Math 104: Introduction to Analysis

Evan Chen

UC Berkeley

Notes for the course MATH 104, instructed by Charles Pugh.

1 August 29, 2013

Hard: #22 in Chapter 1. Consider a *pile of sand* principle. You wish to take away part of it to make it smaller than some number ϵ. Suppose you can take away 5% of what remains on any given day. Eventually you will reduce the pile of sand to less than ϵ.

Today we will cover section 1 and part of section 2.

1.1 Notation

$\forall$ is read "for each", do NOT use "for all".
$\exists$ is read "there exists".
$\Rightarrow$ is read "implies". Then $A \Rightarrow B$. Note that the principle of explosion exists.

Set notation: we use capital letters to denote sets, and small letters to denote the elements of the set (e.g. $x \in X$). Then x is called an *element* (or a *point* of such a set).

Then

$$\{x \in X : \dots\}$$

is read "the set of all x in X such that...".

Common sets:

- $\varnothing$ is the empty set.
- $\mathbb{N} = \{1, 2, 3, \dots\}$ is the set of natural numbers.
- $\mathbb{Z}$ is the set of integers.
- $\mathbb{Q}$ is the set of rational numbers. We can write this as

$$\mathbb{Q} = \left\{ \frac{p}{q} : p \in \mathbb{N} \cup \{0\}, q \in \mathbb{Z} \setminus \{0\} \right\}.$$

This is a terrible definition, probably a better one is

$$\mathbb{Q} = \left\{ \frac{p}{q} : p \in \mathbb{Z}, q \in \mathbb{N} \right\}.$$

Note that $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q}$. Note that in this course, $\subset$ will mean what is usually meant by $\subseteq$.

We will assume all the nice properties of these sets.

The subject of the course, however is $\mathbb{R}$.

1.2 The Real Numbers

In most books, we accept properties of $\mathbb{R}$ as axioms. We will instead *construct* $\mathbb{R}$ and prove that the "axioms" are true.

The most important property of the real numbers is the *least upper bound* property.

Fact 1.1 (Least Upper Bound)**.** If S is a nonempty set of real numbers and there exists an upper bound for S, then there exists a *least upper bound* in $\mathbb{R}$ for S.

Note: we say a upper bound *for* S, not *of* S; the latter suggests the upper bound is contained in S.

Definition 1.2. An upper bound for a set S is a real M such that $M \geq s \forall s \in S$.

Example 1.3. Let $S = \{r \in \mathbb{Q} : r < 2\}$. Then an upper bound for S is 9000, but the least upper bound is 2.

Note that the *least upper bound* property is NOT true for the rationals. For example, consider

$$\left\{r \in \mathbb{Q} : r < 0 \text{ or } 0 \le r^2 < 2\right\}.$$

This has no least upper bound in the rationals, but it is $\sqrt{2}$ in the reals.

1.3 Constructing the Reals

Theorem 1.4 (Irrationality of the Square Root of 2)**.** *For every rational $\frac{p}{q} \in \mathbb{Q}$, we have $\left(\frac{p}{q}\right)^2 \neq 2$.*

Proof. Suppose for contradiction that $\left(\frac{p}{q}\right)^2 = 2$ where $(p, q) = 1$. In particular, p and q are both not even.

For the first case, assume p is odd. But now $p^2 = 2q^2$, which is a contradiction. `

Now suppose p is even. Then q is odd. Now $p^2 = 2q^2$, but comparing th 2-adic evaluations yields a contradiction. `

This is a contradiction in both cases, so we're done. $\square$

Theorem 1.5. *If $\left(\frac{p}{q}\right)^n = k \in \mathbb{N}$ where $m \in \mathbb{N}$, then $q = 1$.*

Proof. Do some blah with prime factorizations. $\square$

TODO: Read Chapter 1, Section 1.

1.4 Sets and Classes

Definition 1.6. A set is a collection of elements.

Definition 1.7. A class is a collection of sets.

For example, $\mathbb{N}$ is a set. Meanwhile, we might consider $\mathcal{F}$ as a collection of all finite subsets of $\mathbb{N}$. For example, $\{1, 6\} \in \mathcal{F}$. On the other hand, $13 \notin \mathcal{F}$, while $\{13\} \in \mathcal{F}$.

1.5 Constructing the Real Numbers

We will construct $\mathbb{R}$ with the method of *Dedekin cuts.*

Definition 1.8. A *cut* in $\mathbb{Q}$ is a division of $\mathbb{Q}$ into two nonempty sets A and B such that

(i) $A \sqcup B = \mathbb{Q}$; i.e. $A \cup B = \mathbb{Q}$ and $A \cap B = \varnothing$.

(ii) $\forall a \in A \forall b \in B, a < b$.

(iii) A has no largest element.

The notation is $A \mid B$.

Definition 1.9. If B has a smallest element $r \in \mathbb{Q}$, we say we have cut the line *at* r, and this is a *rational cut.*

Example 1.10. Take $A = \{r \in \mathbb{Q} : r < 2\}$ and $B = \mathbb{Q} - A$. In this case, $\min B = 2$. So this is a (rational) cut at 2.

Example 1.11. $A = \left\{ r \in \mathbb{Q} : r \le 0 \text{ or } 0 \le r^2 < 2 \right\} \mid B = \mathbb{Q} - A$ is *not* a cut of rational number.

So now we make a definition.

Definition 1.12. A *real number* is a cut in $\mathbb{Q}$. $\mathbb{R}$ is the collection of all cuts in $\mathbb{Q}$.

Now we can impose ordering, etc. on these. We say that $A \mid B \le C \mid D$ if and only if $A \subseteq C$. Also, $A \mid B < C \mid D$ if and only if $A \subseteq C$ but $A \neq C$.

Now we can rewrite the least upper bound property as

Theorem 1.13 (Least Upper Bound)**.** *If $\mathcal{S}$ is a collection of cuts in $\mathbb{Q}$ for which $\mathcal{S}$ is not empty, and there exists an upper bound of $\mathcal{S}$, then there is a least upper bound for $\mathcal{S}$.*

Proof. Take

$$E = \{ r \in QQ : \exists A \mid B \in \mathcal{S} \text{ and } r \in A \}$$

and let $F = \mathbb{Q} - E$. In other words, $E = \cup_{A \mid B \in \mathcal{S}} A$. We claim $E \mid F$ is a least upper bound.

Now E is not empty because $\mathcal{S}$ is not empty. Because $A \subseteq E \forall A \mid B \in \mathcal{S}$, E is certainly an upper bound. But it must also be a least upper bound, because any other cut misses some element of E, which is contained in some A.

Finally we have to check that $E \mid F$ is indeed a cut. $F \neq \varnothing$ because there exists an upper bound for $\mathcal{S}$, and E does not have a maximal element because none of the A do. Finally, check that all elements of F are greater than all elements of E. $\square$

Finally we need to embed $\mathbb{Q}$ in our $\mathbb{R}$. This is not automatic, since $\mathbb{R}$ is actually a collection of cuts. So, for any $c \in \mathbb{Q}$, we associate it with the cut at c, namely

$$c^* \overset{\text{def}}{=} \{ r \in Q \mid r < c \} \mid \{ r \in \mathbb{Q} : r \ge c \} .$$

This allows us to say $\mathbb{Q} \subset \mathbb{R}$, even though formally this is wrong. But these identifications allow us to write this anyways.

One can also check that $c < d \Leftrightarrow c^* < d^*$, so the ordering of $\mathbb{Q}$ agrees with that of $\mathbb{R}$.

1.6 Arithmetic

Now for the hard part... we need to endow $\mathbb{R}$ with an operation.

Well, addition isn't too bad

Definition 1.14. For cuts $A \mid B$ and $C \mid D$, we define

$$(A \mid B) + (C \mid D) \overset{\text{def}}{=} (A + B) \mid (\mathbb{Q} - (A + B)) .$$

One needs to check a LOT of things. First check that it's actually a cut, and that all arithmetic properties (commutative, associative, identity ...) hold.

It's even worse to define $-x$ given $x = A \mid B$. You need to handle the case where B has a smallest element. And so on.

Then you need to prove that $x + (-x) = 0^*$. Now we need to prove equality of sets.

Remark 1.15. The only reliable way to show that $A = B$ is $A \subseteq B$ and $B \subseteq A$.

And then there's $A \mid B \cdot C \mid D$. This will require casework splitting based on sign.

Given two positive guys $x, y > 0$, we define the E set by

$$E \stackrel{\text{def}}{=} \{r \in \mathbb{Q} : r \leq 0 \text{ or } r = ac \text{ where } a, c \in A \times C \text{ and } a, c > 0\}.$$

Next, if $x = 0$ or $y = 0$, then $xy = 0$.

Finally, if $x > 0$, $y < 0$, then define $x \cdot y \stackrel{\text{def}}{=} -(x \cdot (-y))$. And you get the idea for the other cases...

I won't even write down anything for showing $x(y + z) = xy + xz$.

2 September 3, 2013

How to terrify students: ask them to read ahead, and then ask them what a Cauchy sequence is! I should try doing this.

2.1 Notation

$a \in A$ reads a is an element of A

$S \subset T$ reads S is contained in T.

2.2 Ordering of Cuts

Let $x = A \mid B$ and $y = C \mid D$ be cuts. We say $x < y$ if and only if $A \subset C$ and $A \neq C$.

We also had cuts which take place at rational numbers; that is$c \in \mathbb{Q}$ is associated with a cut c^*. In particular,

$$0^* = \mathbb{Q}_{<0} \mid \mathbb{Q}_{\geq 0}.$$

Now we say that $x = A \mid B \in \mathbb{R}$ is positive if $0^* < x$, negative if $0^* > x$, and zero if $x = 0^*$. The terms *nonnegative* and *nonpositive* retain their usual definition.

We also define $|x|$ to be x if $x > 0$, 0 (or 0^*) if $x = 0$ and $-x$ if $x < 0$.

We have some nice properties:

- Trichotomy: exactly one of $x = 0$, $x > 0$, $x < 0$ holds for each $x \in \mathbb{R}$.
- Transitivity: $x < y$ and $y < z$ implies $x < z$.
- Translation: $\forall x, y, z \in \mathbb{R}$¡ we have $x < y \Rightarrow x + z < y + z$.

The first two properties are obvious from the cut definition. The third property follows by the definition of $+$ for cuts.

2.3 Triangle Inequality

Fact 2.1 (Triangle Inequality)**.** For all $x, y \in \mathbb{R}$ we have $|x + y| \leq |x| + |y|$.

Proof. Obviously $x+y \leq |x|+|y|$. On the other hand, $-(x+y) = (-x)+(-y) \leq |x|+|y|$, so $x + y \geq -(|x| + |y|)$. Hence, we're done. $\square$

2.4 Cauchy Sequences

Definition 2.2. A sequence of real numbers is an ordered list $x_1, x_2, x_3, \dots$ of real numbers. It is denoted (x_n) or $(x_n)_{n \in \mathbb{N}}$.

Some books use $\{x_n\}$ for the same thing, but this is bad because it coincides with the set of elements of the sequence.

Definition 2.3. (x_n) converges to $b \in \mathbb{R}$ as $n \to \infty$ if and only if for each $\varepsilon > 0$, $\exists N \in \mathbb{N}$ such that $\forall n \geq N$, $|x_n - b| < \varepsilon$.

This is a good place to point out that we are saying "for each $\varepsilon > 0$" and certainly not "for every $\varepsilon > 0$".

Definition 2.4. A sequence (x_n) satisfies a Cauchy condition (or is Cauchy) if for each $\varepsilon > 0$, $\exists N$ such that if $n, m \geq N$ then $|x_n - x_m| < \varepsilon$.

Theorem 2.5 (Cauchy Convergence Criterion)**.** *For sequences of real numbers, the two definitions are equivalent; that is, a sequence of reals converges if and only if it is Cauchy.*

This is abbreviated CCC. The fact that convergence implies Cauchy is trivial; tho converse is more interesting.

Proof. We prove only the hard direction. It is easy to check that $\{x_m : m \in \mathbb{N}\}$ is bounded (and nonempty), by picking $\varepsilon = 1000$ in the Cauchy condition.

Now let us consider

$$S \overset{\text{def}}{=} \{x \in \mathbb{R} : \exists \text{ inf many } n \in \mathbb{N} \text{ with } x_n \geq x\}.$$

Clearly S is bounded above and nonempty. Let $b = \sup S$. We claim that the limit is b.

We need to show that $\varepsilon > 0$, there is an N such that for each $n \geq N$ we have $|x_n - b| < \varepsilon$.

Now there exists an N_2 such that for every $m, n \geq N_2$ we have $|x_m - x_n| < \frac{1}{2}\varepsilon$. Because $b + \frac{1}{2}\varepsilon$ is not in S, there exists N_3 such that for each $n \geq N_3$ we have $x_n \leq b + \frac{1}{2}\varepsilon$ (because there are only finitely many counterexamples). Meanwhile, because $b - \frac{1}{2}\varepsilon$ is not an upper bound for S, then $b - \frac{1}{2}\varepsilon \in S$, and there are infinitely many n such that $b - \frac{1}{2}\varepsilon \leq x_n \leq b$. Hence there exists one with a large subscript N_4 such that x_{N_4} is in the interval, and $N_4 > \max\{N_1, N_2, N_3\}$. But because $x_{N_4} \in \left(b - \frac{1}{2}\varepsilon, b + \frac{1}{2}\varepsilon\right)$, all the x_n are within $\frac{1}{2}\varepsilon$ of x_{N_4}, which is sufficient. $\square$

2.5 Euclidean Space

We consider $\mathbb{R}^m$. This is a vector space (over $\mathbb{R}$), meaning the elements have an addition and multiplication by scalars. The elements are m-tuples of real numbers, $x = (x_1, x_2, \dots, x_m)$.

