one, called $B_{1/k}$, which must contain an infinite number of x_n.

Thus, let x_{n_1} be one of the x_i in B_1, then x_{n_2} be one of the points in $B_1 \cap B_{1/2}$ (since there must be some $B_{1/2}(x)$ that contains infinitely many points in intersection with B_1), and so on, so one can choose $x_{n_3} \in B_1 \cap B_{1/2} \cap B_{1/3}$, where $B_{1/3}$ was chosen to have infinitely many x_n, and so forth. Then, $\{x_{n_k}\}$ is Cauchy, because x_{n_ℓ} and x_{n_i} both lie in $B_{1/k}$ if $i, \ell \geq k$, so $d(x_{n_\ell}, x_{n_i}) < 2/k$. ⊠

10. Completeness: 5/6/14

We've seen that the real numbers are complete in several equivalent ways, some of which don't generalize (e.g. the supremum definition requires an order). However, the Cauchy criterion does generalize.

One direction is always true.

Claim. In any metric space, if $\{x_n\}$ converges, then it is Cauchy.

Proof. Suppose $\lim_{n\to\infty} x_n = x$; then, by the triangle inequality, $d(x_n, x_m) \leq d(x_n, x) + d(x, x_m)$, so given an $\varepsilon > 0$, there's an N such that when $n, m \geq N$, then $d(x, x_n) < \varepsilon/2$ and $d(x_n, x) < \varepsilon/2$, so the definition follows. ⊠

We care about the converse.

Definition. A metric space M is complete if all Cauchy sequences on M converge in M.

Example 10.1.

- We have proven that $\mathbb{R}$ is complete.
- $\mathbb{R}^n$ is complete, because a sequence $\{x^{(m)}\}$ is Cauchy in $\mathbb{R}^n$ iff its coordinate sequences $\{x_i^{(m)}\}$ are Cauchy in $\mathbb{R}$, and we showed these converge iff $\{x^{(m)}\}$ does, but this follows from Theorem 5.4.

Definition.

- A Banach space is a complete normed linear space.
- A Hilbert space is a complete inner product space.

Example 10.2. This means we can describe more complete spaces: clearly, $\mathbb{R}^n$ is a Hilbert space, but we also have ℓ^1 and ℓ^2, so in particular they're both complete.

If M is a compact metric space, then for any $f \in C(M) = \{f : M \to \mathbb{R} : f \text{ is continuous}\}$, $f(M)$ is bounded, so there's a natural norm called the ℓ^∞ norm, $\|f\|_\infty = \max_{x \in M} |f(x)|$. This makes $C(M)$ into a Banach space, and it carries the topology of uniform convergence (a sequence converges pointwise to a limit iff it converges uniformly to that limit).

Let's prove that ℓ^1 is complete, as an example.

Proof. Suppose $\{a^{(i)}\}_{i=1}^\infty$ is Cauchy in ℓ^1, so that $a^{(i)} = (a_1^{(i)}, a_2^{(i)}, \dots)$ such that the sum of the $|a_n^{(i)}|$ is finite. Then, for all $n \in \mathbb{N}$, $|a_n^{(i)} - a_n^{(j)}| \le \|a^{(i)} - a^{(j)}\|$, so each coordinate sequence converges: let $a_n = \lim_{n\to\infty} a_n^{(i)}$ and $a = \{a_n\}_{n=1}^\infty$. Now, we need to show that $a \in \ell^1$ and $a^{(i)} \to a$.

It's true in any metric space that a Cauchy sequence is bounded (take $\varepsilon = 1$; then, there can only be a finite number of points away more than distance 1 from most points, which leads to a bound). Thus, limits and sums commute in the following finite sum:

$$\sum_{n=1}^N |a_n| = \lim_{i\to\infty} \sum_{n=1}^N |a_n^{(i)}| \le \|a^{(i)}\| \le C$$

for some C. Thus, $a \in \ell^1$.

To show that it's Cauchy, note that since $a^{(i)}$ is Cauchy, then for any $\varepsilon > 0$ there exists an N such that for all $i, j \ge N$ and for all K,

$$\sum_{n=1}^K |a_n^{(i)} - a_n^{(j)}| \le \|a^{(i)} - a^{(j)}\| < \frac{\varepsilon}{2},$$

so taking the limit as $j \to \infty$,

$$\sum_{n=1}^K |a_n^{(i)} - a_n| \le \frac{\varepsilon}{2},$$

so $\|a^{(i)} - a\| \le \varepsilon/2 < \varepsilon$ when $i \ge N$. ⊠

The proof for ℓ^2 is very similar. Notice that in both cases it's important that $a^{(i)}$ is Cauchy, because this property doesn't hold for general bounded sequences.

Proposition 10.1. *If X is a closed subset of a complete metric space M, then X is itself complete.*

Proof. Suppose $\{x_n\}$ is a Cauchy sequence on X, so that it's still Cauchy on M. Thus, it converges to some $x \in M$, but since x is closed, then $x \in X$, so $\{x_n\}$ converges on X. ⊠

Note that there's no criterion on boundedness; this is more general than compactness.

Example 10.3. Here are some examples of incomplete spaces.

- Let D designate the closed unit disk in $\mathbb{R}^2$; then, $M = D \setminus \{(0,0)\}$ is not complete, because Cauchy sequences that should converge to 0 don't converge in M. In this sense, completeness means that there's no holes in the space.
- Let $I = [0,1]$ and consider the space $X = \{f \in C(I) : f \text{ is differentiable}\}$. This space is incomplete, because there exist uniform limits of differentiable functions that are not differentiable. For example, if $f(x) = |x - 1/2|$, then f isn't differentiable at $x = 1/2$, but one can approximate f arbitrarily closely by "rounding out the corner" to obtain differentiable functions. Thus, one can obtain a Cauchy sequence that doesn't converge. However, X is a normed linear space.[13] Hopefully this example is more interesting than the unit disk minus a point.

[13] X is also dense in $C(I)$, i.e. $\overline{X} = C(I)$, so every continuous function can be approximated uniformly by smooth ones. $C(I)$ has lots of interesting dense subspaces, including also the polynomial functions.

Completion of Metric Spaces. From any metric space, it's possible to "complete" it into a complete metric space with the same structure (e.g. an inner product space is completed into a Hilbert space, and a normed linear space is completed into a Banach space).

Definition.

- If M_1 and M_2 are metric spaces, then $F : M_1 \to M_2$ is an isometric embedding if for all $x, y \in M_1$, $d_2(F(x), F(y)) = d_1(x, y)$.
- F is called an isometry if it is a one-to-one and onto isometric embedding. (Note that any isometric embedding is necessarily one-to-one, but might not be onto; M_2 is in general larger than M_1).
- F is a homeomorphism if F is continuous and invertible, and F^{-1} is continuous.

Isometry is the notion of equivalence for metric spaces; two spaces which are isometric are indistinguishable as metric spaces. A homeomorphism can stretch distances, so it doesn't preserve them, but it does preserve any property coming from the topology induced by the metric, e.g. open and closed sets, convergence of sequences, and so on. This is the equivalence notion of topological spaces, but not always metric spaces: a coffee cup and a donut are homeomorphic! Another way to think of it is as providing a bijection between the open sets of the two spaces that respects inclusions.

So what is this mysterious notion of completeness? We want to do something akin to the process that constructed $\mathbb{R}$ out of $\mathbb{Q}$, so we want to find an isometric embedding $f : M \to (M_1, d_1)$ such that (M_1, d_1) is complete; then, $\overline{f(M)}$ is called the completion of M. To explicitly construct this sequence, consider the set of Cauchy sequences on M, and define an equivalence relation $\{x_n\} \sim \{y_n\}$ if $\lim_{n\to\infty} d(x_n, y_n) = 0$.

Lemma 10.2. *If $\{x_n\}, \{y_n\}$ are Cauchy, then the real-valued sequence $\{d(x_n, y_n)\}$ is Cauchy.*

Proof. These kinds of proofs rely on some key inequality.

$$\begin{aligned}
|d(x_n, y_n) - d(x_m, y_m)| &= |d(x_n, y_n) - d(x_n, y_m) - d(x_m, y_n) + d(x_m, y_n)| \\
&\leq |d(x_n, y_n) - d(x_m, y_n)| + |d(x_m, y_n) - d(x_m, y_m)| \\
&\leq d(x_n, x_m) + d(y_n, y_m),
\end{aligned}$$

but since $\{x_n\}$ and $\{y_n\}$ are Cauchy, then each of these can be taken to be less than $\varepsilon/2$ when m, n are large for any $\varepsilon > 0$, so the sequence of distances is also Cauchy. ⊠

Claim. $\sim$ is an equivalence relation.

Proof. It's necessary to check that it's reflexive, symmetric, and transitive. The first two are pretty obvious: $\{x_n\} \sim \{x_n\}$ because $d(x_n, x_n) = 0$, $\{x_n\} \sim \{y_n\}$ iff $\{y_n\} \sim \{x_n\}$ because the distance function is symmetric, and transitivity (that if $\{x_n\} \sim \{y_n\}$ and $\{y_n\} \sim \{z_n\}$, then $\{x_n\} \sim \{z_n\}$) by the triangle inequality: $d(x_n, z_n) \leq d(x_n, y_n) + d(y_n, z_n)$, but if since latter two go to zero, then the left side must as well. ⊠

An equivalence relation partitions its space into a set of disjoint equivalence classes that union to the whole set, so consider the set M_1 of equivalence classes of Cauchy sequences on M, i.e. the set of equivalence classes $X = \{\{\overline{x}_n\} : \{\overline{x}_n\} \sim \{x_n\}\}$ for all Cauchy sequences $\{x_n\}$, and turn M_1 into a metric space by setting

$$d_1(X, Y) = \lim_{n\to\infty} d(x_n, y_n),$$

where $\{x_n\} \in X$ and $\{y_n\} \in Y$. There's quite a bit to check: first, that the limit exists, which is fine, but also that if one chooses different $\{\overline{x}_n\} \in X$ and $\{\overline{y}_n\} \in Y$, then the same limit $d(\overline{x}_n, \overline{y}_n)$ is obtained. This also follows from the triangle inequality:

$$\begin{aligned}
|d(x_n, y_n) - d(\overline{x}_n, \overline{y}_n)| &= |d(x_n, y_n) - d(\overline{x}_n, \overline{y}_n) - d(x_n, \overline{y}_n) + d(x_n, \overline{y}_n)| \\
&\leq d(y_n, \overline{y}_n) + d(x_n, \overline{x}_n).
\end{aligned}$$

Furthermore, we have to check that d_1 is a metric. It's a limit of nonnegative things, so it's nonnegative, but we still need to show that if $d_1(X, Y) = 0$, then $X = Y$. But if $\lim_{n\to\infty} d(x_n, y_n) = 0$, then $\{x_n\} \sim \{y_n\}$, so $X = Y$, so this follows. The second point is symmetry, which follows because the distance is symmetric, and the triangle inequality follows from the corresponding triangle inequality on M.

Next, we want to construct an isometric embedding $F : M \to M_1$. It makes sense to send $x \mapsto (x, x, x, \dots)$ (the equivalence class of Cauchy sequences that converge to x), and this will end up being an isometric embedding, because $d_1(F(x), F(y)) = \lim_{n\to\infty} d(x, y) = d(x, y)$: we can choose any representatives for calculating the distances, so the constant sequence will do.

Finally, it's important to show that $F(M)$ is dense in M_1, so that $M_1 = \overline{F(M)}$. For any $X \in M_1$ and $\{x_n\} \in X$, then take the sequence $\widehat{x}_k^{(n)} = x_n$ for all $k \in \mathbb{N}$. These sequences are in $F(M)$, and if X_n is the equivalence class of $\{\widehat{x}_n^{(n)}\}$, then $\lim_{n\to\infty} X_n = X$. There's something to check here, that $d_1(X_n, X) = \lim_{n\to\infty} d(x_m, x_n) < \varepsilon$, so it is in the closure of $f(M_1)$, so we're almost done!

11. The Baire Category Theorem: 5/8/14

"The Baire Category theorem has a lot of amusing consequences."

Recall that if (M,d) is any metric space, a completion of M is a complete metric space (M_1,d_1) with an isometric embedding $F : M \to F_1$ such that $F(M)$ is dense in M_1 (i.e. $\overline{F(M)} = M_1$). If M isn't complete, then there are "holes" in M, in that there are Cauchy sequences that don't converge, but the images of this sequence do converge in $M_!$.