$\mathbb{R}^m$ also has a dot product (or inner product). This is a map $\langle \cdot, \cdot \rangle : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ by

$$\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i.$$

This product is bilinear, which means that $x, y, z \in \mathbb{R}^m$ then $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$ (this is clear). Likewise, $\langle x, y \rangle = \langle y, x \rangle$ and $\langle tx, y \rangle = t \langle x, y \rangle$. Finally, it is positive definite; that is, $\langle x, x \rangle \geq 0$ and equality occurs if and only if x is the zero vector.

Definition 2.6. For $x \in \mathbb{R}^m$, we define $|x, x| = \sqrt{\langle x, x \rangle}$.

Theorem 2.7 (Cauchy-Schwarz)**.** *For any x and y in $\mathbb{R}^m$,*

$$\langle x, y \rangle^2 \leq |x| \, |y|.$$

Proof. Assume $x, y \neq 0$, otherwise it's obvious.

Let $t \in \mathbb{R}$, and observe

$$0 \leq \langle x + ty, x + ty \rangle = \langle x, x \rangle + 2t \langle x, y \rangle + t^2 \langle y, y \rangle.$$

The right-hand side is a quadratic in t which clearly has a nonpositive discriminant, so taking $B^2 - 4AC \leq 0$ yields the Cauchy-Schwarz inequality. (We note that $\langle y, y \rangle \geq 0$ so this is a quadratic with a positive leading term.) $\square$

This implies good things about $\mathbb{R}^m$; for instance,

$$|x - y| = (\langle x - y, x - y \rangle)^2 = \sqrt{(x_1 - y_1)^2 + \dots + (x_m - y_m)^2}$$

which is the distance between two points in $\mathbb{R}^m$.

Fact 2.8 (Triangle Inequality)**.** For $x, y \in \mathbb{R}^m$, $|x + y| \le |x| + |y|$.

Proof. Square both sides, use bilinearity and it reduces to twice the Cauchy Schwarz inequality. $\square$

Of course, this implies $|x - z| \le |x - y| + |y - z|$.

2.6 Closing Remarks

Definition 2.9. The *unit ball* B^m is defined as $\{x \in \mathbb{R}^m : |x| \le 1\}$. The unit sphere S^{m-1} is defined as $S^{m-1} = \{x \in \mathbb{R}^m : |x| = 1\}$.

Note the differences in the subscripts! S^{m-1} is the boundary of B^m – try to not confuse them.

3 September 5, 2013

Pop quiz: what's the area of a circle? Zero. The disk is the one with positive area.
Today's lecture is about cardinality.

3.1 Function

A function $f : X \to Y$ is a map which associates each $x \in X$ an element $f(x) \in Y$. The set X is called the domain and the set Y is called the target. The range of Y is the set of points which are actually hit by f.

It's much better to say "consider a function f" rather than "consider a function $f(x)$". Furthermore, we define

$$f^{\text{pre}}(y) = \{x \in X : f(x) = y\}$$

Note that this is *not* actually a function from Y to X, but from Y to 2^X. (So the common notation $f^{-1}(x)$ is abusive.) We only use this notation when f is actually one-and-one and onto.

Definition 3.1. For a function f,

- f is *one-to-one* (or injective) if $\forall x, x' \in X$, if $x \neq x'$ then $f(x) \neq f(x')$.
- f is *onto* (or surjective) if $\forall y \in Y$, $\exists x \in X$ (not necessarily unique) such that $y = f(x)$.
- f is a *bijective*, or a bijection, if these are both true.

3.2 Cardinality

Definition 3.2. Sets X and Y *have the same cardinality* if there exists a bijection from X to Y. We write $X \sim Y$ for this relation.

This is an equivalence relation.

Definition 3.3. If $\mathbb{N}$ and X have the same cardinality, then X is said to be *denumerable.*

Claim 3.4. $\mathbb{Z} \sim \mathbb{N}$.

Proof. Take

$$f(n) = \begin{cases} \frac{1}{2}(n+1) & \text{if } n \text{ is odd} \\ -\frac{1}{2}n & \text{if } n \text{ is even.} \end{cases}$$

□

Theorem 3.5 (Cantor's Diagonalisation Argument)**.** *$\mathbb{R}$ is NOT denumerable.*

Proof. Insert Cantor's proof here. □

3.3 Cardinality Continued

If X is denumerable, then it can be exhibited as a list. Conversely, given a sequence of the elements of X, then X is denumerable.

Proposition 3.6. *$\mathbb{N} \times \mathbb{N}$ is denumerable.*

Proof. A list is $(1, 1)$, $(1, 2)$, $(2, 1)$, $(1, 3)$, $(2, 2)$, $(3, 1)$, et cetera is a suitable list. □

It turns out that $\mathbb{Q}$ is also denumerable; make the same list, and cross out repetitions. Because there are infinitely many rational numbers, we are okay. But we will see a different way to do this.

Note that if X and Y are denumerable, then $X \times Y$ is denumerable (since $\mathbb{N} \times \mathbb{N}$ is denumerable). Then we can also write $X \times Y \times Z = (X \times Y) \times Z$, so a similar property holds for 3 sets (or even n sets) by induction. In particular, $\mathbb{N}^k$ is denumerable for all k.

But $\mathbb{N}^\infty$ is not denumerable! It contains all sequences of digits, i.e. $\{0, 9\}^\infty$, so it contains a subset bijective to $\mathbb{R}$ which is already uncountable.

Theorem 3.7. *Suppose X is an infinite set.*

(a) If $\exists f : X \to \mathbb{N}$ an injection, then X is denumerable.

(b) If $\exists g : \mathbb{N} \to X$ a surjection, then X is denumerable.

Proof. For part (a), notice that $f(X)$ is infinite and nonempty. But $\mathbb{N}$ has the least criminal property, so we can get a bijection! Namely, let $\tau(n)$ denote the nth smallest element of $f(X)$; this is defined because $f(X)$ is infinite. Then τ is a bijection.

Part (b) follows from part (a), because for each $x \in X$ we can construction an injection $f : X \to \mathbb{N}$ by selecting an arbitrary element of $g^{\text{pre}}(x)$ (which is nonempty by assumption) for each $f(x)$. $\square$

This provides a "clean" proof that $\mathbb{Q}$ is denumerable, since there is an injection from $\mathbb{Q}$ into $\mathbb{Z} \times \mathbb{N} \sim \mathbb{N}$ (just send $\frac{p}{q}$ to (p, q), where $q > 0$ and the gcd is 1).

Proposition 3.8. *Let $A = \cup_{i=1}^\infty A_i$, where each A_i is denumerable. Then A is denumerable.*

Note that this behaves differently from Cartesian product (since, say, $\mathbb{R}^\infty$ is denumerable).

Proof. There is a surjection from $\mathbb{N} \times \mathbb{N}$ to $\cup A_i$ by taking (i, j) to the jth smallest element of A_i. $\square$

Remark 3.9 (Hotel Story)**.** A hotel has denumerably many hotels, each with finitely many rooms, with all rooms filled. Then you can accommodate one additional guest.

You can use this to show $[a, b) \sim (a, b)$.

4 September 10, 2013

4.1 Metric Spaces

Definition 4.1. A *metric space* is a set M and a function $d : M \times M \to \mathbb{R}$ such that

(i) d is positive definite and symmetric, meaning $d(y, x) = d(x, y) \ge 0 \forall x, y \in M$, with equality if and only if $x = y$, and

(ii) for all $x, y, z \in M$, $d(x, z) \le d(x, y) + d(y, z)$.

The best example of a metric space is $\mathbb{R}^2$, where d is the Euclidean distance; i.e. $d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$. Of course this generalizes to $\mathbb{R}^m$ and $\mathbb{R}$.

Given $M \subseteq \mathbb{R}^2$, we can consider the inherited metric space with the same distance. That means $d(x, y) = d_{\mathbb{R}^2}(x, y)$. In other words, basically everything is a metric space when we give the inherited metric.

Not all distance functions are inherited. For consider S^2. One possible distance is simply to consider $S^2 \subset \mathbb{R}^3$ as a parent metric space. But we can also consider the distance between two points as the great-circle distance; this will be a *different* metric on the sphere.

Another possibility, on the surface of a torus, would be the shortest arc on the surface of the torus.

The upshot is that any claim about metric spaces applies to many objects.

Finally, a much cooler metric.

Example 4.2 (Discrete metric)**.** Given any set M, define

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

4.2 Sequences

Suppose $(x_n) = x_1, x_2, x_3, \dots$ is a sequence of points in a metric space M.

Definition 4.3. We say that (x_n) converges to a point p in M if and only if for each $\varepsilon > 0$, there exists a positive integer N such that if $n \ge N$ then the distance between x_n and p is less than epsilon.

Example 4.4. In $\mathbb{Q}$, the sequence $1, 1.4, 1.41, 1.414, \dots$ converges to $\sqrt{2}$ in $\mathbb{R}$. However, it does *not* converge in $\mathbb{Q}$!

The standard notation is $x_n \to p$ or $\lim_{n\to\infty} x_n = p$, among several other notations. But all of this is done in a metric space, not $\mathbb{R}$. In particular, it does not make sense to "subtract" in a general metric space, or to say one point is less than another. For example, in $\mathbb{R}^2$, the sequence $\left(1 - \frac{1}{n}, 1 + \frac{1}{2} + \dots + \frac{1}{2^n}\right)$ converges $(1, 2)$, but we don't have an ordering.

4.3 Maps between Metric Spaces

Let $F : M \to Y$ be a function, where M and Y are metric spaces equipped with distance functions d_M and d_Y. Take $p \in M$.

Definition 4.5. f is continuous at p if and only if

(i) For every sequence in M that converges to p, the f-image of that sequence converges to $f(p)$ in M.

(ii) For every $\varepsilon > 0$ there exists a $\delta > 0$ such that if $d(x,p) < \delta$ then $d(fx, fp) < 0$.[1]

Theorem 4.6. *These two definitions are equivalent.*

Proof. First we will show the first definition implies the second. Assume (2) is false, then $\exists \varepsilon > 0$ such that $\forall \delta > 0$ there exists a "bad x" in M such that $d(x,p) < \delta$ yet $d(fx, fp) \ge \varepsilon$.

For each positive integer n, let $\delta = \frac{1}{n}$, and let x_n be a point such that $d(x_n, p) < \frac{1}{n}$ and $d(fx_1, fp) \ge \varepsilon$. Now consider the sequence $x_1, x_2, \dots$. Because $d(x_n, p) < \frac{1}{n}$ for each n, we find that x_n converges to p. On the other hand, $d(fx_n, fp) \ge \varepsilon$ for each n, so the f-image does *not* converge to $f(p)$. This is a contradiction, so our assumption was wrong and (2) is true.

Now for the other direction. Assume (2), and suppose that x_n converges to p. We wish to show $f(x_n)$ converges to $f(p)$. That means, for any given $\varepsilon > 0$, we need to find N such that $\forall n \ge N : d(fx_n, fp) < \varepsilon$. However, we know there exists a $\delta > 0$ such that if $d(x,p) < \delta$ then $d(fx, fp) < \varepsilon$ by (2). But $\lim x_n = p$; this means $d(x_n, p) < \delta$ for large enough n, completing the proof. ☐

Note that none of this uses the triangle inequality.

4.4 Continuity Continued

No pun intended.

Definition 4.7. If $f : M \to Y$ and f is continuous at each $p \in M$, then f is continuous.

Proposition 4.8. *Let $f : M \to Y$ and $g : Y \to P$ be continuous maps between metric spaces. Then $g \circ f : M \to P$ is continuous as well.*

Proof. Trivial using sequences. ☐

This is much more obnoxious with the ε-δ definition. This suggests that it is probably best to think of metric spaces in terms of convergent sequences than anything else.

Definition 4.9. If (x_n) is a sequence in M and $1 \le m_1 < m_2 < m_3 < \dots$ is a sequence of integers, then $x_{m_1}, x_{m_2}, \dots$ is a *subsequence* of the original sequence.

Proposition 4.10. *A subsequence of a convergent sequence converges to the same point.*

4.5 Homeomorphism

Definition 4.11. Consider $f : M \to Y$. If f is a bijection and f is continuous, and $f^{-1} : Y \to M$ is also continuous, then f is called a *homeomorphism*.

Example 4.12. Ellipses are homeomorphic to S^1 if they inherit a distance metric from $\mathbb{R}^2$. Function: assume they are cocentric, and map $x \in S^1$ to $y \in E$, where O, x, and y are collinear, and O lies outside the segment xy.

Example 4.13. Doughnuts are homeomorphic to coffee cops.

Example 4.14. The trefoil know is homeomorphic to S^1 because it is trivial to construct an homeomorphism in terms of paths.