Theorem 11.1. *Any two completions of M are isometric.*

Proof. Suppose we have two completions

$$\begin{tikzcd} & M_2 \\ M \arrow[ur, "F_2"] \arrow[dr, "F_1"'] & \\ & M_1 \end{tikzcd}$$

Each of F_1 and F_1 is one-to-one, so let $F = F_2 \circ F_1^{-1}$, which is a function $F(M) \to M_2$. Since the image is dense in both M_1 and M_2, then (it's possible to show that) it extends to a unique continuous $\hat{F} : M_1 \to M_2$ that restricts to F on $F_1(M)$ and is onto. Thus, it's an isometry. ⊠

In some sense, the completion has to be abstract, as it involves creating a bigger space out of pure thinking.[14]

The explicit construction was given last time: that if $\{x_n\}$ and $\{y_n\}$ are Cauchy sequences on M, let $\{x_n\} \sim \{y_n\}$ if $\lim_{n\to\infty} d(x_n,y_n) = 0$. Then, M_1 is the set of equivalence classes of Cauchy sequences under this relation, and $d_1(X,Y) = \lim_{n\to\infty} d(x_n,y_n)$, where $\{x_n\} \in X$ and $\{y_n\} \in Y$. We showed that this is well-defined, and turns M_1 into a metric space. Then, $F : M \to M_1$ sends x to the equivalence class of $\{x_n\}$ where $x_n = x$ for all n. We showed this is isometric and that $F(M)$ is dense in M_1 (because if $X \in M_1$, then there's an $\{x_n\} \in X$ such that $X = \lim_{n\to\infty} F(x_n)$).

Thus, it remains only to show that M_1 is complete.

Lemma 11.2. *Suppose $\{X_n\}$ is Cauchy on M_1, with $X_n \in F(M)$; then, $\{X_n\}$ converges.*

Proof. There exists a Cauchy sequence $\{x_n\}$ on M such that $F(x_n) = X_n$, so $\{X_n\}$ is Cauchy (because F is isometric), and if X is the equivalence class of $\{x_n\}$, then $\lim_{n\to\infty} F(x_n) = X$. ⊠

Now, we can prove directly that M_1 is complete.

Proposition 11.3. (M_1,d_1) *is complete.*

Proof. Suppose $\{X_n\}$ is Cauchy in M_1, so that for all n, there exists a $Y_n \in F(M)$ with $d_1(x_n,y_n) < 1/n$, so $\{Y_n\}$ is also Cauchy. Thus, for any $\varepsilon > 0$, there exists an N such that if $m,n \geq N$, then

$$\begin{aligned} d_1(Y_n,Y_m) &\leq d_1(Y_n,Y_m) + d_1(X_n,X_m) + d_1(X_m,Y_m) \\ &\leq \frac{1}{n} + \frac{\varepsilon}{3} + \frac{1}{m} < \varepsilon. \end{aligned}$$

Thus, $Y_n \to X$ and $X_n \to X$. ⊠

The idea is that the Cauchy sequences in the image of M converge, and since the image is dense, a general Cauchy sequence can be perturbed into a Cauchy sequence in $F(M)$.

Completion preserves a lot of structure: if M is a field (e.g. $\mathbb{Q}$), then so is M_1; if M is a normed linear space, then M_1 is a Banach space; and if M is an inner product space, then M_1 is a Hilbert space. These often involve looking at an object abstractly, as equivalence classes of Cauchy sequences, but sometimes that's true already (e.g. on $\mathbb{R}$, that's effectively what the decimal expansion means).

If M is a complete space, then a subspace $X \subseteq M$ is complete iff it is closed, which leads to the following generalization of the characterization of compact subsets of $\mathbb{R}$.

Theorem 11.4. *If (M,d) is a metric space, then M is compact iff M is complete and totally bounded.*

Proof. In the forward direction, suppose M is compact and $\{x_n\}$ is Cauchy on M. Then, since M is compact, then $\{x_n\}$ has a subsequence $x_{n_k} \to x$ for some x, so, since $\{x_n\}$ is Cauchy and has a convergent subsequence, then $x_n \to x$ as well. Then, we've shown that compactness implies total boundedness in the characterization of sequential compactness (see the proof of Theorem 8.1).

In the other direction, suppose M is complete and totally bounded. If $\{x_n\}$ is any sequence on M, then it has a Cauchy subsequence because M is totally bounded. Since M is complete, then that subsequence converges, so every sequence on M has a convergent subsequence, so M is compact. ⊠

[14]This turns out to have a nice universal property: the completion M_1 of a metric space M is the object such that any continuous function $M_1 \to N$, where N is a complete metric space, factors through a continuous map $M \to N$.

Thus, compactness is much stronger than completeness, and should be thought of as very special. For example, $\mathbb{R}^n$ and ℓ^1 are complete but not compact.

Moving to the Baire Category theorem, suppose (M, d) is complete and consider some interesting kinds of sets.

Definition.

- $U \subseteq M$ is an *open dense set* if it's an open subset such that $\overline{U} = M$, e.g. $\mathbb{R} \subseteq \mathbb{R}$, or $\mathbb{R} \setminus \{x_1, \dots, x_n\} \subset \mathbb{R}$. But there are more interesting examples: let $\{x_n\}$ be an enumeration of the rationals (since they're countable), and for some $\varepsilon > 0$, let
$$U = \bigcup_{n=1}^{\infty} \left(x_n - \varepsilon 2^{-n-1}, x_n + \varepsilon 2^{-n-1}\right).$$
This seems like it might be equal to $\mathbb{R}$, but the length of the n^{th} interval is $\ell_n = \varepsilon 2^{-n}$, so the sum of the total lengths is ε. Thus, this set can be as small as we like! This is related to the idea of a nowhere dense set (even though, confusingly, they can be dense too). Open dense sets are big in some sense, but not in the sense of measure.
- $A \subseteq M$ is *nowhere dense* if $(\overline{A})'$ is dense, or equivalently, the interior of $\overline{A}$ is empty. This means that every point of A is a boundary! These include the complements of open dense sets given above. You can think of these as "thin" or "small."
- A G_δ *set* $A \subseteq M$ is a countable intersection of open sets,
$$A = \bigcap_{n=1}^{\infty} U_n,$$
for U_n open in M. These obviously include open intervals, but also closed intervals and points (take open intervals with endpoints whose limits approach the endpoints, or the point in question). However, not every set is a G_δ set, and the Baire Category theorem will show that this is actually a rather special property.
- A set is *meager* (or of the *first category*) if it's the countable union of nowhere dense sets. These should also be thought of as "small," and include finite collections of points, $\mathbb{Q}$, etc.
- A set is *second category* if it isn't first category.
- A set $A \subseteq M$ is *residual* if A' is of first category. This allows one to talk about "typical" elements of M; for example, this is the notion that most real numbers are irrational, that $\mathbb{R} \setminus \mathbb{Q}$ is residual in $\mathbb{R}$, even though constructing irrationals is tricky.

The notion of residual sets is often bad, leading to the following notion (not a rigorous theorem!).

Murphy's Law of Metric Spaces. *Most elements of a metric space M are bad with respect to most given properties.*

For example, the Baire category theorem will imply that most continuous functions on $[0, 1]$ are nowhere differentiable, and monotone on no interval! But actually constructing an example isn't obvious.

Theorem 11.5 (Baire Category)**.** *A complete metric space is of second category, i.e. it's not a countable union of nowhere dense sets.*

This is usually proved via an easier characterization.

Theorem 11.6. *If M is complete and $\{U_n\}_{n=1}^{\infty}$ is a sequence of open dense sets, then $\bigcap_{n=1}^{\infty} U_n$ is dense.*

Once this is known, the Baire Category theorem follows quickly.

Proof of Theorem 11.5. If M is complete, then suppose it can be written as a countable union of nowhere dense sets A_n; then,
$$M = \bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} \overline{A_n},$$
so therefore
$$\emptyset = M' = \bigcap_{n=1}^{\infty} (\overline{A}_n)',$$
but since these are open dense sets, this contradicts Theorem 11.6. ⊠

Thus, proving Theorem 11.6 is the real meat (or tofu, if you will) of this section.

Proof of Theorem 11.6. Suppose $x \in M$ and $\varepsilon > 0$. Then, we want to find a $y \in \bigcap_{n=1}^{\infty} U_n$ such that $d(x,y) < \varepsilon$.

Since U_1 is open, then there exists a $y_1 \in U_1 \cap B_\varepsilon(x)$, so there exists an $\varepsilon_1 > 0$ such that $B_{\varepsilon_1}(y_1) \subseteq U_1 \cap B_\varepsilon(x)$. Then, let $\delta_1 = \min\{\varepsilon/2, 1\}$ and choose a $y_2 \in U_2 \cap B_{\delta_1}(y_1)$, so that $B_{\varepsilon_2}(y_2) \subseteq U - 2 \cap B_{\delta_1}(y_1)$ (since U_2 is dense open, this can be done).

Then, let $\delta_2 = \min\{\varepsilon_2/2, 1/2\}$, and since U_3 is an open dense subset of M, there's a $y_3 \in U_3 \cap B_{\delta_2}(y_3)$ with $B_{\varepsilon_3}(y_3) \subseteq U_3 \cap B_{\delta_2}(y_2)$, and let $\delta_3 = \min\{\varepsilon_3/2, 1/3\}$.

Thus, it is possible to inductively define sequences y_n and ε_n, and define $\delta_n = \min\{\varepsilon_n/2, 1/n\}$ such that $B_{\varepsilon_{n+1}}(y_{n+1}) \subseteq U_{n+1} \cap B_{\delta_n}(y_n)$, and δ_n is a decreasing sequence (the radius shrinks). Thus, $\{y_n\}$ is Cauchy, so it converges to some limit y. This is the point we want; it just remains to show it's in all of the U_n, as well as in $B_\varepsilon(x)$.

Since $y_n \in B_{\delta_m}(y_m)$ for $m = 1, \dots, n-1$, then $y \in \overline{B_{\delta_n}(y_m)}$ for all m (since it could be on the boundary), but $\delta_n \le \varepsilon_n/2$, so in particular $y \in \overline{B_{\varepsilon_n/2}(y_m)} \subseteq B_{\varepsilon_m}(y_m) \subseteq U_m$ for all m, so therefore $y \in \bigcap_{m=1}^{\infty} U_m$. And since all of this happened within $B_\varepsilon(x)$, then $y \in B_\varepsilon(x)$. ⊠

The argument is a bit messy, but the idea is to construct a Cauchy sequence in which everything has the right properties.

Definition.

- If (M,d) is a metric space, then $x \in M$ is an isolated point if $\{x\}$ is open.
- A perfect subset $F \subseteq M$ is a closed subset with no isolated points.

Theorem 11.7. *If M is complete and has no isolated points, then M is uncountable.*

Corollary 11.8. *A perfect subset of $\mathbb{R}^n$ is uncountable.*

These both follow from The Baire Category theorem.

Within $\mathbb{R}$, the set $\{0\} \cup \{1/n : n \in \mathbb{N}\}$ is interesting: 0 is the only accumulation point (i.e. it's not an isolated point).

Proof of Theorem 11.7. If x is *not* isolated, then $\{x\}$ is nowhere dense, so if M is countable, then $M = \{x_n : n \in \mathbb{N}\}$, so $M = \bigcup_{n=1}^{\infty} \{x_n\}$, contradicting the Baire Category theorem. ⊠

Though it's not immediately obvious that one should invoke the Baire Category theorem, it makes the proof a lot easier; you could try to do it directly, but there would be a lot of fiddling involved.

Theorem 11.9. *$\mathbb{Q}$ is not a G_δ subset of $\mathbb{R}$.*

Proof. If $\mathbb{Q} = \bigcap_{n=1}^{\infty} U_n$ for U_n open, then each U_n is dense, so each U_n' is nowhere dense, so $\mathbb{Q}'$ is a countable union of nowhere dense sets; thus, $\mathbb{R} = \mathbb{Q} \cup \mathbb{Q}'$ is also a countable union of nowhere dense sets. ⊠

This argument applies to any countable subset of a complete metric space.

There are also some basic theorems in functional analysis that rely on the Baire Category theorem (uniform boundedness, closed curve theorem, etc.).

Part 3. **Integration**

12. Riemann Integration: 5/13/14

In Riemann integration, one wants to calculate the area under a curve f along an interval I, denoted $\int_I f$. If $I = [a,b]$, one approximates this by taking a partition $x_0 = a < x_1 < \cdots < x_N = b$: $\mathscr{P} = \{(x_{i-1}, x_i) : i = 1, \dots, N\}$.

Definition. If f is bounded on a region R, then the following quantity is well-defined, and is called the upper Riemann integral:

$$\overline{\int_R} f = \inf_{\mathscr{P}} \sum_{I \in \mathscr{P}} \overline{f}_I |I|;$$

here, $\overline{f}_I$ is the function equal to $\sup_{x \in I} f(x)$ on I and 0 elsewhere, and the infimum is taken over all partitions $\mathscr{P}$ of R.

In the same way the lower Riemann integral is defined to be

$$\underline{\int_R} f = \sup_{\mathscr{P}} \sum_{I \in \mathscr{P}} \underline{f}_I |I|,$$

where $\underline{f}_I = \inf_{x \in I} f(x)$ on I and is 0 elsewhere.

If the lower and upper integrals are the same, this value is called the Riemann integral, and f is said to be Riemann integrable on R.

One of the key ideas that allows this to work is that f is approximated by step functions. Thus, extend this notion to $\mathbb{R}^n$:

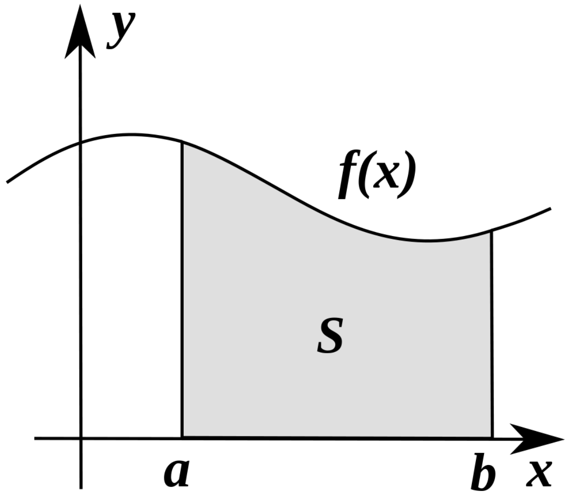

FIGURE 2. The goal of integration is to approximate $S = \int_a^b f(x)\,\mathrm{d}x$. Riemann integration does this by dividing the region into intervals and approximating the area in each interval. Source: `http://en.wikipedia.org/wiki/Riemann_integral`.

Definition.

- An interval on $\mathbb{R}^n$ is a product of intervals on $\mathbb{R}$: $I = I_1 \times \cdots \times I_n$.[15]
- A partition $\mathscr{P}$ of an interval $R \subset \mathbb{R}^n$ is given by writing $R = I_1 \times \cdots \times I_n$ and choosing a partition of each $I_i = [a_i, b_i]$; then, let
$$\mathscr{P} = \left\{[x_{1,i-1}, x_{1,i_1}] \times [x_{2,i_2-1}, x_{2,i_2}] \times \cdots \times [x_{n,i_n-1}, x_{n,i_n}] : 1 \le i_1 \le N_1, \dots, 1 \le i_n \le N_n\right\},$$
where N_i is the number of elements of the i^{th} partition. Thus, it's the set of all products of intervals from the chosen partitions on $\mathbb{R}$.
- φ is a step function in R if there exists a partition $\mathscr{P}$ such that $\varphi(x)$ is constant on the interior of each interval:
$$\varphi(x) = \sum_{I \in \mathscr{P}} a_I \chi_{I^0};$$
that is, the $a_I \in \mathbb{R}$ and χ_{I^0} denotes the characteristic function
$$\chi_A(x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \notin A \end{cases}$$
on the interior of each I. We also require φ to be bounded on $\bigcup_{I \in \mathscr{P}} \partial I$.
- The set $\mathscr{S}(R)$ denotes the set of all such step functions on R.

The idea is that we know what the integral of a step function ought to look like, and then generalize this to other functions on R. It will also be useful to require R to be closed in $\mathbb{R}^n$, so that it's a product of closed intervals in $\mathbb{R}$.

Proposition 12.1. *If φ and ψ are step functions on R, then $a\varphi + b\psi$, $\varphi\psi$, $\max\{\varphi, \psi\}$, and $\min\{\varphi, \psi\}$ are all step functions; thus, $\mathscr{S}(R)$ is an algebra.*

These are all clear if both come from the same partition, but this isn't *a priori* the case. This suggests the following definition.

Definition. A partition $\mathscr{R}$ is a refinement of $\mathscr{P}$ if the collection $\{x_0, \dots, x_N\}$ of vertices (that is, its endpoints) of $\mathscr{R}$ is contained within the vertices $\{y_0, \dots, y_K\}$ of $\mathscr{P}$.

This might not be intuitive, but the idea is that a refinement of a partition $\mathscr{P}$ is another partition given by partitioning each of the intervals into more pieces. The refinement may have more intervals.

Remark. If φ is a step function on R given by a partition $\mathscr{P}$, and $\mathscr{R}$ is a refinement of $\mathscr{P}$, then φ is also a step function given by $\mathscr{R}$: if $I \in \mathscr{P}$ and $J \subseteq I$ is in $\mathscr{R}$, then $\varphi = a_I$ on J^0, so define $b_J = a_I$ whenever $J \subseteq I$. Then,
$$\varphi = \sum_{J \in \mathscr{R}} b_J \chi_{J^0}.$$

[15]This notation is not completely standard, but appears in the course notes.

Another important point is that any two partitions $\mathscr{P}$ and $\mathscr{Q}$ share a common refinement, given by the partition taken from the union of their endpoints.

Proof of Proposition 12.1. This is clear if φ and ψ are given from the same partition (just look at the constants). In general, if they're given by different partitions $\mathscr{P}$ and $\mathscr{Q}$, then let $\mathscr{R}$ be a common refinement of $\mathscr{P}$ and $\mathscr{Q}$, so that φ and ψ are both step functions on $\mathscr{R}$, and thus the results hold. ⊠

Now, we can integrate step functions. If $\varphi \in \mathscr{S}(R)$, then write

$$\varphi = \sum_{I \in \mathscr{P}} a_I \chi_{I^0}$$

as usual; then, define its integral to be

$$\int_R \varphi = \sum_{I \in \mathscr{P}} a_I |I|,$$

where $|R| = (b_1 - a_1) \cdots (b_n - a_n)$ is the measure (generalized volume, including length when $n = 1$ and area when $n = 2$). However, we need to check that this is well-defined independently of the partition $\mathscr{P}$: if it's also possible to write $\varphi = \sum_{J \in \mathscr{Q}} b_J \chi_{J^0}$, then choose a common refinement $\mathscr{R}$ of $\mathscr{P}$ and $\mathscr{Q}$; then, write φ on $\mathscr{R}$ as

$$\begin{aligned}\sum_{K \in \mathscr{R}} c_K |L| &= \sum_{I \in \mathscr{P}} \left(\sum_{K \subseteq I} a_I |K| \right) \\ &= \sum_{I \in \mathscr{P}} a_I \left(\sum_{K \subseteq I} |K| \right) \\ &= \sum_{I \in \mathscr{P}} |I|,\end{aligned}$$

because the boundaries don't matter for the calculation of measure. But then, the same computation can be made for $\mathscr{Q}$ and $\mathscr{R}$, so the integrals agree on $\mathscr{P}$ and $\mathscr{R}$, and $\mathscr{Q}$ and $\mathscr{R}$, so therefore on $\mathscr{P}$ and $\mathscr{Q}$. If this doesn't make sense, drawing a picture helps, or look back at the one-dimensional case.

Proposition 12.2. *Let φ and ψ be step functions on an interval $R \subset \mathbb{R}^n$. Then:*

(1) If $\varphi \le \psi$ everywhere, then $\int_R \varphi \le \int_R \psi$.
(2) The integral is linear: if $a, b, \in \mathbb{R}$, then

$$\int_R a\varphi + b\psi = a \int_R \varphi + b \int_R \psi.$$

These are proven by once again taking a common refinement of their partitions.
Now, the general Riemann integral can be defined.

Definition. If f is bounded on R, define its upper Riemann integral to be

$$\overline{\int}_R f = \inf \left\{ \int_R \varphi : \varphi \in \mathscr{S}(R), \varphi \ge f \right\};$$

similarly, its lower Riemann integral is

$$\underline{\int}_R f = \sup \left\{ \int_R \varphi : \varphi \in \mathscr{S}(R), \varphi \le f \right\}.$$

Then, f is Riemann integrable if these quantities are equal, and its integral is defined to be this common value.

In essence, this definition approximates a well-behaved function f by step functions. The notion of "well-behaved" seems unclear, but in fact quite a large class of functions is Riemann integrable.

Theorem 12.3. *If f is continuous on R, then f is Riemann integrable.*

Proof. Since R is closed and bounded in $\mathbb{R}^n$, then it's compact, so f is also uniformly continuous; thus, for all $\varepsilon > 0$ there exists a $\delta > 0$ such that $|f(x) - f(y)| < \varepsilon$ if $|x - y| < \delta$. Then, choose a partition $\mathscr{P}$ of R such that diam$(I < \delta$ for all $I \in \mathscr{P}$, so if $\overline{f}_I = \sup_I f$ and $\underline{f}_I = \inf_I f$, then $\overline{f}_I - \underline{f}_I < \varepsilon$. Then, let

$$\varphi = \sum_{I \in \mathscr{P}} \underline{f}_I \chi_{I^0} \qquad \text{and} \qquad \psi = \sum_{I \in \mathscr{P}} \overline{f}_I \chi_{I^0},$$

so that $\varphi \le f \le \psi$, so

$$\overline{\int_R} f \le \int_R \psi = \sum_{I \in \mathscr{P}} \overline{f}_I |I| < \sum_{I \in \mathscr{P}} \underline{f}_I |I| + \varepsilon \sum_{I \in \mathscr{P}} |I|$$
$$= \int_R \varphi + \varepsilon |R| \le \underline{\int_R} f + \varepsilon |R|.$$

Thus, as $\varepsilon \to 0$, then the lower and upper integrals must be equal, so f is Riemann integrable. ⊠

In the spirit of this course, let's define a metric on the space $\mathscr{V}_{R,I}$ of Riemann integrable functions on R. Here are some nice properties of $\mathscr{V}_{R,I}$, left to the reader to check.

- $\mathscr{V}_{R,I}$ is a vector space.
- If $f, g \in \mathscr{V}_{R,I}$ such that $f \le g$, then $\int_R f \le \int_R g$.

This is nice: it works for continuous and piecewise continous functions, and it seems like we might be done here. Indeed, one can make $\mathscr{V}_{R,I}$ into a normed linear space, and therefore a metric space, with the L^1 norm

$$\|f\|_{L'} = \int_R |f|.$$

This metric space, sometimes also called $L^1(R)$, is very incomplete; there are many Cauchy sequences that don't converge.

Example 12.1. Let $f = \chi_{\mathbb{Q}}$, so the function is 1 on a given x iff x is rational. This is pretty clearly not Riemann integrable: it is 1 on a dense subset of $\mathbb{R}$, so any step function greater than f must be greater than 1 everywhere, so $\overline{\int_R} f = 1$, and similarly, it's 0 on a dense subset of $\mathbb{R}$, so any step function less than f must be less than 0 everywhere. Thus, $\underline{\int_R} f = 0$.

However, since $\mathbb{Q}$ is countable, one can take an ordering $x_1, x_2, x_3, \dots$ of it and then define

$$f_N(x) = \begin{cases} 1, & \text{if } x \in \{x_1, \dots, x_N\} \\ 0, & \text{otherwise.} \end{cases}$$

Then, each f_N is Riemann integrable and they all have the same L^1 metric, so they're a Cauchy sequence that ought to converge to $f(x)$ for all $x \in [0,1]$. But the limit doesn't lie in $L^1(R)$.

An awful lot of analysis requires complete spaces, which means looking for something better.

13. Measure Zero: 5/15/14

Today's lecture was given by Professor Leon Simon.

Recall that last time, we extended the notion of a closed interval to $\mathbb{R}^n$, where it was defined to be a product of closed intervals on $\mathbb{R}$; then, we defined partitions and used them to construct the Riemann integral. Today, these closed intervals will be called R. The following characterization for Riemann integration is fairly useful.

Lemma 13.1. *If $f : R \to \mathbb{R}$ is bounded, then f is Riemann integrable iff for all $\varepsilon > 0$ there exist step functions φ, ψ such that $\varphi \le f \le \psi$ and*

$$\int_R (\psi - \varphi) < \varepsilon.$$

In order to get to the most important results in Lesbegue theory, it'll be necessary to discuss the notion of measure zero.