It's not possible to do a deformation in $\mathbb{R}^3$, but we can do so in $\mathbb{R}^4$! If you imagine the fourth dimension in color…

[1] Here we are using the shorthand $fx \stackrel{\text{def}}{=} f(x)$.

5 September 12, 2013

Today: closed, open, clopen sets, and so on.

Recall that $f : M \to N$ is a homeomorphism if it is a bi-continuous bijection if it is a bijection which is continuous and the inversive bijection is also continuous.

It's important to note that the f continuous does not imply f^{-1} is continuous. Consider the map $f : [0, 2\pi) \to S^1$ by $x \mapsto e^{ix}$. Then it is not hard to construct a sequence of points in S^1 which approaches $1 \in S^1$, but whose inverses approach 2π.

5.1 Closed Sets

Definition 5.1. Suppose $S \subset M$, where M is a metric space. Define $\lim S$, the *limit set* of S, by

$$\lim S \stackrel{\text{def}}{=} \{p \in M : \exists (x_n) \in S \text{ such that } x_n \to p\}.$$

Remark 5.2. $S \subseteq \lim S$ because $\forall p \in S$, $p, p, p, p, \dots$ is a sequence converging to p.

Definition 5.3. S is closed in M if $\lim S = S$.

Proposition 5.4. *$\lim S$ is closed, regardless of whether S itself is closed. That is, $\lim \lim S = \lim S$.*

Proof. We wish to show that if $y_m \to p \in M$, where $y_m \in \lim S$, then $p \in \lim S$ as well.

For each k, y_k is a limit of some sequence in S, say $(x_{k,n})_{n\ge1}$, converging to y_k. For each k, we can find $x_{k,n(k)} \in S$ for which $d(x_{k,n(k)}, y_k) < \frac{1}{k}$. For each k, we can find $x_{k,n(k)} \in S$ for which $d(x_{k,n(k)}, y_k) < \frac{1}{k}$.

Then for each $\varepsilon > 0$ there exists N such that for all $k > N$,

$$d(y_k, p) < \frac{1}{2}\varepsilon \Rightarrow d(x_{k,n(k)}, p) < \frac{1}{2}\varepsilon + \frac{1}{k}$$

so we deduce that $p \in \lim S$ as required. $\square$

Corollary 5.5. *$\lim S$ is the smallest closed subset of M that contains S.*

So, we say that $\lim S$ is the *closure* of S.

5.2 Open Sets

Definition 5.6. For each r, we define

$$M_r(p) \stackrel{\text{def}}{=} \{x \in M : d(x, p) < r\}.$$

$M_r(p)$ is called the *r-neighborhood* of p.

Definition 5.7. A set $U \subset M$ is *open* in M if for each $p \in U$, there exists $r > 0$ such that $M_r(p) \subseteq U$.

Example 5.8. $(0, 1)$ is open in $\mathbb{R}$ but not $\mathbb{R}^2$.

Theorem 5.9. *Open and closed are dual concepts: if S is closed then $M \setminus S$ is open, and vice-versa.*

Proof. Let $S^c = M \setminus S$. First we show S closed implies S^c open. Let $p \in S^c$ be given. Suppose not, and for each $r > 0$ and $M_r(p)$ fails to be in S^c. Then for each n, by taking $r = \frac{1}{n}$ there exists $x_n \notin S^c$ but $x_n \in M_{\frac{1}{n}} p$. But now $x_n \in S$ is a sequence converging to p, yet $p \in S^c$, contradiction. $\square$

Example 5.10. Let $S \subset M$, where M is a metric space equipped with the discrete metric. Then S is closed. Furthermore, S^c is closed so S is open.

Definition 5.11. A set is *clopen* if it is both open and closed.

Problem 5.12. If every set is clopen, then it is homeomorphic to the discrete space.

5.3 Topology

Definition 5.13. Consider the collection of all open sets in M. This is called the *topology* $\mathcal{T}(M)$ of M.

Proposition 5.14. *For any metric space M, then*

(i) $\varnothing, M \in \mathcal{T}$.

(ii) Any union of members of $\mathcal{T}$ belongs to $\mathcal{T}$, even infinitely many, countable or uncountable.

(iii) Any finite intersection of open sets of $\mathcal{T}$ is open.

Proof. The first two are trivial. The third fact follows from the fact that the intersection of two open sets is open – if $p \in U \cap V$, where U and V are open, then $\exists r, s > 0$ for which $M_r(p) \in U$ and $M_s(p) \in V$, whence $M_{\min\{r,s\}}(p) \in U \cap V$ as required. □

It is also not hard to show that if $f : M \to N$ is a homeomorphism, then the topologies are bijected as well (so we obtain a bijection from $\mathcal{T}(M) \to \mathcal{T}(N)$). In other words, homeomorphisms preserve topologies.

By the way, taking De Morgan's laws gives the properties

(i) M and $\varnothing$ are closed.

(ii) The finite union of closed sets is closed.

(iii) All intersections of closed sets are closed.

5.4 More on Closed and Open Sets

Example 5.15. Let $S = \{x \in \mathbb{Q} : x \le 0\}$. S is open in $\mathbb{Q}$, because if a sequence in S converges to a limit **in the rational numbers** converges to something in S. The fact that there is a sequence converging to irrational numbers is irrelevant because our metric space is $\mathbb{Q}$.

But of course S is not closed in $\mathbb{Q}$; consider the interval around S.

Example 5.16. $\left\{x \in \mathbb{Q} : x < \sqrt{2}\right\}$ is clopen in $\mathbb{Q}$.

So, to re-iterate, we care A LOT about what M is!

Theorem 5.17. *Every open set $S \in \mathbb{R}$ is a countable disjoint union of open intervals (including rays).*

Proof. Clearly $S = \varnothing$ "works" because it is a "union" of zero intervals.

For each $x \in S$, consider $b_x \overset{\text{def}}{=} \sup\{y : (x, y) \subseteq S\}$. This is okay because the set is nonempty. Note that it is possible that $b_x = +\infty$. Similarly, let $a_x \overset{\text{def}}{=} \lim\{y : (y, x) \subseteq S\}$. Now define

$$\forall x \in S, \quad I_x \overset{\text{def}}{=} (a_x, b_x).$$

We claim that $\forall x, x' \in S$, either $I_x = I_{x'}$ or $I_x \cap I_{x'} = \varnothing$. This is easy to show via contradiction.

So S consists of a disjoint union of open intervals. To show there are countably many, take a rational number in each interval. Hence we see the intervals correspond to some subset of the rationals. There are only countably many rationals, so there are countably many intervals.[2] $\square$

[2] This is a commonly used technique.

6 September 17, 2013

Definition 6.1. The *boundary* of a set $S \subseteq M$ is defined by

$$\partial S = \overline{S} \cap \overline{M \setminus S}.$$

6.1 Continuity and the Open Set Condition

Recall the ε-δ definition of continuity; f is continuous iff $\forall p \in M, \varepsilon > 0$, $\exists \delta > 0$, we have $d_M(p,q) < \delta \Rightarrow d(fp, fq) < \varepsilon$.

Recall that we define $M_\lambda(p) = \{q \in M : d(p,q) < \lambda\}$. So we can rephrase the definition as follows: for each $\varepsilon > 0$ and $p \in M$, there exists $\delta > 0$ such that $f(M_\delta p) \subset N_\varepsilon p$.

Fact 6.2. $M_\lambda(p)$ is always an open set.

Proof. If $q \in M_\lambda(p)$, then we have $d(p,q) < \lambda$. Let $s = \lambda - d(p,q) > 0$. It suffices to show that $M_s(q) \subset M_\lambda(p)$. In fact,

$$d(x,p) \le d(x,y) + d(q,p) < s + d(p,q) = \lambda.$$

$\square$

Now we use a definition of continuity using only open sets.

Theorem 6.3. *$f : M \to N$ is continuous if and only if for every open set $V \subseteq N$, the pre-image of V is open in M.*

Proof. The easy part is to prove that the open set condition implies continuity. We choose V to be $N_\varepsilon(fp)$; the pre-image is some open set $U \subseteq M$ containing p. Because U is open, there is some neighborhood of p contained in U; that gets mapped into V as desired.

The hard part is to show the ε-δ implies the open set condition. Let U be the pre-image of an open $V \subseteq N$. We wish to show U is open as well. Pick any $p \in U$. Because V is open, there exists $\lambda > 0$ for which $N_\lambda(fp) \subseteq V$. Now continuity implies for some $\delta > 0$ we have $f(M_\delta(p)) \subseteq N_\lambda(fp)$. But $N_\lambda(fp) \subseteq V$ so $M_\delta(p) \subseteq U$, as required. (Uh because $\varnothing$ is clopen, it's okay if either U or V are empty.) $\square$

Remark 6.4. This is very nice. Continuity can be done entirely with open sets, without any regard to a metric.

Corollary 6.5. *The open set condition is equivalent to the closed set condition*[3].

Proof. Take complements. Use the fact that $f^{\text{pre}}(N \setminus K) = M \setminus f^{\text{pre}}(K)$. $\square$

6.2 Homeomorphisms Again

Proposition 6.6. *Consider a homeomorphism $f : M \to N$. Suppose $U \subseteq M$ is open. Then $f(U) \subseteq N$ is open.*

This is not true if f is not a homeomorphism. For example, take a map $x \mapsto x^2$ from $\mathbb{R}$ to $\mathbb{R}$. Then $f((-1,1)) = [0,1)$.

Because f is thus a bijection between open sets, this is also called a *topological equivalence.*

[3] Which says that the pre-image of every closed set is itself closed

6.3 Closures and Interiors

Given $S \subseteq M$, $\lim S = \{p \in M : \exists (x_n) \in S, x_n \to p\}$ We know that $\lim S$ is closed and $S \subseteq \lim S$. Furthermore,

Proposition 6.7. *$\lim S$ is the smallest closed set containing S, in the sense that if $S \subseteq K \subseteq M$ and K is closed, then $\lim S \subseteq K$.*

Proof. $S \subseteq K \Rightarrow \lim S \subseteq \lim K = K$. □

For this reason, we call $\lim S$ the *closure* of S, often denoted $\overline{S}$.

Proposition 6.8. *For any S,*

$$\overline{S} = \bigcap_{S \subseteq K, K \text{ closed}} K.$$

Proof. Obvious. □

Now let's look at the dual situation with open sets.

Definition 6.9. The *interior* of S is

$$\mathring{S} \stackrel{\text{def}}{=} \bigcup_{W \subseteq S, W \text{ open}} W.$$

Definition 6.10. The boundary of S is

$$\partial S \stackrel{\text{def}}{=} \bar{S} \setminus \mathring{S} = \overline{S} \cap \overline{S^c}.$$

Example 6.11. Here is a nontrivial example. Consider $S = \mathbb{Q} \subseteq \mathbb{R}$. Then $\lim S = \mathbb{R}$. On the other hand, $\mathring{S} = \varnothing$. Hence, the boundary of S is $\mathbb{R}$.

Example 6.12. Consider the discrete space M and any subset S. Because all subsets of M are closed, $\overline{S} = S$; i.e. S is closed. Furthermore, all sets are open so $\mathring{S} = S$. Once again $\partial S = \varnothing$.

6.4 Clustering

Definition 6.13. We say a set S clusters at $p \in M$ if for each $\varepsilon > 0$, $S \cap M_\varepsilon(p)$ is infinitely large.

Remark 6.14. Finite sets do not contain cluster points. Consequently, this is subtly different from limits.

Definition 6.15. The set S', the *cluster set* of S, is the set of all cluster points for S. Note that this need not be a subset of S.

Remark 6.16. The union of S' and the *isolated points* of S – that is, the points p which have some neighborhood disjoint from $S - \{p\}$ – is equal to the closure of S. That is, the limit points which are not cluster points are isolated points.

Definition 6.17. A point $p \in S$ is a *condensation point* if there are uncountably many points of S in each $M_\varepsilon(p)$. The set of condensation of points is denote S^*.

Example 6.18. Take $\mathbb{Q}$ as a subset of $\mathbb{R}$ again. $\mathbb{Q}' = \mathbb{R}$, but because $\mathbb{Q}$ is countable we have $\mathbb{Q}^* = \varnothing$.

Then, if $S = \mathbb{R} \setminus \mathbb{Q}$ then $S' = S^* = \mathbb{R}$.

7 September 19, 2013

Exercise 7.1. Show that the intersection of infinitely many open sets need not be open.

Solution.

$$\bigcap_{n\geq 1} \left(\frac{1}{2} - \frac{1}{2^n}, \frac{1}{2} + \frac{1}{2^n} \right) = \left\{ \frac{1}{2} \right\}.$$

□

7.1 Subspaces

Let $N \subseteq M$, where M is a metric space, and suppose we define a metric $d_N : N^2 \to \mathbb{R}$ by $d_N(x, y) = d_M(x, y)$. This is called the *inherited metric*. In that case,

Definition 7.2. N with d_N is a *submetric space* of M.

Remark 7.3. Let $S = \mathbb{Q} \cap [0, 1] \subset \mathbb{Q} \subset \mathbb{R}$. Then S is closed in $\mathbb{Q}$ but not in $\mathbb{R}$. So this is not trivial.