Definition. In Riemann theory, a set $S \subseteq \mathbb{R}^n$ has volume zero if for all $\varepsilon > 0$, there exist open intervals $I_1, \dots, I_n$ such that

$$S \subset \bigcup_{j=1}^{n} I_j \quad \text{and} \quad \sum_{j=1}^{n} |I_j| < \varepsilon.$$

For the Lesbegue-theoretic notion, this is extended to infinite collections of intervals. This looks very similar, but makes a huge difference!

Definition. $S \subseteq \mathbb{R}^n$ has (Lesbegue) measure zero if for all $\varepsilon > 0$ there exist open intervals $I_1, I_2, \dots$ such that

$$S \subset \bigcup_{j=1}^{\infty} I_j \quad \text{and} \quad \sum_{j=1}^{\infty} |I_j| < \varepsilon.$$

Example 13.1. Look at the coordinate hyperplane $S = \{x = (x_1, \dots, x_n) \in \mathbb{R}^n : x_n = c\}$ for some $c \in \mathbb{R}$. Then, S has measure zero: let $\varepsilon > 0$ and let

$$I_j = (-j, j) \times \cdots \times (-j, j) \times \left(c - \frac{\varepsilon}{j^{n-1}2^{j+1}}, c + \frac{\varepsilon}{j^{n-1}2^{j+1}} \right),$$

so that $|I_j| = j^{n-1}2^{n-1}\varepsilon/j^{n-1}2^j = 2^{n-1-j}\varepsilon$, so $\sum |I_j| = 2^{n-1}\varepsilon < \varepsilon$.

Corollary 13.2. *If $I = [c_1, d_1] \times \cdots \times [c_n, d_n]$, then ∂I has measure zero.*

Proof.

$$\partial I \subset \bigcup_{j=1}^{n} \left(\{x : x_j = c_j\} \cup \{x : x_j = d_j\} \right),$$

so ∂I is a finite union of sets of measure zero, which can easily be shown to still have measure zero. ⊠

But we can do fancier things than finite unions.

Lemma 13.3. *Suppose $A_1, A_2, \dots$ is a sequence of sets on $\mathbb{R}^n$ such that each A_i has measure zero; then, $\bigcup_{i=1}^{\infty} A_i$ also has measure zero.*

Notice this is completely impossible for Riemann volume zero. Already this is a noticeable difference.

Proof of Lemma 13.3. Let $\varepsilon > 0$, and pick open intervals $I_{i,j}$, with $j = 1, 2, \dots$ with

$$A_i \subset \bigcup_{j=1}^{\infty} I_{i,j} \quad \text{and} \quad \sum_{j=1}^{\infty} |I_{i,j}| < \frac{\varepsilon}{2^i}.$$

Thus,

$$\bigcup_{i=1}^{\infty} A_i \subset \bigcup_{i,j=1}^{\infty} I_{i,j},$$

and

$$\begin{aligned} \sum_{i,j=1}^{\infty} |I_{i,j}| &= \sum_{i=1}^{\infty} \left(\sum_{j=1}^{\infty} |I_{i,j}| \right) \\ &< \sum_{i=1}^{\infty} \frac{\varepsilon}{2^i} = \varepsilon. \end{aligned}$$

This uses the fact that the rearrangement of an absolutely convergent series doesn't change its sum. ⊠

When Lesbegue was developing his theory, he astounded everyone by giving a precise criterion for which functions are Riemann integrable — nobody else had done this before. This is the content of the following theorem.

Theorem 13.4 (Lesbegue)**.** *Suppose $f : R \to \mathbb{R}$ is bounded. Then, f is Riemann integrable iff there exists a set $A \subset \mathbb{R}$ with Lesbegue measure zero such that f is continuous at each point of $R \setminus A$.*

In other words, f is Riemann integrable iff its set of discontinuities has measure zero.

One might try to recast this as "$f_{R \setminus A}$ is continuous;" this is implied by the theorem condition, but the converse is false: consider the function $f(x) = 1$ if $x \in \mathbb{Q}$ and 0 if $x \notin \mathbb{Q}$. Since $\mathbb{Q}$ is the countable union of singleton sets, then it has measure zero, so f is continuous on $\mathbb{R} \setminus \mathbb{Q}$, but f is continuous nowhere on $\mathbb{R}$. Don't conflate these: this function isn't Riemann integrable, because its upper integral on $[a, b]$ is $b - a$, and its lower integral is 0.

Proof of Theorem 13.4. In the forward direction, let $\breve{R}$ denote the interior of R; then, for every $x \in \breve{R}$, f is continuous at x iff for all $\varepsilon > 0$ there exists a $\delta > 0$ with $B_\delta(x) \subset \breve{R}$ and $|f(y) - f(x)| < \varepsilon$ for all $y \in B_\delta(x)$. This is further equivalent to saying that for all $\varepsilon > 0$ there exists a $\delta > 0$ with $B_\delta(x) \subset \breve{R}$ and $\sup_{B_\delta(x)} f - \inf_{B_\delta(x)} f < \varepsilon$, which is equivalent to saying that for all $\varepsilon > 0$, there's an open interval $I \subset \breve{R}$ with $x \in I$ and $\sup_I f - \inf_I f < \varepsilon$ (there is something to check here, but not much).

Thus, if f is discontinuous at some $x \in \breve{R}$, then there's an $\varepsilon_0 > 0$ such that for every open interval $I \subset \breve{R}$ such that $x \in I$, then $\sup_I f - \inf_I f \geq \varepsilon_0$.

Let $S = \{x \in \breve{R} : f$ is discontinuous at $x\}$; then, using the above result, one can write $S = \bigcup_{j=1}^{\infty} S_j$, where $S_j = \{x \in \breve{R} : \sup_I f - \inf_I f \geq 1/j$ for all open intervals $I \subset \breve{R}, x \in I\}$. (This doesn't deal with ∂R, but this has measure zero anyways, so this is all right.)

Now, if we can show that each S_j has measure zero, then S is a countable union of measure-zero sets and thus also has measure zero. By Lemma 13.1, since $f : R \to \mathbb{R}$ is Riemann integrable, then for all $\varepsilon > 0$, there exist step functions ψ, φ

such that $\varphi \le f \le \psi$ and $\int_R(\psi - \varphi) < \varepsilon/j$. Since φ and ψ are step functions, then there are partitions $\mathscr{Q}$ and $\mathscr{R}$ of R such that

$$\varphi = \sum_{I\in\mathscr{Q}} a_I \chi_{\breve{I}} \quad \text{and} \quad \psi = \sum_{J\in\mathscr{R}} b_J \chi_{\breve{J}}.$$

We don't care what the step functions do on the boundaries of the $J \in \mathscr{R}$ or $I \in \mathscr{Q}$, since these have measure zero anyways; in general, though, this is one of the subtleties of step functions. Let $\mathscr{P}$ be a common refinement of $\mathscr{Q}$ and $\mathscr{R}$; thus,

$$\varphi = \sum_{I\in\mathscr{P}} c_I \chi_{\breve{I}} \text{ on } R \setminus \bigcup_{I\in\mathscr{P}} \partial I, \text{ and}$$
$$\psi = \sum_{I\in\mathscr{P}} d_I \chi_{\breve{I}} \text{ on } R \setminus \bigcup_{I\in\mathscr{P}} \partial I.$$

Since $\varphi \le f$, then $c_I \le f$ on $\breve{I}$, and in particular $c_I \le \inf_{\breve{I}} f$. Similarly, $f \le \psi$, so $\sup_{\breve{I}} \le d_I$. Putting these together, $\sup_{\breve{I}} f - \inf_{\breve{I}} f \le d_I - c_I$ for all $I \in \mathscr{P}$.

$\int_R(\psi - \varphi) < \varepsilon/j$ is equivalent to

$$\begin{aligned}
&\sum_{I\in\mathscr{P}} (d_I - c_I)|I| < \frac{\varepsilon}{j} \\
\implies &\sum_{I\in\mathscr{P}} \left(\sup_{\breve{I}} f - \inf_{\breve{I}} f\right)|I| < \frac{\varepsilon}{j} \\
\implies \frac{1}{j} \sum_{\substack{I\in\mathscr{P}\\ \breve{I}\cap S_j\neq\emptyset}} |I| \le &\sum_{\substack{I\in\mathscr{P}\\ \breve{I}\cap S_j\neq\emptyset}} \left(\sup_{\breve{I}} f - \inf_{\breve{I}} f\right)|I| < \frac{\varepsilon}{j}.
\end{aligned}$$

Then,

$$S_j \subset \left(\bigcup_{\substack{I\in\mathscr{P}\\ \breve{I}\cap S_j\neq\emptyset}} \breve{I}\right) \cup \left(\bigcup_{I\in\mathscr{P}} \partial I\right).$$

The latter set has measure zero, as we already saw in Example 13.1, and the left side has volume less than ε, so S_j is contained in the union of countably many open intervals with total volume 2ε. Thus, S_j has measure zero, so S does as well (since it's a countable union of measure-zero sets).

Now on to the converse: let S be the set of discontinuities of f, which is of measure zero, and let $\varepsilon > 0$. Then, by definition, S can be covered by a countable collection of open intervals $I_1, I_2, \ldots$ such that the sum of the volumes of these intervals is less than ε. Let

$$K = R \setminus \bigcup_{j=1}^{\infty} I_j,$$

which is closed (since R is closed and the I_j are all open) and bounded (since R is), so it is compact. Thus, f is uniformly continuous on K: there exists a $\delta > 0$ such that $|f(x) - f(y)| < \varepsilon$ for all $x \in K$ and $y \in R$ such that $|x - y| < \delta$. This is a little bit stronger than the uniformity statement, so we'll have to show it.[16] Specifically, the usual theorem just speaks for $y \in K$.

So let's prove it: if this result fails, then this fails with $\delta = 1/j$, $j = 1, 2, \ldots$; that is, for every $j \in \mathbb{N}$, there exist points $x_j \in K$ and $y_j \in R$ such that $|x_j - y_j| < 1/j$, but $|f(x) - f(y)| > \varepsilon$. Thus, by Bolzano-Weierstrass, there exists a subsequence $\{x_{j_k}\}$ with $x_{j_k} \to x \in K$, since K is closed. But this implies that $|x_{j_k} - y_k| < 1/j_k < 1/k$, so $y_{j_k} \to x$ as well. But since f is continuous at x, since $x \in K$, then $f(x_{j_k}) \to f(x)$ and $f(y_{j_k}) \to f(x)$ as well, so $|f(x_{j_k}) - f(y_{j_k})| \to 0$ as well, which is a contradiction; thus, the claim is true.

We've only a couple of lines more: take a partition $\mathscr{P}$ such that $\operatorname{diam}(I) < \delta$ for every $I \in \mathscr{P}$. Then,

$$\sum_{I\in\mathscr{P}} \left(\sup_I - \inf_I f\right)|I| = \sum_{\substack{I\in\mathscr{P}\\ \breve{I}\cap K\neq\emptyset}} \left(\sup_I - \inf_I f\right)|I| + \sum_{\substack{I\in\mathscr{P}\\ \breve{I}\cap K=\emptyset}} \left(\sup_I - \inf_I f\right)|I|.$$

Now, let's eyeball these two carefully. For the first term, if $I \cap K \neq \emptyset$, then there's a $y_I \in I \cap K$, so

$$\begin{aligned}
\sup_I f - \inf_I f &= \sup_{z_1,z_2\in I}(f(z_1) - f(z_2)) \\
&= \sup_{z_1,z_2\in I}\left(f(z_1) - f(y_I) + f(y_I) - f(z_2)\right) \le 2\varepsilon.
\end{aligned}$$

For the second term, $\breve{I} \subset R \setminus \left(R \setminus \bigcup I_j\right)$, so it's covered by the I_j, and thus it can be bounded by $2\varepsilon(|R| + M)$. ⊠

[16]"It's not covered by the previous theorem... or for that matter, by K."

"When I was an undergraduate, I had an instructor who said, 'The key to doing a proof is often moving your mouth in the right way.' I haven't really moved my mouth in the right way here."

First, there will be a minor point of clarification regarding the WIM draft.

Definition. If (M,d) is a metric space and $X \subseteq M$, and all Cauchy sequences in X converge to some $x \in M$, then X is called pre-compact.

Then, it's "almost obvious" (but not actually that hard) to show that X is precompact iff $\overline{X}$ is compact.

Last time, we formulated and proved Lesbegue's theorem, that a function is Riemann integrable iff its discontinuities are a set of measure zero. Thus, these functions are continuous at a lot of points; however, in the Lesbegue sense of integration, the typical integrable function is nowhere continuous!

A typical approach to this (e.g. in Math 205A) constructs the Lesbegue integral with measure theory, but we'll take a different approach, due to Daniel and generalized by Stone, which constructs Lesbegue-integrable functions as appropriate limits of a set of functions (we'll use step functions, but it also works for continuous functions). This ends up being equivalent to the abstract completion of L^1, but this approach makes it much easier to approach theorems used all the time in analysis (the Monotone Convergence Theorem and so on).

First, we'll need two finicky lemmas to approach this completion. Recall that $\mathscr{S}(R)$ is the set of step functions on a closed interval R.

Lemma 14.1. *If $\{\varphi_k\}$ is a decreasing sequence in $\mathscr{S}(R)$ (i.e. for all $x \in R$, $\varphi_{k+1}(x) \le \varphi(x)$) with $\varphi_k \ge 0$, then $\int_R \varphi_k \to 0$ iff $\lim_{k\to\infty} \varphi_k(x) = 0$ almost everywhere.*

Lemma 14.2. *Suppose $\{\varphi_k\}$ is pointwise increasing and $\lim_{k\to\infty} \int_R \varphi_k$ is finite; then, $\lim_{k\to\infty} \varphi_k(x)$ is finite almost everywhere.*

Proof of Lemma 14.1. The forward direction turns out to be easier, so assume $\lim_{k\to\infty} \int_R k\varphi_k = 0$, and let $S = \{x : \lim_{k\to\infty} \varphi_k(x) > 0\}$. Thus, if $S_j = \{x : \lim \varphi_k(x) > 1/j\}$, then $S = \bigcup_{j=1}^n S_j$, so it suffices to show that S_j has measure zero.

In fact, for any $\alpha > 0$, define $S_\alpha = \{x : \lim \varphi_k(x) > \alpha\}$. For any $\varepsilon > 0$, there exists a k such that $\int_R \varphi_k < \varepsilon$. Let $\mathscr{P}$ be a partition for φ_k; then, let $\mathscr{F} = \{I \in \mathscr{P} : \varphi_k|_{I^0} > \alpha\}$, so that

$$S_\alpha \subseteq \left(\bigcup_{I\in\mathscr{F}} I^0\right) \cup \left(\bigcup_{I\in\mathscr{P}} \partial I\right).$$

But we know that ∂I has measure zero. Thus, by the definition of the integral for a step function,

$$\alpha \sum_{I\in\mathscr{F}} |I| < \int_R \varphi_k < \varepsilon,$$

so $\sum_{I\in\mathscr{F}} |I| < \varepsilon/\alpha$. And since

$$\bigcup_{I\in\mathscr{P}} \partial I \subseteq \bigcup_{j=1}^\infty I_j,$$

then $\sum_{j=1}^\infty |I_j| < \varepsilon$, so it's possible to cover S_α by the set

$$\mathscr{U} = \{I^0 : I \in \mathscr{F}\} \cup \{I_j : j = 1,2,\dots\},$$

and

$$\sum_{I\in\mathscr{U}} |I| < \frac{\varepsilon}{\alpha} + \varepsilon,$$

which can be made arbitrarily small as $\varepsilon \to 0$, so S_α has measure zero, and thus all of the S_j do.

In the other direction, assume that $\{x : \lim \varphi_k(x) > 0\}$ has measure zero, and we want to show that $\lim \int_R \varphi_k = 0$. This will use a compactness argument, and depend on the following lemma.

Lemma 14.3. *Let $\{f_k\}$ be a monotone and bounded sequence of continuous functions on a compact set K, and suppose that $f(x) = \lim_{k\to\infty} f_k(x)$ is continuous.*[17] *Then, $\{f_k\}$ converges uniformly.*

This will be proven on the homework.

Returning to the proof, notice that if $\mathscr{P}_k$ is a partition for φ_k, then

$$\widehat{S} = S \cup \bigcup_{\substack{I\in\mathscr{P}_k \\ k\in\mathbb{N}}} \partial I$$

[17] This is not true in the general case, and it's easy to cook up counterexamples.

still has measure zero, because S does and the rest is a countable collection of boundaries ∂I, which also have measure zero. Thus, there's an open set $\mathscr{U}$ such that for any $\varepsilon > 0$,

$$\widehat{S} \subseteq \bigcup_{I \in \mathscr{U}} I \qquad \text{and} \qquad \sum_{I \in \mathscr{U}} |I| < \varepsilon.$$

Let $K = R \setminus \left(\bigcup_{I \in \mathscr{O}} I\right)$; then, it's compact, and each φ_k is locally constant on K (and therefore continuous), and $\lim \varphi_k(x) = 0$ for all $x \in K$. Thus, the sequence converges uniformly by Lemma 14.3, which is sometimes written $\varphi_k \rightrightarrows 0$.

Thus, there exists a k large enough such that $\varphi_k(x) < \varepsilon$ for all $x \in K$, so for any $x \in K$, there exists an $I_x \in \mathscr{P}_k$ with $x \in I_x$. Thus,

$$\begin{aligned}
\int_R \varphi_k &= \sum_{I^0 \cap K \neq \emptyset} (\varphi_k|_{I^0})\, |I| + \sum_{I^0 \cap K = \emptyset} (\varphi_k|_{I^0})\, |I| \\
&< \varepsilon |R| + M \sum_{I \in \mathscr{U}} |I| \\
&< \varepsilon (|R| + M).
\end{aligned}$$

There's just a little more to be said here, but this polishes off the other direction. ⊠

Proof of Lemma 14.2. Suppose $\{\varphi_k\}$ is increasing and $\lim_R \varphi_k$ is finite, and let $S = \{x : \lim \varphi_k(x) \text{ is infinite}\}$; then, we want to show that S has measure zero.

Since we have a convergent sequence, we can choose a subsequence which converges as fast as we want; so we'll choose one that converges very fast. That is, there exists a subsequence $\{\varphi_{k_j}\}$ such that for all j,

$$\int_R \left(\varphi_{k_{j+1}} - \varphi_{k_j}\right) < 2^{-2j}.$$

Let $\alpha \gg 1$, and let $S_j = \{x : \varphi_{k_{j+1}}(x) - \varphi_{k_j}(x) \geq 2^{-j}\alpha\}$, so that

$$S \subseteq \bigcup_{j=1}^{\infty} S_j,$$

because if $x \in S$ and $x \notin S_j$ for any j, then

$$\sum_{j=1}^{\infty} \left(\varphi_{k_{j+1}}(x) - \varphi_{k_j}(x)\right) < 2^{-j}\alpha,$$

but this is a telescoping series, so it's easily seen to simplify to $(\lim \varphi_k(x)) - \varphi_{k_1}(x) < \alpha$, which is a contradiction. Thus, the S_j cover S.

Let $\mathscr{P}_j$ be a partition such that $\varphi_{k_{j+1}} - \varphi_{k_j}$ is constant on I^0 for all $I \in \mathscr{P}_j$, and define

$$\mathscr{F}_j = \left\{ I \in \mathscr{P}_j : \left(\varphi_{k_{j+1}} - \varphi_{k_j}\right)|_{I^0} \geq 2^{-j}\alpha \right\}.$$

Thus, there's a covering

$$S_j \subseteq \left(\bigcup_{I \in \mathscr{F}_j} I^0\right) \cup \left(\bigcup_{I \in \mathscr{P}_j} \partial I\right).$$

But the second term has measure zero and the first one's measure satisfies

$$2^{-j}\alpha \sum_{I \in \mathscr{F}_j} |I| \leq \int_R \varphi_{k_{j+1}} - \varphi_{k_j} < 2^{-2j},$$

which implies that

$$\begin{aligned}
&\sum_{j=1}^{\infty} \left(\sum_{I \in \mathscr{F}_j} |I|\right) < \sum_{j=1}^{\infty} \alpha^{-1} 2^{-j}, \\
\implies &\sum_{I \in \bigcup_{j=1}^{\infty} \mathscr{F}_j} |I| < \alpha^{-1}.
\end{aligned}$$

Since α is arbitrarily large, then this is arbitrarily small, so this covering can be made arbitrarily small, and thus S has measure zero. ⊠

One nice property of this approach is that it's relatively easy to characterize the Riemann integrable functions in terms of the Lesbegue ones, even easier than Lesbegue's theorem.

15. The Lesbegue Integral II: 5/22/14

Proposition 15.1. *If $J_1, \dots, J_N$ are disjoint and open, and $I_1, I_2, \dots$ are open, then if*

$$\bigcup_{i=1}^{N} J_i \subseteq \bigcup_{k=1}^{\infty} I_k,$$

then

$$\sum_{i=1}^{N} |J_i| \le \sum_{k=1}^{\infty} |I_k|.$$

Proof. Choose some $\widehat{J}_i \subseteq J_i$ such that $\widehat{J}_i$ is closed and $|\widehat{J}_i| > |J_i| - \varepsilon$. Then, since a finite union of finite closed intervals is compact, then there are $I_1, \dots, I_K$ such that

$$\bigcup_{i=1}^{N} \widehat{J}_i \subseteq \bigcup_{k=1}^{K} I_k.$$

Now, consider the characteristic functions $\widehat{\chi}_i(x) = 1$ if $x \in \widehat{J}_i$ and 0 if not, and $\chi_k(x) = 1$ if $x \in I_k$ and 0 if not. Then, since these are step functions, then we can use our previous results on step functions to show that $\sum_{i=1}^{N} \widehat{\chi}_i \le \sum_{k=1}^{K} \chi_{I_k}$, and therefore $\int_R \sum_{i=1}^{N} \widehat{\chi}_i \le \int_R \sum_{k=1}^{K} \chi_{I_k}$, so as $\varepsilon \to 0$,

$$\begin{aligned} \sum_{i=1}^{N} |J_i| - N\varepsilon &\le \sum_{i=1}^{N} |\widehat{J}_i| \\ &\le \sum_{k=1}^{K} |I_k| \\ &\le \sum_{k=1}^{\infty} |I_k|. \end{aligned}$$ ⊠

This is sufficient to finish off the two lemmas from last lecture.

Now, we can define the notion of Lesbegue-integrable functions as the completion of the space of step functions. First, define $\mathscr{S}_+(R) = \{\varphi \in \mathscr{S}(R) : \varphi \ge 0\}$, and then let $f \in L_+(R)$ if there exists an increasing sequence $\{\varphi_k\} \subseteq \mathscr{S}_+(R)$ with $\lim_{k\to\infty} \int_R \varphi_k$ is finite, and such that $\lim_{k\to\infty} \varphi_k(x) = f(x)$ almost everywhere.

It's not hard to see that if $f, g \in L^+$ and $\alpha, \beta \ge 0$, then $\alpha f + \beta g \in L^+$, as are $\max\{f, g\}$ and $\min\{f, g\}$.

Definition. If $f \in L^+$ then there is a sequence $\{\varphi_k\}$ in $\mathscr{S}_+(R)$ that converges to f. Then,

$$\int_R f = \lim_{n\to\infty} \int_R \varphi_k.$$

Claim. $\int_R f$ is well-defined.

Proof. Suppose $\{\varphi_k\}$ and $\{\psi_k\}$ are two increasing sequences on $\mathscr{S}_+$ whose integrals converge to some finite value, and such that $\lim_{k\to\infty} \varphi_k(f) = \lim_{k\to\infty} \psi_k(x)$ almost everywhere. Then, we wish to show that $\lim \int_R \varphi_k = \lim \int_R \psi_k$.

Introduce the notation $f_+ = \max\{f, 0\}$ and $f_- = \max\{-f, 0\}$. Then, for a fixed k, $(\varphi_k - \psi_\ell)_+$ tends to 0 (in ℓ) almost everywhere (because $\{\psi_\ell\}$ is increasing to a limit that φ_k is strictly smaller than), and thus by Lemma 14.1, $\lim_{\ell\to\infty} (\varphi_k - \psi_\ell)_+ = 0$. Thus, since $f \le f_+$, then

$$\lim_{\ell\to\infty} \int_R (\varphi_k - \psi_\ell) \le \lim_{\ell\to\infty} \int_R (\varphi_k - \varphi_\ell)_+ = 0,$$

so for all k, $\int_R \varphi_k \le \lim_{\ell\to\infty} \int_R \psi_\ell$. Thus, $\lim \int_R \varphi_k \le \lim \int_R \psi_\ell$; then, to get the opposite inequality, switch φ_k and ψ_ℓ in the above proof, since the initial choice was arbitrary; thus, the limits are in fact equal. ⊠

Definition. $L^1(R) = L_+(R) - L_+(R) = \{f : f = g - h, g, h \in L_+\}$.