Theorem 7.4 (Inheritance Theorem)**.** *Let $K \subset N \subset M$, where M and N are subspaces. Suppose K is closed in N if and only if there exists a closed L in M such that $K = L \cap N$.*

This implies that "closed sets are inherited with intersection".

Proof. Suppose that L is closed in M, and consider $K = L \cap N$. Consider any sequence (x_n) in K which converges in $x \in N$. Then $x_n \to x$ in M as well, so $x \in L$ by closure of L; therefore $x \in L \cap N = K$.

Conversely, suppose K is closed in N. We claim that we can take L to be the closure of K in M; that is, $L = \lim_M K$. Then $L \cap N = \lim_N K = K$; after all, $L \cap N$ are just those limit points of K which lie in N, which is equal to K by closure. □

Corollary 7.5. *Open sets in N are inherited from M; that is, a set V is open in N if and only if there exists U open in M such that $V = U \cap N$.*

Proof. Take complements. Let $K = N \setminus V$, and consider closed L such that $K = L \cap N$. Then $U = M \setminus L$; note

$$M = L \sqcup U.$$

Then $N = (L \cap N) \sqcup (U \cap N)$. Because $L \cap N$ is closed, we deduce $U \cap N$ is open. □

7.2 Five-Minute Break

Exercise 7.6. Show that a metric space is homeomorphic to the discrete space if and only if all sets are closed.

Solution. Let M be the space, and let D be the metric space with the discrete metric on it. Let f be an arbitrary bijection. Because all sets in D are open, all sets in M are open if and only if f is a homeomorphism. □

7.3 Continuation

Proposition 7.7. *Let $K \subset N \subset M$ where N and M are metric spaces. Suppose further than N is closed in M. Then K is closed in N if and only if K is closed in M.*

Proof. Direct corollary of previous theorem. □

Example 7.8. Consider $[0, 1] \in \mathbb{R}$. If $K \subset [0, 1]$ is closed, then K is closed in $\mathbb{R}$.

On the other hand suppose $V \subset [0, 1] \subset \mathbb{R}$ is open. Then $(\frac{1}{2}, 1] = \left(\frac{1}{2}, 3\right) \cap [0, 1]$ is an open subset of $[0, 1]$. So a half-open interval is open!

7.4 Product Metrics

Let M and N be metric spaces and consider the Cartesian product

$$M \times N.$$

There are a few ways to view $M \times N$ as a metric space. In what follows, we consider $a = (x, y)$ and $b = (x', y')$, so that $a, b \in M \times N$.

- $d_{max}(a, b) = \max\{d_M(x, x'), d_N(y, y')\}$.
- $d_E(a, b) = \sqrt{d_M(x, x')^2 + d_N(y, y')^2}$. This is the *Euclidean* metric, hence the subscript d_E.
- $d_{sum}(a, b) = d_M(x, x') + d_N(y, y')$.

It is easy to check that these are all metrics.

Fact 7.9. For any $a, b \in M \times N$, we have

$$d_{max}(a, b) < d_E(a, b) < d_{sum}(a, b) < 2d_{max}(a, b)$$

Proof. Totally and utterly trivial. □

Why do we care? We have a much better result.

Theorem 7.10. *Consider a sequence in $M \times N$. If it converges to $p \in M \times N$ with respect to any metric, it converges with respect to the other two.*

Musing: what if I select some silly or even asymmetric metric like $\frac{2}{3}d(x, y) + \frac{1}{3}d(x', y')$. Are they comparable? Do they lead to any interesting results?

Proposition 7.11. *$(x_n, y_n) \to (x, y)$ in $M \times N$ equipped with d_{max}, d_E, d_{sum} if and only if $x_n \to x$ and $y_n \to y$.*

Proof. Use d_{max} and remark that $\max\{d(x_n, x), d(y_n, y)\} \to 0$ if and only if $d(x_n, x) \to 0$ and $d(y_n, y) \to 0$. □

7.5 Continuity of Arithmetic

Proposition 7.12. *The map $+ : \mathbb{R}^2 \to \mathbb{R}$ by $(x, y) \mapsto x + y$ is continuous. So are $-$, $\times$, and $\div$, although the domain of $\div$ is $\mathbb{R} - \{0\}$.*

Proof. Many epsilons and deltas appeared. Remark that we can pick any of the three d's; let's use $d_{\max}$. $+$ and $-$ are trivial.

How about $\times$? Suppose $(x_0, y_0) \in \mathbb{R}^2$ and $\varepsilon > 0$ we wish to find δ such that $|x - x_0| < \delta$ and $|y - y_0| < \delta$, then $|xy - x_0y_0| < \varepsilon$. But

$$\begin{aligned} |xy - x_0y_0| &= |xy - xy_0 + xy_0 - x_0y_0| \\ &= |x(y - y_0) + y_0(x - x_0)| \\ &\le |x|\,|y - y_0| + |x - x_0|\,|y| \\ &< \delta\,(|x| + |y_0|) \\ &\le \delta\,(\delta + |x_0| + |y_0|) \end{aligned}$$

Let $\delta = \min\left\{5, \frac{\varepsilon}{5+|x_0|+|y_0|}\right\}$. So, the quantity becomes

$$\begin{aligned} &\le \delta\,(5 + |x_0| + |y_0|) \\ &= \varepsilon. \end{aligned}$$

Similar calculations give $y \mapsto \frac{1}{y}$ is continuous (where $y \neq 0$). □

Lemma 7.13. *The product of continuous function is continuous.*

Proof. Take $M \to \mathbb{R}^2 \to \mathbb{R}$. Take $x \mapsto (f(x), g(x)) \mapsto f(x)g(x)$. Each component is continuous, and the composition of continuous functions is also continuous. $\square$

The manipulation $|xy - xy_0 + xy_0 - x_0y_0|$ is very powerful, supposedly.

8 September 24, 2013

Today: compactness!

For this lecture assume that M is a metric space with a distance function d.

8.1 Completeness

Definition 8.1. A sequence (x_n) in M is Cauchy if $\forall \varepsilon > 0$, $\exists N$ such that $\forall n, m \geq N$ we have

$$d(x_n, x_m) < \varepsilon.$$

Definition 8.2. The space M is *complete* if every Cauchy sequence converges.

We proved that $M = \mathbb{R}$ is complete. It is also easy to check that if M is finite, then M must be complete as well, because we can take $\varepsilon < \min_{x \neq y} d(x, y)$ to force Cauchy sequences to become constant, and thus converge. Similarly, if d is the discrete metric, then M is complete.

On the other hand $M = \mathbb{Q}$ is not complete. For example, we can easily converge to $\sqrt{2} \notin \mathbb{Q}$ in $\mathbb{R}$.

Remark 8.3. Any convergent sequence is Cauchy, but not vice-versa.

Fact 8.4. $\mathbb{R}^2$ is complete, and in general, the product of two complete spaces is product.

Proof. Consider (x_n, y_n) a Cauchy sequence. We find that (x_n) and (y_n) are Cauchy, so $x_n \to x$ and $y_n \to y$ for some x, y, whence $(x_n, y_n) \to (x, y)$. □

Corollary 8.5. *$\mathbb{R}^m$ is complete for each m.*

Proof. Induct. □

Remark 8.6. Homeomorphism does not preserve completeness because $\mathbb{R} \cong (-1, 1)$ through the map $x \mapsto \frac{2}{\pi} \arctan x$. In other words, completeness is not a topological property.

Proposition 8.7. *If S is closed in M and M is complete, then S is complete.*

Proof. If (x_n) is Cauchy in S, it is Cauchy in M, so $x_n \to x \in M$. Because x is closed, $x \in S$. □

8.2 Boundedness

Definition 8.8. Let S be a subset of M. S is bounded if there exists $R > 0$ and $p \in M$ such that $S \subset M_R(p)$.

Remark 8.9. You can strengthen the condition to "for each $p \in M$"; these definitions are equivalent.

Proposition 8.10. *If (x_n) is Cauchy, then $\{x_n : n \in \mathbb{N}\}$ is bounded.*

Proof. Totally obvious! Take $\varepsilon = 1000$, and then note that there are finitely many points not within a distance of 2013 of a giving sufficiently large point.

Here it is written out. IF $d(x_n, x_m) < 1000$ for all $m, n \geq N$ – and such an N exists by the Cauchy condition – then

$$d(x_n, x_N) < 2013 + \max_{k=1,2,\dots,N} d(x_k, x_N).$$

the latter which is a constant. □

Remark 8.11. Being bounded is not a topological property by the same example $\mathbb{R} \cong (-1, 1)$.

Remark 8.12. In fact, for any metric space M, one can find a metric space N such that M is a dense subset in N and inherits its metric for N. This is called the *completion* of M.

8.3 Compactness

Definition 8.13. Let M be a metric space and suppose A is a subset of M. A is *compact* iff for every sequence (x_n) of A, there exists a subsequence $(x_{n_k})_{k\in\mathbb{N}}$ converging to a limit in A as $k \to \infty$.

Theorem 8.14. *Every compact is closed and bounded.*

Remark 8.15. This is important, partially because it is true in every metric space ever.

Proof. Let (x_n) be a sequence in A that converges ta $p \in M$. Now (x_n) has a subsequence such that $x_{n_k} \to x \in A$. But we must have $p = x$. Thus $p \in A$. Consequently (x_n) is closed.

Now suppose for contradiction that A is not bounded. Then $\exists p \in M$ such that for every $R > 0$, $M_R(p)$ does not contain A. Thus there exists (x_n) in A such that $d(p, x_n) \geq n$ by taking R to be successive integers. By compactness we can find a subsequence $x_{n_k} \to q \in A$. Now $d(x_{n_k}, p) \geq n_k$. Now

$$d(x_{n_k}, q) \geq d(x_{n_k}, p) - d(p, q)$$

Because $d(p, q)$ is fixed and $d(x_{n_k}, p)$ grows arbitrarily large, the distance $d(x_{n_k}, q)$ must grow without bound, contradicting $x_{n_k} \to q$. Hence our assumption was wrong and A is bounded. $\square$

Very importantly, the converse is false! Here is an example.

Example 8.16. Take $\mathbb{N}$ with the discrete metric, and let $A = \mathbb{N}$. Because $\mathbb{N}$ is discrete, A is clearly bounded. In fact $\mathbb{N}$ is even complete. But this is far from compact: take the sequence $1, 2, \dots$.

8.4 Trickier Compactness

Here is the "best" example of a compact set. This is a very fundamental theorem; it is the first nontrivial compact set. It is a consequence of the least upper bound property, something very special about $\mathbb{R}$.

Theorem 8.17. $[a, b] \subset \mathbb{R}$ *is compact.*

Proof. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence. Consider

$$S = \{c \in [a, b] : x_n \geq c \text{ for infinitely many } n\}.$$

Obviously $a \in S$ bounded by b. So let $x = \sup S$.

By our selection of x, we find $x - \delta \in S$ and $x + \delta \not{i}nS$. So only finitely many n have $x_n \geq x + \delta$, while infinitely many have $x_n \geq x - \delta$, so it must be the case that $[x - \delta, x + \delta]$ has infinitely many points.

Thus x is a cluster point; this yields our sequence. Here's how. For every $k \geq 1$ we can find $n_k \geq n_{k-1}$ (set $n_0 = 0$ for convenience), with $x_{n_k} \in \left[x - \frac{1}{k}, x + \frac{1}{k}\right]$ (after all, there are infinitely many). So $x_{n_k} \to x$ as required. $\square$

Proposition 8.18. *Let $A \subset M$ and $B \subset N$. If A and B are compact, then so is $A \times B$ as a subset of $M \times N$.*

Proof. Let (x_n, y_n) be a sequence in $A \times B$. There is a subsequence $x_{n_k} \to x \in A$. Now (y_{n_k}) is a sequence in B, so there is a subsequence $y_{n_{k_\ell}} \to y \in B$. $\square$

Triple subscripts!

Theorem 8.19. *Let $S \subset A$ be a closed subset of $A \subset M$. If A is compact then S is compact.*

Proof. Repeat the proof of Cauchy-ness. $\square$

As a result we deduce the Heine-Borel Theorem.

Theorem 8.20 (Heine-Borel Theorem)**.** *Consider $A \subset \mathbb{R}^m$. A is compact if and only if it is closed and bounded in $\mathbb{R}^m$.*

9 September 26, 2013

Exercise 9.1. Give an example of a closed and bounded set which is not compact.

Solution. Take the discrete metric $\mathbb{N}$, and consider the set $\mathbb{N}$. It is not compact because $1, 2, \dots$ fails to have a convergent subsequence. $\square$

The product of two compacts is compact. The triple subscripts apparently confused people, so the proof is being repeated. Because I understood it the first time I am too lazy to copy it down again.

9.1 Compactness Continued

Theorem 9.2 (Bolzano-Weierstrass Theorem). *Every bounded sequence (p_n) in $\mathbb{R}^m$ has a convergent subsequence.*

Proof. There exists a box which contains (p_n). The box is compact. $\square$

Here are some examples of compact sets!