This will be a complete space, which we'll have to prove. But right now, we can see that if $f, g \in L^1$ and $\alpha, \beta \ge 0$, then $\alpha f + \beta g \in L^1$ and $-f \in L^1$, so L^1 is a vector space.

Definition. If $f \in L^1(R)$, so that $f = g - h$ for $g, h \in L_+(R)$, then $\int_R f = \int_R g - \int_R h$.

Claim. This integral is also well-defined.

Proof. This is a bit easier than last time: suppose $f \in L^1$ can be written as $f = g - h = \widetilde{g} - \widetilde{h}$. Thus, $g + \widetilde{h} = \widetilde{g} + h$, so

$$\begin{aligned} \int_R g + \int_R \widetilde{h} &= \int_R \widetilde{g} + \int_R h \\ \implies \int_R g - \int_R h &= \int_R \widetilde{g} - \int_R \widetilde{h}. \end{aligned} \qquad \boxtimes$$

One might expect that if $f \in L^1$, then $f = f_+ - f_-$, so $f_+, f_- \in L^+$, but this is *not* true! This will rear its head on the homework.

It's useful to have some easy properties of the Lesbegue integral.

- If $f = 0$ almost everywhere, then $f_+, f_- \in L_+(R)$ and $\int_R f = \int_R f_- = \int_R f_+ = 0$. This is because f can be approximated by the step functions $\varphi_k = 0$.
- If $f, \widetilde{f} \in L^1$ are such that $f = \widetilde{f}$ almost everywhere, then $\int_R f = \int_R \widetilde{f}$, since $f = \widetilde{f} + (f - \widetilde{f}) = (f - \widetilde{f}) + -(f - \widetilde{f})_-$, but the latter two are zero almost everywhere, then the integrals must be the same.
- If $f \geq g$, then $\int_R f \geq \int_R g$.
- If $f \in L^1$, then $|f| \in L^1$, because if $f = g - h$ for $g, h \in L^+$, then $|f| = \max\{g, h\} - \min\{g, h\}$, both of which are in L_+.
- Linearity of the integral: if $f, g \in L^1$ and $\alpha, \beta \in \mathbb{R}$, then
$$\int_R \alpha f + \beta g = \alpha \int_R f + \beta \int_R g.$$

In some sense, if two functions agree almost everywhere, then Lesbegue theory considers them to be identical.

We will show that $L^1(R)$ is complete in the L^1 norm $\|f\|_1 = \int_R |f|$. Technically, this isn't a norm: it satisfies nonnegativity and the triangle inequality, and it's nonnegative, but there exist $f, g \in L^1$ such that $f \neq g$ but $\|f - g\|_1 = 0$ (this is true iff $f = g$ almost everywhere). Thus, this is called a seminorm.[18]

Returning to the Riemann integral, define $\mathscr{R}(R)$ to be the set of $f : R \to \mathbb{R}$ that are Riemann integrable. We showed this is a vector space with similar basic properties to L^1 (i.e. linearity, maximum and minimum properties, and so forth). However, this space is not complete: $\mathscr{R}(R)$ is much bigger than the space of continuous functions, but it's not big enough.

Theorem 15.2. *Let $\widetilde{L}_+(R)$ be the set of functions that can be approximated by (in the sense that they're pointwise limits of) an increasing sequence of step functions (but not necessarily nonnegative), and $\widetilde{L}_-(R)$ is the same, but for decreasing sequences of step functions.*

Then,

(1) $\mathscr{R}(R) \subseteq \widetilde{L}_+(R) \cap \widetilde{L}_-(R)$.
(2) If $f \in \widetilde{L}_+ \cap \widetilde{L}_-$, then there exists an $\widetilde{f} \in \mathscr{R}(R)$ such that $f = \widetilde{f}$ almost everywhere.

In some sense, this just boils down to the notation, but it's an interesting characterization: Lesbegue functions are those that can be approximated on one side, and Riemann functions require approximating on both sides.

On the homework, we'll show that if $K \subseteq R$ is closed, then its characteristic function $\chi_K(x) \in \widetilde{L}_-(R)$, even though it's positive, and there are some K for which it's not Riemann integrable.

Proof of Theorem 15.2. For (1), let $f \in \mathscr{R}(R)$, so that for any $\varepsilon > 0$, there exist step functions φ, ψ such that $\varphi \leq f \leq \psi$, with $\int_R \psi - \int_R \varphi < \varepsilon$. Thus, for any k there exist φ_k, ψ_k such that $\varphi_k \leq f \leq \psi_k$ and $\int_R \varphi_k - \int_R \varphi_k < 1/k$, so the sequences $\{\varphi_k\}$ and $\{\psi_k\}$ are what we want, except that they're not monotone.

Thus, we'll use the lattice property: let $\widetilde{\varphi}_k = \max\{\varphi_1, \dots, \varphi_k\}$, and $\widetilde{\psi}_k = \min\{\psi_1, \dots, \psi_k\}$, so they're monotone in the way we want and $\widetilde{\varphi}_k \leq f \leq \widetilde{\psi}_k$, and since $\widetilde{\varphi}_k \geq \varphi_k$ and $\widetilde{\psi}_k \leq \psi_k$, then

$$\int_R \widetilde{\psi}_k - \widetilde{\varphi}_k \leq \int_R \psi_k - \varphi_k < \frac{1}{k}.$$

Thus,

$$\lim_{k\to\infty} \int_R \left(\widetilde{\psi}_k - \widetilde{\varphi}_k\right) = 0,$$

and the function $\lim_{k\to\infty} \left(\widetilde{\psi}_k - \widetilde{\varphi}_k\right) = 0$ almost everywhere. Thus, $\lim_{k\to\infty} \widetilde{\psi}_k = \lim_{k\to\infty} \widetilde{\varphi}_k = f$ almost everywhere, so $f \in \widetilde{L}_+(R)$ and $f \in \widetilde{L}_-(R)$.

[18] To turn this into a real norm, one has to define $f \sim g$ if $f = g$ almost everywhere, and then define the norm on equivalence classes of functions.

Thus, (1) is done, so look at (2), a sort of converse. If $f \in \widetilde{L}_- \cap \widetilde{L}_+$, then there exists an increasing sequence $\{\varphi_k\}$ and a decreasing sequence $\{\psi_k\}$ that both converge to f almost everywhere. Let $\mathscr{P}_k$ be a partition simultaneously for φ_k and ψ_k; thus, there exists an $M > 0$ such that we can modify these to

$$\widetilde{\varphi} + k(x) = \begin{cases} \varphi_k(x), & x \in \bigcup_{I \in \mathscr{P}_k} I \\ -M, & x \in \bigcup_{I \in \mathscr{P}_k} \partial I, \end{cases} \qquad \text{and} \qquad \widetilde{\psi}_k(x) = \begin{cases} \psi_k(x), & x \in \bigcup_{I \in \mathscr{P}_k} I \\ M, & x \in \bigcup_{I \in \mathscr{P}_k} \partial I. \end{cases}$$

Thus, $\widetilde{\varphi}_k \le \widetilde{\psi}_\ell$ for all k, ℓ, but we've completely destroyed monotonicity. This will be dealt with next time.

16. The Convergence Theorems: 5/27/14

"I blame the blackboard."

Before stating and proving the convergence theorems, there's just a bit more to be done on the characterization of Riemann-integrable functions in Theorem 15.2, i.e. that the intersection of $\widetilde{L}_+(R) \cap \widetilde{L}_-(R)$ is essentially the same as $\mathscr{R}(R)$, the set of Riemann integrable functions (where "essentially" means that there's an $\widetilde{f} \in \mathscr{R}(R)$ such that $\widetilde{f} = f$ almost everywhere).

The definitions of these $\widetilde{L}_+(R)$ and $\widetilde{L}_-(R)$ are that there's an increasing (resp. decreasing) sequence of step functions whose limit is f almost everywhere, bounding f almost everywhere on one side.

Let $\mathscr{P}_k$ be a partition for φ_k and

$$\widetilde{\varphi}_k(x) = \begin{cases} \varphi_k, & x \in I \text{ for some } I \in \mathscr{P}_k \\ -M, & x \in \partial I \text{ for some } I \in \mathscr{P}_k. \end{cases}$$

Construct $\widetilde{\psi}_k$ in the same way, but with M instead of ∂M; here, M is a bound we can place on f. Thus, $\widetilde{\varphi}_k \le \widetilde{\psi}_j$ for all j and k if M is sufficiently large, so $\lim \widetilde{\psi}_k = f$ almost everywhere, as with $\lim \widetilde{\psi}_k$.

However, these sequences aren't monotone yet. The standard trick will suffice:

$$\widehat{\varphi}_k = \max\{\widetilde{\varphi}_1, \dots, \widetilde{\varphi}_k\}.$$

Thus, these still converge to $\widetilde{f}$ almost everywhere, but $\{\widehat{\varphi}_k\}$ is increasing and $\int_R \widehat{\psi}_k \le \int_R \widetilde{\psi}_k$. Then, by one of the lemmas proved a few lectures ago, since $\widehat{\varphi}_k \to \widetilde{f}$ almost everywhere and $\int_R \widehat{\varphi}_k \to \int_R \widetilde{f}$, then $\widetilde{f} = f$ almost everywhere. Then, since for all j and k $\widehat{\varphi}_j \le \widetilde{f} \le \widetilde{\psi}_k$ and $\int_R (\widetilde{\psi}_k - \widehat{\varphi}_k) \to 0$ as $k \to \infty$, then $\widetilde{f}$ is Riemann integrable.

The Convergence Theorems. These theorems concern themselves with sequences of functions, and ask when the limit is integrable, and whether it's equal to the limit of the integrals.

One unfortunate aspect of Riemann integrability is that there are sequences of continuous, smooth, etc. functions that converge to discontinuous functions, and so on. In particular, there are sequences of Riemann-integrable functions with non-Riemann-integrable limits. This is not true for the Lesbegue integral: the limit of a series of Lesbegue-integrable functions ought to still be Lesbegue integrable, but the limit of the integrals might not be the integral of the limit.

Example 16.1. On $R = [0, 1]$, consider the function $f_k(x)$ that is given by the isosceles triangle with area 1 and base $[0, 2/k]$. Then, $\lim f_k(x) = 0$ (check pointwise), but $\int_R f_k = 1$, so $\lim \int_R f_k \ne \int_R (\lim f_k)$.

The above is a useful example to have in mind; for the physics majors in the audience, this sequence converges in some sense to a delta "function" (actually a measure): it's concentrated at the origin, with infinite mass at that point, but with total integral 1.

Theorem 16.1 (Monotone Convergence Theorem)**.** *Suppose $\{f_k\} \subseteq L^1(R)$. If $\{f_k\}$ is monotone and $\left\{\int_R f_k\right\}$ is uniformly bounded, then there exists an $f \in L^1(R)$ such that $\lim f_k = f$ and $\int_R f = \lim \int_R f_k$.*

Example 16.1 demonstrates a sequence that isn't monotone, but otherwise satisfies the theorem.

Lemma 16.2 (Fatou)**.** *If $\{f_k\} \subseteq L^1(R)$, $\left\{\int_R |f_k|\right\}$ is bounded, and $\lim f_k = f$ almost everywhere, then $f \in L^1(R)$. Additionally, if $f_k \ge 0$, then*

$$\int_R f \le \varliminf \int_R f_k.$$

This is remarkably convenient: it allows one to prove that a function is in $L^1(R)$ by constructing a nice sequence of functions to approximate it.

Theorem 16.3 (Dominating Convergence Theorem)**.** *If $\{f_k\} \subseteq L^1(R)$ is such that $\lim f_k = f$ almost everywhere and there exists an $F \in L^1(R)$ that dominates f_k (i.e. for all k, $|f_k| \le F$ almost everywhere), then $f \in L^1(R)$ and $\int_R f = \lim \int_R f_k$.*

In Example 16.1, there is no uniformly integrable F that dominates the functions f_k (which isn't entirely obvious).

The Monotone Convergence Theorem is the hardest to prove, so we'll assume it for now in order to prove the others.

Proof of Lemma 16.2. Once again define $f_{k+} = \max\{f_k, 0\}$ and $f_{k-} = -\min\{f_k, 0\}$; then, both $\{f_{k+}\}$ and $\{f_{k-}\}$ satisfy the hypotheses of the lemma, since $\int_R f_{k\pm} \le \int_R |f_k|$ and $\lim f_{k\pm} = f_\pm$ almost everywhere. Thus, it suffices to consider $f_k \ge 0$, because $f \in L^1(R)$ iff $f_+, f_- \in L^1(R)$.

Note that if $\lim f_k(x) = f(x)$, then this is also the lim inf: $\underline{\lim} f_k(x) = f(x)$. But this is related to monotone sequences, which is the key: if $F_k = \inf\{f_k, f_{k+1}, \dots\}$, then $\{F_k\}$ is increasing, and $f(x) = \lim F_k(x)$ almost everywhere. Then, let $F_{k,j} = \inf\{f_k, \dots, f_j\}$, so that $\{F_{k,j}\}$ is decreasing in j and $F_k = \lim F_{k,j}$. Since the infimum of a finite set of Lesbegue-integrable functions is still integrable, then all of these $F_{k,j} \in L^1(R)$.

Thus, by Theorem 16.1, $F_k \in L^1(R)$ and

$$\int_R F_k = \lim_{j\to\infty} \int_R F_{k,j}.$$

Then, since

$$\int_R F_{k,j} \le \min\left\{\int_R f_k, \dots, \int_R f_j\right\},$$

so passing to the limit,

$$\int_R F_K \le \inf\left\{\int_R f_k, \int_R f_{k+1}, \dots\right\}.$$

Furthermore, by hypothesis, these functions are uniformly bounded. Thus, we're in the situation to apply Theorem 16.1 again, to show that $\lim F_k = \underline{\lim} f_k = f$ almost everywhere and $\int_R f = \lim \int_R f_k = \underline{\lim} \int_R f_k$. ⊠

The idea is to apply the Monotone Convergence Theorem first to the tail end of the sequence, and then to the entire sequence.

Then, the Dominating Convergence Theorem follows relatively cleanly from Fatou's lemma.

Proof of Theorem 16.3. Since F dominates f_k, then $\{F - f_k\}$ and $\{F + f_k\}$ are nonnegative, so (with the other hypotheses satisfied) by Lemma 16.2, then $F - f \in L^1(R)$ and $F + f \in L^1(R)$, and

$$\begin{aligned}\int_R F - f &\le \underline{\lim}_{k\to\infty} \int_R F - f_k = \int_R F + \underline{\lim} \int_R -f_k \\ &= \int_R f - \overline{\lim} \int_R f_k. \\ \int_R F + f &\le \underline{\lim} \int_R F + f_k = \int_R F + \underline{\lim} \int_R f_k.\end{aligned}$$

Thus,

$$\overline{\lim} \int_R f_k \le \int_R f \le \underline{\lim} \int_R f_k,$$

so the lim sup and lim inf must agree, and thus the sequence converges. ⊠

The Dominating Convergence Theorem is probably the most commonly used of all three theorems.

To prove the Monotone Convergence Theorem, it's necessary to return to the basic definition of $L^1(R)$.

Proof of Theorem 16.1. We will first prove the theorem for $\{f_k\} \subseteq L_+(R)$. Then, there exist sequences $\{\varphi_{k,j}\} \subseteq \mathscr{S}_+(R)$ such that $\lim_{j\to\infty} \varphi_{k,j} = f_k$ almost everywhere, and

$$\int_R f_k = \lim_{j\to\infty} \int_R \varphi_{k,j}.$$

Furthermore, by starting further along the series, one can assume $\varphi_k = \varphi_{k,1}$ satisfies $\int_R f_k - \varphi_k < 2^{-k}$, (starting with a better approximation). Thus, for all $j \in \mathbb{N}$,

$$\int_R f_k - \varphi_{k,j} \le \int_R f_k - \varphi_{k,1} < 2^{-k}.$$

In some sense, the immediate goal is to bound f_k. Let

$$\psi_j = \sum_{k=1}^{j} \left(\varphi_{k,j} - \varphi_k\right),$$

so it's an increasing sequence in $\mathscr{S}_+(R)$. Additionally,

$$\int_R \psi_j \leq \sum_{k=1}^{j} \int_R \left(\varphi_{k,j} - \varphi_k\right) < 1.$$

Thus, by one of the lemmas we proved earlier, this bound implies a pointwise bound C_x on

$$\sum_{k=1}^{\infty} \varphi_{k,j}(x) - \varphi_k(x) \leq C_x$$

for almost every x. Thus, this is true (with the same constant) for each finite partial sum, since every term is positive:

$$\sum_{k=1}^{N} \varphi_{k,j}(x) - \varphi_k(x) \leq C_x,$$

so letting $j \to \infty$, one obtains the N^{th} partial sum for

$$\sum_{k=1}^{N} f_k(x) - \varphi_k(x) \leq C_x,$$

so the above bound is also true as $N \to \infty$:

$$\sum_{k=1}^{\infty} f_k(x) - \varphi_k(x) \leq C_x.$$

In particular, for almost all x, $\lim(f_k(x) - \varphi_k(x)) = 0$.

Now, let $\psi_k = \max\{\varphi_1, \dots, \varphi_k\}$, so that $\psi_k \leq f_k$ for all k (since each φ_k is); thus, ψ_k is an increasing sequence of nonnegative step functions uniformly bounded: $\int_R \psi_k \leq \int_R f_k$. Thus, there exists an f such that $f = \lim \psi_k$, and moreover, $\int_R f = \lim \int_R \psi_k$. The proof will be continued tomorrow; there's not much more to show.

17. $L^1(R)$ is Complete: 5/29/14

"I understand these arguments are a bit demanding, but they're also demanding for me to present. But you don't need to feel sorry for me."

Recall that we were in the middle of proving the Monotone Convergence Theorem, Theorem 16.1.

The first case that we proved, where $\{f_k\} \subseteq L_+(R)$. We know via one of the basic lemmas that this theorem holds if $\{f_k\} \in \mathscr{S}_+(R)$, so in general approximate $\{f_k\}$ by $\{\varphi_k\} \subseteq \mathscr{S}_+(R)$ such that $\int_R f_k - \varphi_k < 2^{-k}$ and $\varphi_k \leq f_k$ almost everywhere; this implies that $\lim f_k - \varphi_k = 0$ almost everywhere.

The φ_k aren't monotone yet, so the usual trick will work: let $\psi_k = \max\{\varphi_1, \dots, \varphi_k\}$, so that ψ_k is a sequence of increasing step functions, and since f_k is monotone, then $\varphi_k \leq \psi_k \leq f_k$ almost everywhere. Thus, $\int_R \psi_k \leq \int_R f_k$ acts as a uniform bound. Thus, by one of the lemmas, there exists an $f \in L_+(R)$ such that $\lim \psi_k(x) = f(x)$ almost everywhere and $\lim \int_R \psi_k = \int_R f$.

However, $f_k - \psi_k \leq f_k - \varphi_k$, and the latter goes to zero almost everywhere, so $f(x) = \lim f_k(x)$ almost everywhere. In particular, this implies that $\int_R f_k - \psi_k < 2^{-k}$, so $\int_R f = \lim \int_R f_k$.

Then, the second case, where the sequence $\{f_k\}$ is generally in $L^1(R)$, with $\lim \int_R f_k$ finite, and define $f_0 = f_1$ for the purpose of indices of summation. Then, there exist $g_k, h_k \in L_+(R)$ for each k such that $f_k = g_k - h_k$, and since f is monotone, then these can be chosen such that $f_k - f_{k-1} = g_k - h_k$. This means that f_k can be recovered from the k^{th} partial sum.

Choose $\varphi_k \in \mathscr{S}_+(R)$ with $\varphi_k \leq h_k$ almost everywhere and such that $\int_R h_k - \varphi_k < 2^{-k}$. Thus, $g_k - \varphi_k$ is also nonnegative, as are $h_k - \varphi_k$ (almost everywhere), so

$$\begin{aligned} g_k - \varphi_k &= (g_k - h_k) + (h_k - \varphi_k) \\ &= (f_k - f_{k-1}) + (h_k - \varphi_k), \end{aligned}$$

and this is greater than or equal to 0 almost everywhere. Thus, $g_k - \varphi_k \in L^+(R)$, so one can apply the first case to it: the function $G = \sum_{k=1}^{\infty} g_k - \varphi_i$ is in $L_+(R)$ and

$$\sum_{k=1}^{\infty} \int_R (g_k - \varphi_k) = \int_R G.$$

But the series is telescoping, so it converges:

$$\int_R (g_k - \varphi_k) = \int_R (f_k - f_{k-1}) + \left| \int_R h_k - \varphi_k \right|,$$

and the second term is bounded by 2^{-k}, so the sequence $g_k - \varphi_k$ is an increasing sequence of functions in $L_+(R)$ with bounded integral. The same thing can be done to $h_k - \varphi_k$: it's in $L_+(R)$ and the integrals converge to that of $H = \sum_{k=1}^\infty h_k - \varphi_k \in L_+(R)$:

$$\int_R H = \sum_{k=1}^\infty \int_R h_k - \varphi_k.$$

Case 1 has been used twice, but now we're almost done:

$$\begin{aligned}\sum_{k=1}^N \big((g_k - \varphi_k) - (h_k - \varphi_k)\big) &= g_k - h_k \\ &= \sum_{k=1}^N (f_k - f_{k-1}) \\ &= f_N - f_1.\end{aligned}$$

Thus,

$$f_N = \sum_{k=1}^N (g_k - \varphi_k) - \sum_{k=1}^N (h_k - \varphi_k).$$

Thus, $f_N \to G - H + f_1$ almost everywhere, but this is equal to f, so it's in $L^1(R)$, and $\int_R f_n \to \int_R f$. ⊠

Notice that the same tricks come up in all of these proofs: the telescoping trick is new, but the differences of sequences, etc. should look familiar.

The Lesbegue integral makes sense on other compact regions $X \subset \mathbb{R}^n$, i.e. by extending to a rectangle $R \supseteq X$ and defining $f = 0$ on $R \setminus X$. It's also possible to generalize to noncompact subsets, but this is harder and we won't get into it.

The point of constructing the Lesbegue integral was to make the space of integrable functions complete. We can now go back and prove that this works.

Recall that we defined a "norm" (actually a seminorm) on $L^1(R)$, where $\|f\|_1 = \int_R |f|$. Then, $\|f\|_1 = 0$ iff $f = 0$ almost everywhere. Since nonzero functions exist that are zero almost everywhere, then this isn't truly a norm (so it can be made into one by defining it on equivalence classes of functions), but it satisfies all of the other norm axioms: $\|f\|_1 \geq 0$, and it's homogeneous: if $\alpha \in \mathbb{R}$, then $\|\alpha f\|_1 = |\alpha|\|f\|_1$. Finally, the triangle inequality is met:

Theorem 17.1. *If $\{f_k\}$ is Cauchy in $L^1(R)$, then there exists an $f \in L^1(R)$ with $\lim \|f_k - f\|_1 = 0$, i.e. $\lim f_k = f$ in the L^1-norm. Moreover, $\{f_k\}$ has a subsequence $\{f_{k_j}\}$ such that $\lim_{j\to\infty} f_{k_j}(x) = f(x)$ for almost all x.*

Proof. Since $\{f_k\}$ is Cauchy, then there exists an $\{f_{k_j}\}$ such that $\|f_{k_{j+1}} - f_{k_j}\|_1 < 2^{-j}$ for $j = 1, 2, 3, \ldots$, as shown on the homework. Thus, by Theorem 16.1, it's possible to construct the functions

$$\begin{aligned} g &= \sum_{j=1}^\infty \left(f_{k_{j+1}} - f_{j_k}\right)_+ \\ h &= \sum_{j=1}^\infty \left(f_{k_{j+1}} - f_{k_j}\right)_- . \end{aligned}$$

This does require checking that the integrals are uniformly bounded, but fortunately this is true:

$$\begin{aligned}\int_R \sum_{j=1}^N \left(f_{k_{j+1}} - f_{k_j}\right)_+ &\leq \int_R \sum_{j=1}^N |f_{k_{j+1}} - f_{k_j}| \\ &< \sum_{j=1}^n 2^{-j} < 1,\end{aligned}$$

and the same argument works for h.