1. Finite sets. This follows by the Pigeonhole principle.
2. $S^2 = \left\{(x, y, z) : x^2 + y^2 + z^2 = 1\right\}$ is compact because it is closed and bounded.
3. The unit ball is closed in $\mathbb{R}^3$.
4. The Hawaiian earring. Set $H_n \subset \mathbb{R}^2$ is the circle with radius n^{-1} and center $(n^{-1}, 0)$, wher $n \ge 0$. Let $H_n^* = H_{-n}$ for each $n \ge 1$, then the Hawaiian earring is
$$H \overset{\text{def}}{=} \bigcup_{n \in \mathbb{Z}^+} (H_n \cup H_n^*).$$
H is clearly bounded, and with some effort we can show H is closed. Hence H is compact.

Definition 9.3. A sequence (A_n) of sets is *nested decreasing* if $A_{n+1} \subset A_n$ for each n.

Theorem 9.4. *If (A_n) is nested decreasing, and each A_n is compact, then $A = \bigcap A_n$ is compact.*

Proof. This is a closed subset of the compact A_1. $\square$

Proposition 9.5. *If each A_n is nonempty, then $A = \bigcap A_n$ is nonempty.*

Proof. Choose an arbitrary $a_n \in A_n$ for each n. Then (a_n) is a sequence in A_1, so (a_n) has a convergent subsequence converging to some point in A_1.

Consider a sequence $(a_{n_k})_k$ converging to some a. Now observe that for any m, $a_m, a_{m+1}, \dots$ all belong to A_m, because A_m is closed. This forces $a \in A_m$ for each m. Hence $a \in A$. $\square$

Theorem 9.6. *If A_n is a nested decreasing nonempty compact and the diameter[4] approaches zero, then $\bigcap A_n$ is a singleton.*

Proof. By the previous proposition, then A is nonempty. Finally remark that any set of more than two points has positive diameter. Because the diameters decrease, we find $\operatorname{diam}(A) \le \operatorname{diam}(A_n)$ which forces $\operatorname{diam} A_n \to 0$. $\square$

[4]The diameter of a set S is the supremum of the distances between pairs of points.

The conclusion becomes false if we drop the condition. For a open and bounded example, define

$$A_n = S^2 \cap \left\{ (x,y) \in \mathbb{R}^2 \mid y \ge x^2 + 1 - \frac{1}{n} \right\}.$$

9.2 Continuity

Proposition 9.7. *Let $f : M \to N$ be continuous and $A \subset M$ compact. Then $f(A)$ is compact.*

Note that this is *not* true for any of the analogous phrases "open", "closed", "continuous" in place of "compact"! Compactness is very nice.

Proof. Let (y_n) be any sequence in B. For each n consider any point $x_n \in A$ such that $f(x_n) \in A$; because $B = f(A)$ at least one such point exists. Then $x_{n_k} \to p \in A$ for some n_k. By continuity, $y_{n_k} = f(x_{n_k}) \to f(p) \in B$ as required. □

Proposition 9.8. *Let $f : M \to N$ be a continuous bijection. If M is compact, then f^{-1} is continuous, and f is a homeomorphism.*

Proof. Suppose on the contrary that fore some $(y_n) \in N$, $y_n \to y \in N$ but $f^{-1}(y_n)$ does not converge to $f^{-1}(y)$.

Let $x_n = f^{-1}(y_n)$ for each n and $x = f^{-1}(y)$. Evidently there exists $\varepsilon > 0$ such that $x_n \notin M_\varepsilon(x)$ for arbitrarily large n. Now let x_{n_k} be the subsequence of x_n of points lying outside this neighborhood. Then there exists a subsequence of *that*, $x_{n_{k_\ell}}$, converging to some x'; clearly $x' \neq x$. Then $y_{n_{k_\ell}}$ converges to $fx' \neq fx = y$, contradicting the fact that y_n converges to y. □

Remark 9.9. Compactness is a topological property. In particular, $\mathbb{R} \ncong [0,1]$.

Definition 9.10. $f : M \to N$ is *uniformly continuous* if for each $\varepsilon > 0$ there is a $\delta > 0$ such that for all $x, y \in M$ with $d_M(x,y) < \delta$, then $d_N(fx, fy) < \varepsilon$.

Note that here, the point δ may depend only on ε! For the usual continuity, δ depends on ε and x.

Theorem 9.11. *If $f : M \to N$ is continuous and M is compact, then f is uniformly continuous.*

Proof next time.

10 October 1, 2013

10.1 Loose Ends

Recall the definition of uniform continuity from the end of the previous lecture.

Example 10.1. Let $f(x) = x^2$, $f : \mathbb{R} \to \mathbb{R}$. Clearly f is continuous, but f is not uniformly continuous – f' grows without bound.

Now we prove our theorem from last time. Again we use sequences.

Proof. Suppose not. Then there exists $f : M \to N$ continuous, with M compact, but f is not uniformly continuous. That means we can find an $\varepsilon > 0$ such that for each $\delta > 0$, we can find $x, y \in M$ such that $d_M(x, y) < \delta$ but $d_N(fx, fy) \ge \varepsilon$.

We pick $\delta = n^{-1}$ for each $n \in \mathbb{Z}^+$. Thus, we find $\exists x_n, y_n \in M$ such that $d_M(x_n, y_n) < n^{-1}$ but $d_N(fx_n, fy_n) \ge \varepsilon$. By compactness, we obtain $x_{n_k} \to x$. Because $d(x_{n_k}, y_{n_k}) \to 0$, we derive that $y_{n_k} \to x$ as well.

By continuity, $d_N(fx_{n_k}, fx) \to 0$ and $d_N(fy_{n_k}, fx) \to 0$. Yet $d_N(fx_{n_k}, fy_{n_k}) \ge \varepsilon$, which is a contradiction. $\square$

Let M be a (nonempty) compact, and consider $f : M \to \mathbb{R}$ be continuous. Then $f(M)$ is compact (and nonempty) in $\mathbb{R}$; hence it is closed and bounded. Then $\sup f(M)$ is well-defined and belongs to f itself; that is, f achieves a maximum and minimum.

10.2 Connectedness

Let M be a metric space.

Definition 10.2. A *separation* of a subset $S \subset M$ is a division of S as $S = A \sqcup B$ such that $A, B \neq \varnothing$, but $\overline{A} \cap B = A \cap \overline{B} = \varnothing$, where the closures are taken with respect to M.

Definition 10.3. A set S is *connected* if no separation of S exists. Else it is *disconnected.*

Remark 10.4. Because $\overline{A} \subset S = A \sqcup B$, we find that $\overline{A} \cap B = \varnothing \Leftrightarrow \overline{A} \cap S = A$; that is, A is closed in S; hence B is open in S. Similarly, B is closed and A is open (in S again). So actually, it's equivalent to say A and B are clopen in S.

Example 10.5. $\mathbb{R}$ is connected.

Proof. Suppose on the contrary that $\mathbb{R} = A \sqcup B$, where A and B are both clopen. Consider any maximal (by inclusion) open interval $(a, b) \subseteq B$. We claim $b = +\infty$. Assume otherwise. Because B is closed, $b \in B$. This forces $b \in B$. Now, because B is open, $\exists \delta > 0$ such that $(b-\delta, b+\delta) \in B$. Then $(a, b+\delta)$ is a larger interval, contradiction. So $b = +\infty$. Similarly $a = -\infty$. Thus $B = \mathbb{R}$, which is absurd. $\square$

Proposition 10.6. *If $f : M \to N$ is continuous, and M is connected, then $f(M)$ is connected as well.*

Proof. We show that if $f(M)$ is disconnected as $A \sqcup B$, then M is disconnected. Note that A and B are clopen in $f(M)$. Now $f^{\text{pre}}(A)$ is clopen, as is $f^{\text{pre}}(B)$. Yet they are disjoint. Hence M is disconnected. $\square$

Corollary 10.7. *Connectedness is a topological property.*

Remark 10.8. The converse is not true; that is, $f : M \to N$ and N connected does not imply M connected, even if f is surjective. For, what if M consisted of two disjoint disks, and N was the singleton set?

Example 10.9. Let $f : \mathbb{R} \to \mathbb{R}$ by $x \mapsto \sin x$. Thus $[-1,1] = \sin \mathbb{R}$ is connected.

Example 10.10. Let $f : \mathbb{R} \to \mathbb{R}^2$ by $x \mapsto (\cos x, \sin x)$. Then the image S^1 is connected.

Proposition 10.11. *Let $S \subseteq M$ be connected. Then for any $S \subseteq T \subseteq \overline{S}$, then T is connected.*

The proof of this is in the book. It's not very surprising.

Remark 10.12. $(0,1)$ is connected because it is homeomorphic to $\mathbb{R}$. Hence $(0,1]$ is connected.

Example 10.13. Consider

$$S = \left\{x, y \in \mathbb{R}^2 \mid 0 < x \le 1, y = \sin x^{-1}\right\}.$$

This is a connected set. What is its closure? It turns out that $\overline{S} - S$ is $\{0\} \times [-1,1]$, and so counterintuitively, the set

$$T = S \cup \{0\} \times [-1,1].$$

is connected!

10.3 Path-Connectedness

Definition 10.14. A *path* in M is a continuous function $f : [a,b] \to M$.

Definition 10.15. M is *path-connected* if $\forall p, q \in M$, there exists a path $f : [a,b] \to M$ for which $f(a) = p$, $f(b) = q$.

Example 10.16. The above set T is connected, but not path-connected!

Proposition 10.17. *Let $S = \bigcup_\alpha S_\alpha$, where each $S_\alpha \subseteq M$ and $\exists p \in S$ such that $p \in S_\alpha$ for each α. Then S is connected.*

Proof. If not, suppose $S = A \sqcup B$ and assume without loss of generality that $p \in A$. Observe

$$(A \cap S_\alpha) \sqcup (B \cap S_\alpha) = S_\alpha.$$

It is not hard to check that these are both clopen in S_α because A and B are open. Now $p \in A \cap S_\alpha$, so this forces $B \cap S_\alpha$ for each α. This implies $B = \varnothing$, contradiction. $\square$

Remark 10.18. This is versatile since the union need not be countable. Hence, S^2 is connected, because it can be expressed by unions of S^1 (great circles) each passing through the South Pole.

Proposition 10.19. *Let U be open in $\mathbb{R}^m$. If U is connected, then U is path-connected.*

This seems intuitively obvious.

Proof. Pick $p \in U$. Consider

$$V = \{q \in U : \exists \text{ path in } U \text{ from } p \text{ to } q\}.$$

We claim V is clopen in U. The fact that V is open follows from the fact that U is open. For closedness, assume q is a limit of V. Again by openness, we can find a neighborhood $M_r(q)$ contained in U, and because q is a limit point some point in that neighborhood is path-connected to p. Hence V is clopen in U; because $p \in V$ this forces $V = U$, as desired. $\square$

This is called a *open-and-closed* argument. This is kind of a "control theory" approach.

11 October 4, 2013

Today we discuss open coverings, and how to define compactness from that perspective.

11.1 Open Coverings

Definition 11.1. A collection $\mathcal{U}$ of open sets whose union contains a set A is called an *open covering* of A. The open sets are called *scraps*.

Definition 11.2. A is *open covering compact* if each open covering of A induces to a finite subcovering; i.e. $\forall \mathcal{U}$, there exists $U_1, U_2, \dots, U_n \subseteq \mathcal{U}$ such that $A \subseteq \bigcup_{i=1}^n U_i$.

Example 11.3. Consider $B = \left\{ \frac{1}{n} \mid n = 1, 2, \dots \right\}$ living in $\mathbb{R}$, and let $A = B \cup \{0\}$. Show that A is open covering compact, but

Proof. Some open set $(u, v) \in \mathcal{U}$ covers 0. Only finitely many elements of A now exist outside $(-u, v)$, and we can cover the elements of A exceeding v easily.

On the other hand, one can construct open intervals around each element of B which do not contain any other element of B. □

11.2 The Main Result

Theorem 11.4. *A is open covering compact if and only if it is sequentially compact.*

Remark 11.5. Afterwards, we will just use *compact* to refer to both.

One direction is easy.

Proof that covering compact implies sequentially compact. Assume (a_n) is a sequence in A, yet no subsequence converges. That implies that for each $p \in A$, some neighborhood of p contains only finitely many points of (a_n); that is,

$$\forall p \in A \exists r(p) > 0 : M_{r(p)}(p) \cap \{a_n\} \text{ is finite}$$

Now consider the collection of $M_r(p)$; that is

$$\mathcal{U} \stackrel{\text{def}}{=} \{M_r(p) \mid r = r(p), p \in A\}.$$

This is an open covering, so it has a finite subcovering, say $M_{r_i}(p_i)$ for $i = 1, 2, \dots, m$. Now each of these scraps only contains finitely many $\{a_n\}$, so their union only contains finitely many, which is a contradiction. □

As usual I'm ignoring the stupid edge case where some number appears infinitely many times, because such sequences are compact anyways. But there is a slight difference between "contains infinitely many a_n" and "contains a_n for infinitely many n".

11.3 Lebesgue Numbers

First, we state something obvious.