The same telescoping trick works again:

$$\begin{aligned} g_N - h_N &= \sum_{j=1}^{N} \left(f_{k_{j+1}} - f_{k_j}\right)_+ - \left(f_{k_{j+1}} - f_{k_j}\right)_- \\ &= f_{k_{N+1}} - f_{k_1}. \end{aligned}$$

Therefore $f_{k_{N+1}} = g_N - h_N + f_{k_1}$, so almost everywhere, it goes to $g - h + f_{k_1}$. But since $|f_{k_{N+1}} - f| = |(g_n - g) - (h_N - h)|$, then $f = g - h + f_{k_1}$, so $f_{k_j} \to f$. And it's a general fact in metric spaces that if a subsequence of a Cauchy sequence converges to a limit, then the entire Cauchy sequence converges to that limit, so $f_k \to f$. ⊠

Thus, $L^1(R)$ is a complete normed space (well, except that $\|\cdot\|_1$ is a seminorm, but other than that everything works).

Definition. The space of square-integrable functions is $L^2(R) = \{f \in L^1(R) : f^2 \in L^1(R)\}$.

This is a linear subspace of $L^1(R)$, and is dense in the L^1-norm, but has its own norm, induced from the L^2 inner product: if $f, g \in L^2(R)$, then define

$$\langle f, g\rangle = \int_R f g.$$

This makes sense because if $f, g \in L^2(R)$, one can show that $f g \in L^1(R)$.

Once again, this is almost an inner product:

- This is nonnegative: $\langle f, f\rangle \ge 0$ and $\langle f, f\rangle = 0$ iff $f = 0$ almost everywhere (so this axiom isn't quite met).
- It's bilinear: if $\alpha_1, \alpha_2 \in \mathbb{R}$ and $f_1, f_2, g \in L^2(R)$, then

  $$\langle \alpha_1 f_1 + \alpha_2 f_2, g\rangle = \alpha_1\langle f_1, g\rangle + \alpha_2\langle f_2, g\rangle.$$

  This is also true in the second argument, which follows from the symmetry below.
- The inner product is symmetric: $\langle f, g\rangle = \langle g, f\rangle$ for all $f, g \in L^2(R)$.

Thus, the required conditions are met for the Schwarz inequality to hold (since the proof didn't require the strict positivity of the norm): $|\langle f, g\rangle| \le \|f\|_2\|g\|_2$. This turns $L^2(R)$ into an (almost) normed space, and thus also a metric space. There's a very similar result to Theorem 17.1 for $L^2(R)$, but the proof is essentially the same.

Theorem 17.2. *If $\{f_k\} \subseteq L^2(R)$ is Cauchy in the L^2-norm, then there exists an $f \in L^2(R)$ such that $\|f_k - f\|_2 \to 0$ as $k \to \infty$. Moreover, there exists a subsequence $\{f_{k_j}\}$ converging to f almost everywhere.*

Thus, Riemann-integrable functions can be characterized in either the L^1-norm and the L^2-norm.

From the Lesbegue integral it's possible to recover the Lesbegue measure, though this isn't usually the way it's done.

Measurable Sets. One might also wonder whether every function is Lesbegue integrable. It turns out that this depends on the axioms of set theory one chooses! In the usual formulation, there are non-integrable functions, but it takes some effort to produce them, and in analysis these kinds of functions don't come up often at all (discounting unbounded integrals like $\int 1/x$).

The goal of measure theory is to assign a length (or volume) to every subset of a space. Let R be a rectangle in $\mathbb{R}^n$ and $A \subseteq R$; the goal is to assign a measure to A, generalizing length (area, volume, etc.).

Definition. A is measurable if its characteristic function $\chi_A \in L^1(R)$. If A is measurable, then it measure is defined to be

$$|A| = \int_R \chi_A.$$

This characteristic function (sometimes also called the indicator function) is defined by

$$\chi_A(x) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A. \end{cases}$$

The measure has some nice properties.

(1) If A is measurable, then so is its complement A' (since its characteristic function is $\chi_{A'} = 1 - \chi_A$ and L^1 is closed under addition and subtraction).
(2) $\emptyset$ and $\mathbb{R}$ are measurable (since they have constant characteristic functions).
(3) If $\{A_k\}$ is a sequence of measurable sets, then $\bigcup_{k=1}^{\infty} A_k$ and $\bigcap_{k=1}^{\infty} A_k$ are both measurable, and if the $\{A_k\}$ are pairwise disjoint, then

$$\left|\bigcup_{k=1}^{\infty} A_k\right| = \sum_{k=1}^{\infty} |A_k|.$$

Proof of (3). Call $B_N = \bigcap_{k=1}^N A_k$, so that $\chi_{B_N} = \min\{\chi_{A_1}, \dots, \chi_{A_k}\}$; since each $\chi_{A_j} \in L^1$, then so is χ_{B_N}. Thus, a finite intersection of measurable sets is measurable; to generalize we'll need to use the Monotone Convergence Theorem.

Notice that $\chi_B = \lim \chi_{B_N}$, and $\{\chi_{B_N}\}$ is a decreasing nonnegative sequence in $L^1(R)$, so using Theorem 16.1, $|B| = \lim_{N\to\infty}|B_N|$, and in particular it exists.

Turning to the union, suppose first that the $\{A_k\}$ are pairwise disjoint, so that if $B_N = \bigcup_{k=1}^N A_k$, then $\chi_{B_N} = \sum_{k=1}^N \chi_{A_k}$. Thus, the conditions for the Monotone Convergence Theorem are satisfied (e.g. everything is bounded above by 1), so $B = \bigcup_{k=1}^\infty A_k$ is measurable and $|B| = \sum_{k=1}^\infty |A_k|$, so $\chi_B \in L^1(R)$. ⊠

In the homework, we show that closed sets are also measurable, and thus their complements, open sets, are measurable. This is a huge class of measurable sets, but we'll see that there exist non-measurable sets.

18. Non-Measurable Sets: 6/3/14

Recall that last time, we used the Lesbegue integral to define the measure of a set: if $R \subseteq \mathbb{R}^n$ is an interval and $A \subseteq R$, then A is measurable if $\chi_A \in L^1(R)$ (i.e. $\chi_A(x) = 1$ if $x \in A$ and is 0 otherwise). Then, we showed that if A is measurable, then so is its complement A', and that $\emptyset$, R open sets, and closed sets are all measurable. Furthermore, if $A_1, A_2 \dots$ are all measurable sets, then $\bigcup_{i=1}^\infty A_i$ and $\bigcap_{i=1}^\infty A_i$ are both measurable. Thus, there are lots and lots of measurable sets. These properties used the Monotone Convergence Theorem, so they're far from obvious.

Proposition 18.1 (Countable Additivity)**.** *Suppose $\{A_k\}$ is a sequence of measurable sets that are pairwise disjoint, i.e. $A_k \cap A_\ell = \emptyset$ if $k \neq \ell$. Then, $\bigcup_{i=1}^n A_k$ is measurable and*

$$\left|\bigcup_{k=1}^\infty A_k\right| = \sum_{k=1}^\infty |A_k|.$$

Proof. Let $f_k = \chi_{\bigcup_{j=1}^k A_j}$, the characteristic function of the first k A_j. Thus, if $f_k(x) = 1$, then x is in exactly one A_j, so $f_k = \sum_{j=1}^k \chi_{A_j}$. In particular, $\{f_k\}$ is an increasing sequence and that $\int_R f_k \le |R|$, since each $A_j \subseteq R$. Thus, by Theorem 16.1, if $f = \chi_{\bigcup_{j=1}^\infty A_j}$ is the characteristic function of the whole set, then $\lim f_k = f$ and

$$\int_R f = \lim_{k\to\infty} \int_R f_k = \lim_{k\to\infty} \sum_{j=1}^k |A_j|.$$

$$\implies \left|\bigcup_{j=1}^\infty A_j\right| = \sum_{j=1}^\infty |A_j|.$$ ⊠

The Lesbegue measure is also translation-invariant, though for simplicity we'll show this only on $R = [0,1)$. If $t \in \mathbb{R}$ and $x \in R$, define $x \mathbin{\overset{\circ}{+}} t = (x+t) \bmod 1 \in R$, and similarly $x \mathbin{\overset{\circ}{-}} t = (x-t) \bmod 1$. Then, if $A \subseteq R$, define $A_t = A \mathbin{\overset{\circ}{+}} t = \{a \mathbin{\overset{\circ}{+}} t : a \in A\}$: this represents translating A by distance t along this circle. It's possible to do this on $\mathbb{R}^n$ in general, but since we've only dealt with measure on intervals, we'll prove the more specific case.

Proposition 18.2. *If A is measurable, then $|A_t| = |A|$.*

Proof. This is quite easy to check: let $f_t(x) = f(x \mathbin{\overset{\circ}{-}} t)$ for any $f \in L^1(R)$ and $t \in \mathbb{R}$. Assuming that $f_t \in L^1(R)$ and $\int_R f_t = \int_R f$, then the rest follows, because if A is measurable, then $\chi_A \in L^1(R)$ and $\chi_{A_t}(x) = \chi_A(x \mathbin{\overset{\circ}{-}} t) = (\chi_A)_t(x)$, so by the assumed claim,

$$|A_t| = \int_R \chi_{A_t} = \int_R \chi_A = |A|.$$

Thus, it remains only to check the property on L^1, and since step functions are dense in L^1, then it's enough to check it on step functions: suppose $\varphi \in \mathscr{S}(R)$. Then, $\varphi_t \in \mathscr{S}(R)$ as well, since if $\mathscr{P}$ is a partition for φ, then its translate is for φ_t. Then, write

$$\varphi = \sum_{I\in\mathscr{P}} a_I \chi_I,$$

so the problem reduces to showing it for the characteristic functions of intervals, since the integrals will be the same if $|I \mathbin{\overset{\circ}{+}} t| = |I|$. But this is clear. ⊠

We will now give a standard construction for non-measurable sets. In analysis, these are usually recognized, but not thought much about: they exist off in logic somewhere. Nonetheless, they exist, and the following construction is very general.

Claim. Not all sets are measurable.

Proof. For $x, y \in [0,1)$, let $x \sim y$ if $x - y \in \mathbb{Q}$. This is an equivalence relation, because $\mathbb{Q}$ is a field. This equivalence relation divides $[0,1)$ into equivalence classes. One of these equivalence classes is $\mathbb{Q}$, and for any $a \notin \mathbb{Q}$, $a + \mathbb{Q}$ is as well. Thus, $[0,1)$ is the disjoint, uncountable union of these equivalence classes.

Let E be a set that has exactly one point from each equivalence class. Wait, can we do that? It's clearly possible with a finite number of nonempty sets, or even (by induction) if there are countably many. But since there are uncountably many, this boils down to an axiom of set theory, called the Axiom of Choice. This is normally assumed, unless one does constructive mathematics — and it's a perfectly reasonable thing to assume. This set E, if we assume it exists, can be proven to be non-measurable.

This set E has the odd property that if $t \in \mathbb{Q} \setminus 0$, then $E \overset{\circ}{+} t$ is disjoint from E (since $x - t \sim x$, so they are in the same equivalence class, and thus are not both in E), and if $t_1, t_2, \dots$ is an enumeration of $\mathbb{Q} \cap [0,1]$, then

$$\bigcup_{i=1}^{\infty} E_{t_i} = [0,1).$$

This is true because all $x \in [0,1)$ lie in some equivalence class, and thus $x \sim y$ for some $y \in E$, so $x - y \in \mathbb{Q}$ and $x \in E_{x-y}$.

From the axiom of choice, we constructed an uncountable set whose countable disjoint translates cover the whole interval. This is kind of strange. Now, suppose E were measurable, so that

$$1 = |[0,1)| = \sum_{i=1}^{\infty} |E_{t_i}| = \sum_{i=1}^{\infty} |E|.$$

There is no way to add a real number to itself a countable number of times and get 1, so this is impossible. Thus, E cannot be measurable. ⊠

The axiom of choice for uncountable collections is needed to produce non-measurable sets. In fact, there's a very interesting theorem in set theory that says that if the axiom of choice is taken to be false, then every set is measurable! But analysts usually assume the Axiom of Choice and generally have sets that are measurable.

If one develops Lesbegue theory in the standard way, one constructs the measure first, and then defines measure on functions, which induces the Lesbegue integral. The generating idea is that of simple functions (constant on more general measurable sets than just intervals) rather than step functions, and leads to the fuller theory.