Fact 11.6. Let $p \in M$ and $x \in M_{\frac{1}{2}r}(p)$. Then

$$p \in M_{\frac{1}{2}r}(x) \subset M_r(p).$$

Proof. Totally and utterly trivial. Obviously $p \in M_{\frac{1}{2}r}(x)$, and apply the triangle inequality to any $q \in M_{\frac{1}{2}r}(x)$. □

Definition 11.7. If $\mathcal{U}$ is an open covering of $A \subset M$. A real number $\lambda > 0$ is called a *Lebesgue number* $\mathcal{U}$ if for each $p \in A$, there exists a scrap $U \in \mathcal{U}$ for which $M_\lambda(p) \subset U$.

This is a measure of the "coarseness" of the open covering $\mathcal{U}$.

Lemma 11.8 (Lebesgue Number Lemma)**.** *If A is sequentially compact set, then all open coverings of A have a Lebesgue number.*

Proof. Suppose on the contrary that for every $\lambda > 0$, there exists a $p \in A$ such that $M_\lambda(p)$ is not contained in any scrap. Then by taking $\lambda = \frac{1}{n}$ for $n = 1, 2, \dots$, we can find a point $a_n \in A$ such that $M_{\frac{1}{n}}$ is not in any scrap.

Then there exists a subsequence $a_{n_k} \to p \in A$. Then $\exists U_p \in \mathcal{U}$ such that $p \in U_p$. Now there exists $r > 0$ such that $M_r(p) \subset U_p$. Evidently we can find an a_{n_k} such that $M_{\frac{1}{n_k}}(a_{n_k}) \subset M_r(p)$ because $d(a_{n_k}, p) \to 0$ and $\frac{1}{n_k} \to 0$. This becomes a contradiction. □

11.4 Finishing the Proof

Using this, we can now establish the main result.

Proof that sequentially compact implies covering compact. Let $\mathcal{U}$ be a open covering of A, and let λ be a Lebesgue number. Choose $a_1 \in A$ and $U_1 \in \mathcal{U}$ such that $M_\lambda(a_1) \subseteq U_1$. If $\{U_1\}$ is a finite subcover by some stroke of luck, then we're done. Otherwise pick $a_2 \in A \setminus U_1$ and pick a U_2 such that $M_\lambda(a_2) \subset U_2$. Rinse and repeat, selecting points $a_1, a_2, \dots$ and $U_1, U_2, \dots$ such that $M_\lambda(a_i) \subset U_i$ and $a_i \in A - \cup_{j=1}^{i-1} U_i$ for all i.

Assume on the contrary that this results in an infinite sequence (a_n). Then compactness of A implies $a_{n_k} \to p \in A$. This is immediately bad because a pair of the a_{n_k} grow close to each other, but they must be at least λ apart. □

Musing: you can weaken this to "every open covering has a Lebesgue number" and "every sequence has a Cauchy subsequence".

Note that this means we can replace completeness completely in terms of open sets, ignoring sequences. Once again, this is exactly what's done in general topology.

11.5 Generalizing the Heine-Borel Theorem

Recall the Heine-Borel theorem which states that for subspaces of $\mathbb{R}^m$, "closed and bounded" is equivalent to compactness. We want to generalize.

Definition 11.9. A is *totally bounded* if for each $r > 0$, there exists a finite covering of A by neighborhoods of radius r.

Theorem 11.10. *Suppose that M is complete. Then A is closed and totally bounded if and only if A is compact.*

Proof. First suppose A is compact. Then A is closed, and for any $r > 0$, the open covering $\mathcal{U} = \{M_r(p) \mid p \in A\}$ has a finite subcover as desired.

The converse is NOT EASY. □

12 October 8, 2013

Today: Cantor sets!

12.1 Perfection

Definition 12.1. A metric space is *perfect* if every point is a cluster point.

Example 12.2. Both $\mathbb{R}$ and $\mathbb{Q}$ are perfect.

Theorem 12.3. *Let M be a perfect, nonempty complete metric space. Then M is uncountable.*

Proof. Assume not. Evidently M is denumerable, so let $M = \{x_1, x_2, x_3, \dots\}$.

Define
$$\hat{M}_r(p) = \{x \in M \mid d(x,p) \le r\}$$
which we colloquially call a *closed neighborhood*. Choose any point $y_1 \in M - \{x_1\}$ and define $r_1 = \min\left\{1, \frac{1}{2}d(x_1, y_1)\right\}$. We see that $x_1 \in \hat{M}_{r_1}(y_1)$.

Now choose $y_2 \in M_{r_1}(y_1)$ such that $y_2 \neq x_2$; this happens to be possible because the space is perfect. Certainly $y_2 \neq x_1$. Then select
$$r_2 = \min\left\{\frac{1}{2}, \frac{1}{2}d(x_2, y_2), r_1 - d(y_2, y_2)\right\}.$$
Evidently
$$\hat{M}_{r_2}(y_2) \subset \hat{M}_{r_1}(y_1).$$
Repeating this construction, we derive a sequence of points $y_1, y_2, \dots$ and radii $r_1, r_2, \dots$ for which
$$\hat{M}_{r_1}(y_1) \supset \hat{M}_{r_2}(y_2) \supset \hat{M}_{r_3}(y_3) \supset \dots$$
This is an infinite nested sequence of closed neighborhoods whose radii tend to zero. Consider the sequence (y_n). Evidently $d(y_m, y_n) \to 0$ as $m, n \to \infty$; that is, (y_n) is Cauchy. Then $y_n \to p$, and by closure we deduce that p belongs to $\bigcap \hat{M}_{r_i}(y_i)$.

Yet $p \neq x_n$ for any n, because $x_n \notin \hat{M}_{r_n}(y_n)$ by construction, breaking the assumption that x_i was a full enumeration of the points of M. □

Corollary 12.4. *$\mathbb{R}$ is uncountable.*

12.2 Cantor Dust

Definition 12.5. A metric space is *totally disconnected* if for each point $p \in M$ and $r > 0$ there exists a nonempty clopen subset U such that $p \in U \subset M_r(p)$.

Theorem 12.6. *There exists a subset $C \subset \mathbb{R}$ which is nonempty, compact, perfect, and totally disconnected.*

In particular, C is complete, C is also uncountable. The last property implies that C is totally disconnected.

Remark 12.7. While this is called the Cantor Set, a man named Smith in England discovered the set and published a paper 20 years before Cantor. Regrettably, no one paid attention to it.

Remark 12.8. Let C be the Cantor Set, and let $\chi_C(x) = 1$ if $x \in C$ and 0 otherwise. $\chi_C(x)$ is discontinuous at infinitely many points, and yet is Riemann integrable.

Here is the construction for the notorious Cantor Set. Begin with $C^0 = [0, 1]$. Delete the interval $\left(\frac{1}{3}, \frac{2}{3}\right)$. That is,

$$C^1 = \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right].$$

We repeat this procedure ad infinitum – delete the middle thirds to obtain a sequence $C^2, C^3, \dots$ of intervals. Here, C^n has 2^n interval each of length 3^{-n}.

Evidently $C^1 \supset C^2 \supset \dots$. Now we define the Cantor Set

Definition 12.9. The Cantor Set C is defined by $C = \bigcap_{n=1}^{\infty} C^n$.

Claim 12.10. C is the desired set.

Proof. We check this.

(a) $0 \in C$, so C is nonempty.

(b) C is closed (as it is the intersection of closed sets)

(c) We wish to show that for any $p \in C$, $r > 0$, we have $(p-r, p+r) \cap C$ is infinite. Just select n such that $3^{-n} < \varepsilon$. Since $p \in C^n$, p is in an interval $I \subset C^n$ contained in $(p-r, p+r)$. Now just remark that each interval contains infinitely many endpoints.

(d) The proof that C is totally disconnected is similar to the above. Pick n such that $3^{-n} < r$, and select an interval I in C^n containing p. Obviously I is closed in C^n, and yet I is also open in C^n! So I is a clopen subset of C^n. We claim $I \cap C$ is clopen in C. Reason: inheritance principle. C is a subset of C^n, and so I is clopen in C.

$\square$

Now here's something scary: the sot of all endpoints of intervals is finite. But the Cantor set is uncountable!

Remark 12.11. You can show directly that C contains no interval simply from the fact that C has length 0, while any intervals have positive length.

12.3 Shadows of the Cantor Set

Here is a ridiculous theorem.

Theorem 12.12. *Let M be a metric space which is nonempty, compact, perfect, and totally disconnected. Then M is hoeomorphic to the Cantor set.*

Corollary 12.13. *If C is the Cantor set, then $C \cong C \times C$!*

Remark 12.14. The Chinese multiplication table rhymes, according to someone in the class.

12.4 Addresses

Let's assign each point in the Cantor Set with an address, according to the rules

$$0 = \text{left} \quad \text{and} \quad 2 = \text{right}.$$

A sequence of length n, consisting of 0's and 2's, will now designate an interval in C^n. For example, C_{00} is the leftmost interval in C^2.

Now let's take an infinite string ω of 0's and 2's. Then we can find a point $p(\omega) \in C$. Let $\omega|_n$ denote the first n characters of ω. Now the point

$$p(\omega) = \bigcap C_{\omega|_n}$$

is indeed contained in each C^n for each n; in fact this is a nested decreasing sequence with decreasing radii. Hence $p(\omega)$ is uniquely defined and indeed belongs to C.

Conversely, every point corresponds to an address. (Check this.) Thusly we obtain a bijection from C to infinite zip codes.

Remark 12.15. In fact, $p(\omega)$ corresponds to the number in $[0, 1]$ whose base-3 expansion corresponds to ω.

13 October 10, 2013

More about Cantor sets!

Again define $C = \bigcap_{n\geq 1} C^n$ the middle-thirds Cantor set.

13.1 Zero Sets

Definition 13.1. For a set $S \subset \mathbb{R}$, we say S is a *zero set* if $\forall \varepsilon > 0$ there exists a covering of S by countably many open intervals whose total length is less than ε.

If we let the intervals be denoted be (a_i, b_i), then the length is merely $\sum_{i\geq 1} b_i - a_i$.

Proposition 13.2. *C is a zero set.*

Proof. For any fixed $\varepsilon > 0$, we can find an n for which C^n has length less than $\frac{1}{2}\varepsilon$. These intervals are closed, so if we dilate the closed intervals by a factor of 2 to an open interval, this will cover the Cantor set. $\square$

Definition 13.3. Let F denote the *fat Cantor set* as follows: at the nth step, we delete the middle fraction of each remaining interval in such a way that the total discard at each step is $\frac{1}{3}\left(\frac{1}{6}\right)^n$ (really?).

Remark 13.4. Being a zero set is not a topological property. $C \cong F$ but F has positive measure.

13.2 Cantor Surjection Theorem

The following theorem is terrifying.

Theorem 13.5 (Cantor Surjection Theorem)**.** *Let M be a compact metric space. Then there exists a surjective continuous map $\sigma : C \to M$, where C is the standard middle-thirds Cantor set.*

Definition 13.6. A function $\tau : [a, b] \to RR$ is called a *Peano curve* if $\tau([a, b])$ has non-empty interior.

The existence of the Peano curves is counterintuitive, and yet it is an easy consequence of the Cantor Surjection Theorem. Here is how it is done. A *gap interval* is an interval (a, b) for which $a, b \in C$ but $(a, b) \cap C = \varnothing$. Evidently we can find a function $\sigma C \to B^2$ surjective. Then we extend σ to $\tau : [0, 1] \to B^2$ by

$$\tau(x) = \begin{cases} \sigma(x) & \text{if } x \in C \\ (1-t)\sigma(a) + t\sigma(b) & \text{if } x \text{ lies in a gap interval } (a, b) \text{ and } x = (1-t)a + tb,\ 0 \le t \le 1. \end{cases}$$

It is immediate that τ is continuous, and this yields a Peano curve.

Here is the idea of the proof of the surjection theorem. The goal is to obtain M as the intersection of nested decreasing intervals, and then try to link the constructions.

We need some machinery for this...

13.3 Pieces and Filtrations

Definition 13.7. A *piece* of M is a compact nonempty subset.

Lemma 13.8. *M can be "broken" into small pieces; i.e. $\forall \varepsilon > 0$ there exists finitely many pieces (not necessarily disjoint) of M whose union is M and each piece has diameter less than ε.*

Proof. Look at the open covering of M via

$$\left\{ M_{\frac{1}{3}\varepsilon}(x) \mid x \in M \right\}.$$

This has a finite subcover $M_{\frac{1}{3}\varepsilon}(x_i)$ as $i = 1, 2, \dots, n$. The diameter is less than $\frac{2}{3}\varepsilon$. Now close them; the diameter is still at most $\frac{2}{3}\varepsilon < \varepsilon$. □

Decompose M into pieces, and let $\mathcal{M}_1$ be the collection of pieces when $\varepsilon = 1$. Then let $\mathcal{M}_2$ be the division of each of the pieces of M_1 by pieces with diameter less than $\frac{1}{2}$, and repeat this procedure to obtain a sequence $(\mathcal{M}_k)_{k \in \mathbb{N}}$. By construction, each member of $\mathcal{M}_n$ is a union of members of $\mathcal{M}_{n+1}$ and a subset of some piece of $\mathcal{M}_{n-1}$.

Definition 13.9. The sequence $(\mathcal{M}_k)$ is called a *filtration* of M.

13.4 Words

Definition 13.10. Let $W(n)$ be the sequence of words in 2 letters 0 and 2 of length n.

Example 13.11. $W(2) = \{00, 02, 20, 22\}$.

Evidently $|W(n)| = 2^n$ for any n. Now we remark that if $\#S \le 2^n$, then there exists a surjection $W(n) \to S$.

So there exists an n_1 so that for which we can construct a surjection $W(n_1) \to \mathcal{M}_1$; we can label each of the pieces of $\mathcal{M}_1$ by labels such that each label is is used. Now we pick an n_2 sufficiently large and construct an extended surjection

$$W(n_1 + n_2) \to \mathcal{M}_2$$

such that the first n_1 characters of any letter identifies a corresponding parent piece.

Remark that the diameters of the pieces tends to zero, and it's compact and nested. To be explicit, consider an address $\alpha = \alpha_1 \alpha_2 \dots$ and denote $\alpha|_k = \alpha_1 \alpha_2 \dots \alpha_{n_1+n_2+\dots+n_k}$. Let $M_{\alpha|_k}$ denote the associated piece. We obtain

$$M_{\alpha|_k} \supset M_{\alpha|_{k+1}} \supset \dots$$

and so each point α is associated to the unique point

$$p(\alpha) = \bigcap_{k=1}^{\infty} M_{\alpha|_k}.$$

Because each point can be expressed with an address, this gives us a map. To be even more explicit, let $\beta(x)$ denote the address of a point $x \in C$. Then we simply compute

$$x \mapsto \beta(x) \mapsto p(\beta(x)) \in M.$$

This map is clearly surjective, since each point of M has an address, and β is a bijection.

Finally we need to check this is continuous. Just use sequences. Two points which are "close" have similar addresses, meaning they are in the same piece in M.

13.5 Other Properties

This proof can also be adapted to show that any set which is (i) compact, (ii) nonempty, (iii) perfect, and (iv) totally disconnected. As we said before,

$$C \cong C \times C.$$

In particular, $\dim C = 0$ since $\dim C = \dim(C \times C) = 2 \dim C \Rightarrow \dim C = 0$.

It is possible to draw a space homeomorphic to the Cantor set in $\mathbb{R}^2$ such that each vertical line hits the Cantor set.

Antoine's necklace is even worse.

14 October 15, 2013

Review.

14.1 Midterm Review for Chapter 1

Recall that $\mathbb{R}$ is defined as the set of cuts in $\mathbb{Q}$ (via Dedekind).

- The least upper bound property.
- All Cauchy sequences converge.
- Any bounded monotone sequence converges.

Surprisingly, these are all equivalent.

Other things from Chapter 1.

- Recall the definition of $\mathbb{R}^m$, and the notion of the *dot product*.
- Cauchy-Schwartz, which holds for any inner product[5]. The proof is on the discriminant $0 \le \langle x+ty, x+ty \rangle$.
- Convexity in $\mathbb{R}^m$.
- Cardinality - injection, surjection, bijections.
- $\mathbb{R}$ is uncountable while $\mathbb{Q}$ is countable. [6]

14.2 Midterm Review for Chapter 2

Metric spaces:

- The "best" metric spaces are $\mathbb{R}$ and $\mathbb{R}^m$, and its subsets with the *inherited metric* (e.g. Hawaiian earing.)
- The "worst" metric space is the discrete space.

Remember information about sequences, subsequences, and the definition of convergence. Know the definition of a continuous function $f : M \to N$. There are four continuity definitions:

- Sequences. Continuity of f is equivalent to $x_n \to x$ then $f(x_n) \to f(x)$.
- ε-δ definition. For all $\varepsilon > 0$ and $p \in M$, $\exists \delta > 0$ such that $d_M(p,q) < \delta \Rightarrow d_N(fp, fq) < \varepsilon$.
- Open set condition. The pre-image of any open set is open.
- Closed set condition. The pre-image of any closed set is closed.

We also have the notion of *uniform continuity*. For every $\varepsilon > 0$, there exists $\delta > 0$ such that for each p, q with $d_M(p,q) < \delta$, we must have $d_N(fp, fq) < \varepsilon$. An example of a function failing this criteria is $x \mapsto x^2$ with $\mathbb{R} \to \mathbb{R}$. An example of a *bounded* function failing this criterion is $\sin \frac{1}{x}$.

[5] symmetric bilinear

[6] Schroeder Bernstein Theorem.

Know the definition of a homeomorphism. Homeomorphisms are surprisingly hard to prove. For example, it is not easy to prove that $B^m \ncong B^k$ when $m \neq k$ (here B^n is a ball). Note that any property defined solely in terms of open/closed sets and cardinality is automatically a topological property. This is because homeomorphisms biject their respective topologies!

Key properties of *open sets*, which permit general topology to take form:

1. $\varnothing$ and M are both open.
2. Any union of open sets is open.
3. Finite intersections of open sets are open.

Needless to say, know the definition of an open set.

Closed sets can be defined in two ways – as the complement of an open set, or as points which contain all their limit points. Recall that for a set $S \subset M$, $\lim S$ is the set of limits of all sequences in S that converge in M. Then S is closed if and only if $S = \lim S$. We derive that

1. $\varnothing$ and M are both closed.
2. Any intersection of closed sets is closed.
3. Any finite union of closed sets is closed.

For $\mathbb{R}$, the open sets are very special - they are countable unions of disjoint open intervals.

Exercise 14.1. What can be said about continuous functions $f : \mathbb{R} \to \mathbb{Q}$?

Recall that the *connected image is connected* of a continuous function.

Recall that M is disconnected if and only if M has proper clopen subset. A subset $S \subset M$ is disconnected if S has a separation $S = A \sqcup B$ into sets A and B so that $\overline{A} \cap B = A \cap \overline{B} = \varnothing$.

14.3 Midterm Review: Compactness

Definition 14.2. A set $S \subseteq M$ is compact if each sequence (x_n) in S has a convergent subsequence.

Definition 14.3. A set $S \subseteq M$ is compact if every open cover has a finite subcover.

Compact implies closed and bounded. The converse is false. Continuous images of compacts are compact, so compactness is a topological property.

Fact 14.4. The Cartesian product of two compacts is compact.

Of course, we have Heine-Borel and Bolzano-Weierstrass.

Nested decreasing sequences: suppose $(S_n)_{n\in\mathbb{N}}$ is nested decreasing, and each is compact. Then $\cap S_n$ is clearly compact as a closed subset of a compact. Furthermore, if each $S_n \neq \varnothing$, then $\bigcap S_n$ is nonempty.

The continuous image of a compact is compact. A continuous function on a compact is uniformly continuous.

Cantor Set lore. Perfect sets: $M' = M$. Totally disconnected: every neighborhood contains a proper clopen subset.

15 October 17

Midterm.

Average score: 45.

My score: 97.

16 October 22

16.1 Definition

We will now be returning to real-valued functions.

Definition 16.1. A function $f : (a, b) \to \mathbb{R}$ is differentiable at x if and only if

$$\lim_{t \to x} \frac{f(t) - f(x)}{t - x} = L \in \mathbb{R}$$

exists; in that case, L is the *derivative.*

Other notations of this include $\frac{\Delta f}{\Delta x} = \frac{f(x+\delta x)-f(x)}{\delta x}$.

16.2 Immediate Consequences

There are a few "basic" facts about a derivative.

Fact 16.2. f differentiable at x implies that f is continuous at x.

Proof. For the limit to exist, $f(t) - f(x) \to 0$ must hold as $t \to x$. □

Fact 16.3 (Sum Rule)**.** If f and g are differentiable at x, then

(i) $f + g$ is differentiable at x and $(f + g)'(x) = f'(x) + g'(x)$.

(ii) $f \cdot g$ is differentiable at x, and $(f \cdot g)'(x) = f'(x)g(x) + f(x)g'(x)$.

(iii) f/g is differentiable at x assuming $g(x) \neq 0$. Some expression.

Fact 16.4 (Chain Rule)**.** Consider $f : (a, b) \to \mathbb{R}$ differentiable at x and $g : (c, d) \to \mathbb{R}$. Suppose g is differentiable at y and $y = f(x)$. Then $(g \circ f)$ is differentiable at x and

$$(g \circ f)'(x) = g'(y) \cdot f'(x).$$

Proof. We would hope that

$$\frac{\Delta g}{\Delta x} = \frac{\Delta g}{\Delta f}\frac{\Delta f}{\Delta x}$$

so that $\frac{\Delta g}{\Delta f} \to g'(y)$ and $\frac{\Delta f}{\Delta x} \to f'(x)$. The danger here is if Δf is zero.

Let us write

$$\Delta g = g(y + \Delta y) - g(y)$$

where we ignore the connection between y and f. Then whenever $\Delta y \neq 0$, we have

$$\frac{\Delta g}{\Delta y} = g'(y) + \sigma(\Delta y)$$

where σ is some "remainder" function. As $\Delta y \to 0$, $\sigma(\Delta y) \to 0$ as well.

Now we *define* $\sigma(0) = 0$! Then

$$\Delta g = (g'(y) + \sigma(\Delta y))\Delta y$$

holds true for all Δy, even $\Delta y = 0$.

Okay, let us now write

$$\frac{\Delta g}{\Delta x} = \frac{\Delta y(g'(y) + \sigma(\Delta y))}{\Delta x} = \left(g'(y) + \sigma(\Delta y)\right)\frac{\Delta f}{\Delta x}.$$

Here $\Delta f = \Delta y$. Note that we have made no limits so far. This is all very innocent. Now the point is that as $\Delta x \to 0$, then $\Delta y \to 0$ by continuity of f at x. Now,

$$g'(y) + \sigma(\Delta y) \to g'(y)$$

and

$$\frac{\Delta f}{\Delta x} \to f'(x).$$

□

As an aside, this proof works in higher dimensions.

16.3 Geometry of Derivatives

Definition 16.5. We say that $f : (a, b) \to \mathbb{R}$ is *differentiable* if it is differentiable at each point.

Theorem 16.6 (Mean Value Theorem)**.** *Suppose $f : [a, b] \to \mathbb{R}$ is continuous and the restriction of f to (a, b) is differentiable. Then there exists a $\theta \in (a, b)$ such that*

$$f(b) - f(a) = f'(\theta)(b - a).$$

We say a function has the *mean value property* if this turns out to be true.

Proof. The intuition is secants. We refer the reader to the diagram in the book.

Define

$$\phi(x) = f(x) - S(x - a).$$

Evidently $\phi(a) = f(a)$ and $\phi(b) = f(a)$. Furthermore, ϕ is differentiable, since f and $S(x - a)$ are both differentiable. Moreover, ϕ is continuous on $[a, b]$.

By previous work with general metric spaces, ϕ attains a maximum value. If ϕ is constant, then nothing is interesting. Otherwise, there is either a local maximum or a local minimum in (a, b) (we need $\phi(b) = \phi(a)$ for this; if a minimum is at a then it's also at b, and hence neither is the maximum).

Let $\theta \in (a, b)$ be a point with $M = \phi(\theta)$ an absolute maximum, say. Then one can check that this forces $\phi'(\theta) = 0$. After all,

$$\frac{\phi(t) - \phi(\theta)}{t - \theta} \to \phi'(\theta) \quad t \to \theta.$$

The numerator is always nonpositive. The limit must now be zero, because if $t \to \theta^+$ the fraction is nonpositive; if $t \to \theta^-$ the fraction is nonnegative.

Thus, $0 = \phi'(\theta) = f(\theta) - S$ so $f(\theta) = S$ as desired. □

16.4 L'Hospital's Rule

The MVT can be modified as follows.

Theorem 16.7 (Ratio MVT)**.** *Let $f, g : [a, b] \to \mathbb{R}$ be continuous and differentiable on (a, b). Then there exists $\theta \in (a, b)$ such that*

$$\Delta f g'(\theta) = f'(\theta)\Delta g$$

where $\Delta f = f(b) - f(a)$ and $\Delta g = g(b) - g(a)$.

In particular, if $\Delta g \neq 0$ and $g'(x) \neq 0\ \forall x$ then

$$\frac{\Delta f}{\Delta g} = \frac{f'(\theta)}{g'(\theta)}$$

Proof. Consider

$$\Phi(x) = \Delta f(g(x) - g(a)) - (f(x) - f(a))\Delta g.$$

Clearly Φ is "well-behaved"; it is differentiable. Evidently $\Phi(a) = \Phi(b) = 0$.

Then IVT implies that there exists $\theta \in (a, b)$ such that

$$0 = \Phi'(\theta) \cdot (b - a).$$

Therefore,

$$0 = \Phi'(\theta) = \Delta f g'(\theta) - \Delta g f'(\theta)$$

and we are done. □

Now we present L'Hospital's Rule. Try to not miss conditions!

Theorem 16.8 (L'Hospital's Rule)**.** *Let f and g be functions differentiable on (a, b). Suppose that $f(x) \to 0$ and $g(x) \to 0$ as $x \to b$. Furthermore, suppose $g(x) \neq 0$ and $g'(x) \neq 0$ for any $x \in (a, b)$.*

Suppose finally that $\frac{f'(x)}{g'(x)} \to L$ as $x \to b$, then

$$\frac{f(x)}{g(x)} \to L$$

as $x \to b$.

The condition $g(x) \neq 0$ is IMPORTANT here!

Here is an intuitive description of the proof. Consider a fixed x. Then we can find a t much closer to b than x. Because $g \to 0$, then $g(t)$ is very close to zero, and negligible in comparison to the $g(x) \neq 0$. (Note that $g(x) \neq 0$ is important here!) Likewise, $f(x) - f(t)$ is close. So, for any x, we can find a $t = t(x)$ such that

$$\frac{f(x)}{g(x)} = \frac{f(x) - 0}{g(x) - 0} \approx \frac{f(x) - f(t)}{g(x) - g(t)} = \frac{f'(\theta)}{g'(\theta)}$$

for some $\theta \in (x, t)$. As $x \to b$ and $\theta \to b$ much faster, we have $\theta \to b$ and so $\frac{f'(x)}{g'(x)} \to L \Rightarrow \frac{f(x)}{g(x)} \to L$.

This proof is extensible to $x \to a$ and $x \to \pm\infty$. You can also modify the proof to work with $f(x) \to \infty$ and $g(x) \to \infty$. The key idea in all of these is the *advance guard* metaphor.

Be very careful to ensure $f(x) \to 0$ and $g(x) \to 0$ before applying this rule.

16.5 Continuity of the Derivative

The derivative $f'(x)$, even if it exists, need not be continuous. However, it turns out that the intermediate value property still holds.

Definition 16.9. A function with the Intermediate Value Property is called *Darboux continuous.*

That is,

Theorem 16.10. *If $f : (a, b) \to \mathbb{R}$ is differentiable, then f' is Darboux continuous. That is, if $f'(x_1) < \alpha < f'(x_2)$, then $f'(\theta) = \alpha$ for some α between x_1 and x_2.*

Proof. Fix a positive $h > 0$. Then let $S(x)$ be the slope of the secant joining $(x, f(x))$ and $(x + h, f(x + h))$.

For some sufficiently small $h > 0$, we have $S(x_1) < \alpha < S(x_2)$. Then, by IVT, we have that some $S(\theta) = \alpha$. Now we apply MVT to the interval $(\theta, \theta + h)$ and we are done! □

17 October 24, 2013

We discuss some more properties of the derivative.

Recall that $f : (a, b) \to \mathbb{R}$ is differentiable if $\forall x \in (a, b)$, $f'(x)$ exists.

17.1 Unpleasant Examples

$f'(x)$ is always Darboux continuous; that is, the IVT is true for it too. However, f differentiable does not imply $f'(x)$ continuous. In particular, Darboux continuity does not imply continuity.

The function $f : (0, 1] \to \mathbb{R}$ by $x \mapsto \sin\frac{1}{x}$ is not continuous at $x = 0$. The function $x \mapsto x \sin\frac{1}{x}$ is, but not differentiable. On the other hand,

$$f(x) = \begin{cases} x^2 \sin\left(\frac{1}{x}\right) & \text{if } 0 < x \le 1 \\ 0 & \text{if } x = 0. \end{cases}$$

Is f differentiable at $x = 0$? Yes. Notice that

$$\frac{f(t) - f(0)}{t - 0} = \frac{f(t)}{t} = t \sin\frac{1}{t} \to 0$$

since sin is bounded. So, $f'(0)$ exists and is equal to zero. Furthermore,

$$f'(x) = 2x \sin\frac{1}{x} + x^2 \left(-\frac{1}{x^2} \cos\frac{1}{x}\right) = 2x \sin\frac{1}{x} - \cos\frac{1}{x}.$$

Accordingly we discover $f'(x)$ is not continuous at $x = 0$, as $\lim_{x\to 0} f'(x)$ does not even exist.

In fact, we can create even worse examples. The function defined by $f(x) = x^{\frac{3}{2}} \sin\frac{1}{x}$ $\forall x > 0$ and $f(0) = 0$ is even more pathological.

We can create two unpleasant points instead of one. Just consider $g : (0, 1) \to \mathbb{R}$ by

$$x \mapsto x^{1.5} \sin\frac{1}{x} \cdot (1 - x)^{1.5} \sin\frac{1}{1 - x}.$$

Now we take the Cantor set and paste copies of the above function into the gaps in the middle thirds Cantor set. By shrinking the amplitudes of g according to the size of the interval we're sticking it in, we obtain a function G such that

- $G'(x)$ exists everywhere.
- $G'(x)$ is discontinuous at uncountably many points.

This gets worse – we can place G inside the gaps rather than just g. We can eventually get a dense uncountable set of discontinuity in this manner.

Remarkably, you cannot get every point discontinuous. Darn.

Remark 17.1. In complex analysis, there is almost no pathology, unlike in the reals.

17.2 Higher Derivatives

Let $f : (a, b) \to \mathbb{R}$ be differentiable. It's quite possible that $f'(x)$ is differentiable.

Definition 17.2. Then $f''(x)$ is defined to be the $(f')'(x)$.If $f''(x)$ exists, we say f is *second order differentiable.*

Definition 17.3. Analogously, we can say a function is *third order differentiable*, and so on. When f is rth order differentiable, and we write $f^{(r)}(x)$ to denote this derivative.

With the notation above, $f^{(0)} = f$, $f^{(1)} = f'$, $f^{(2)} = f''$ and so on.

Definition 17.4. If f is rth-order differentiable for all r, we say f is *smooth*.

Let us see if we can get a multi-derivative chain rule. We already know

$$(g \circ f)' = (g' \circ f) \cdot f'.$$

Then the derivative of that is

$$\begin{aligned} \big((g' \circ f) \cdot f'\big)' &= \big((g' \circ f)' \cdot f'\big) + \big((g' \circ f) \cdot f''\big) \\ &= \big(\big(g'' \circ f\big) \cdot f'\big) \cdot f' + \big(g' \circ f\big) \cdot f'' \\ &= (g'' \circ f) \cdot f'^2 + (g' \circ f) \cdot f'' \end{aligned}$$

It is not easy to get a general formula. Sorry.

17.3 Nicer Functions

Definition 17.5. f is continuously differentiable if $f'(x)$ exists and is continuous as a function of x.

Definition 17.6. The collection of all such functions is denoted C^1, read "see-won" functions. We can also put $C^1((a,b),\mathbb{R})$ to be clear about the domain and range.

Definition 17.7. In general, C^r is the set of functions that $f^{(r)}$ exists and is continuous.

In particular, C^0 is the set of continuous functions, while C^∞ is the set of smooth functions. Note that continuity of $f^{(r)}$ implies continuity of $f^{(r-1)}$. Hence,

$$C^0 \supset C^1 \supset C^2 \supset \dots$$

and $C^\infty = \bigcap_{r \ge 0} C^r$.

We can do even better.

Definition 17.8. Let C^ω denote the set of *analytic functions*.

An analytic function, loosely, is a function which can be expressed as a power series. More formally,

Definition 17.9. A function $f : (a,b) \to \mathbb{R}$ is analytic if for each $x \in (a,b)$ there exists a power series

$$\sum_{r=0}^{\infty} a_r h^r \quad (a_r) \in \mathbb{R}$$

and a $\delta > 0$ such that if $|h| < \delta$, then $f(x+h) = \sum_{r=0}^{\infty} a_r h^r$.

You can show the following theorem, to be proved later.

Theorem 17.10. *In a power series as above, $a_r = \frac{1}{r!} f^{(r)}(x)$. In particular, (a_r) depends only on f.*

17.4 The Bump Function

It turns out C^ω is a strict subset of C^∞! This should come as surprising – smoothness seems like it should be good enough.

Here is the standard example. Define

$$\mathrm{e}(x) = \begin{cases} e^{-1/x} & \text{if } x > 0 \\ 0 & \text{if } x \le 0. \end{cases}.$$

Let us first check that this is not C^ω. Take for granted it is C^∞. Moreover, if it did, then at $x = 0$ we can find a series such that

$$f(h) = \sum_{r=0}^{\infty} a_r h^r$$

for all sufficiently small h. Moreover, $a_r = \frac{1}{r!} f^{(r)}(0)$.

Now it's obvious that the derivatives have to be zero! Just look at the left of the function – it has a straight tail. Now $a_r = 0$ for all r, so $f(h) \equiv 0$; this is clearly absurd.

Let us now check that $\mathrm{e}(x)$ is C^∞. Verify that when $x > 0$,

$$\mathrm{e}'(x) = e^{-1/x} \frac{1}{x^2}$$

and

$$\mathrm{e}''(x) = e^{-1/x} \frac{1}{x^4} - e^{-1/x} \frac{2}{x^3}.$$

In general, $\mathrm{e}^{(r)}(x)$ when $x > 0$ is a finite sum of terms of the form $e^{-1/x} \frac{1}{x^n}$, where n is a positive integer.

We just want to show that this approaches 0 as $x \to 0$. We could try to let L'Hopital's rule.as

$$\lim_{x \to 0} \frac{e^{-1/x}}{x^n} = \lim_{x \to 0} \frac{e^{-1/x} \cdot \frac{1}{x^2}}{nx^{n-1}} = \lim_{x \to 0} \frac{e^{-1/x}}{nx^{n+1}}.$$

Oh, that's quite unfortunate. This is getting worse: the numerator is increasing. This won't go anywhere.

But let's instead set $y = \frac{1}{x}$. Then we wish to compute

$$\lim_{y \to \infty} \frac{e^{-y}}{\left(\frac{1}{y}\right)^n} = \lim_{y \to \infty} \frac{y^n}{e^y}.$$

The conclusion is now obvious: exponentials grow much faster. Anyways, you can verify this with L'Hopital's rule.

We are now basically done. You can easily extend this to show that e is indeed in C^∞.

By extending this construction, you can eventually get a function that is analytic nowhere but smooth everywhere.

Remark 17.11. Note that we have not actually shown that analytic functions are smooth. We will prove this.

17.5 Taylor's Theorem

Let $f : (a, b) \to \mathbb{R}$ be a function differentiable at $x \in (a, b)$. We wish to approximate f near x by a polynomial. That is, we would want

$$f(x+h) = P(h) + R(h)$$

where $P(h)$ is a polynomial and $R(h)$ is a "small" remainder.

How do you do this! Natural.

$$P(h) = f(x) + h \cdot f'(x) + \frac{1}{2}h^2 f''(x) + \cdots + \frac{h^r}{r!} f^{(r)}(x).$$

Here $h > 0$. So what properties will $R(h)$ have if we put in this value of $P(h)$?

Definition 17.12. We say that $P(h)$ is the *Taylor polynomial* for f at x.

You should view x as fixed.

Theorem 17.13. *Fix x and $f : (a, b) \to \mathbb{R}$, where f is rth order differentiable at x. Define $R(h) = f(x + h) - P(h)$. Then $\frac{R(h)}{h^r} \to 0$ as $h \to 0$ if and only if $P(h)$ is the Taylor polynomial.*

Note that this suggest $R(h)$ is *really small.* $h^r \to 0$ is very fast for large r. We say that $R(h)$ is *rth order flat.*

Let us prove that $R(h)$ is flat when P is the Taylor polynomial.

Proof. Clearly $R(0) = 0$. Furthermore, by MVT, we know $R(h) - R(0) = R'(\theta_1) \cdot h$ for some $\theta_1 \in (0, h)$.

Now we can easily check $R'(0) = 0$ by construction. Then by the same logic,

$$R'(\theta_1) - R'(0) = R''(\theta_2) \cdot \theta_1$$

for some θ_2 with $0 < \theta_2 < \theta_1$.

By continuing the chain $R(h) = R'(\theta_1)h = R''(\theta_2)\theta_1 h = \dots$ we derive that

$$R(h) = \left(R^{(r-1)}(\theta_{r-1}) - R^{r-1}(0)\right) h\theta_1\theta_2 \dots \theta_{r-2}.$$

But $h\theta_1\theta_2 \dots \theta_{r-2} \le h^{r-1}$. Follow through. $\square$

18 October 29, 2013

Integration.

18.1 Riemann Sums

Suppose we wish to integrate a function $f : [a, b] \to \mathbb{R}$. Define a partition P of real numbers

$$a = x_0 < x_1 < \cdots < x_n = b.$$

Then, define $T = \{t_1, \ldots, t_n\}$, where $t_i \in [x_{i-1}, x_i]$ for each i.

Definition 18.1. The *Riemann sum* is defined by

$$R(f, P, T) = \sum_{i=1}^{m} f(t_i)\Delta x_i$$

where $\Delta x_i = x_i - x_{i-1}$.

Definition 18.2. The *mesh*, or *norm* of a partition, is $\max 1 \le i \le n \Delta x_i$. When used with the second name, we sometimes denote this by $\|P\|$.

Definition 18.3. We say f is *Riemann integrable* if and only if there exists an $I \in \mathbb{R}$ with the following property: for each $\varepsilon > 0$ there exists a $\delta > 0$ such that for all partitions P and sample T with the mesh of P less than Δ, we have the inequality

$$|R(f, P, T) - I| < \varepsilon.$$

We then write $I = \int_a^b f(x)\, dx$.

We will let $\mathcal{R}$ denote the set of Riemann integrable functions (on some interval $[a, b]$.)

Proposition 18.4. *If $f \in \mathcal{R}$, then f is bounded.*

Proof. Suppose not. Then $\exists I \in \mathbb{R}, \delta.0$ such that $\|P\| < \delta \Rightarrow |R(f, P, T) - I| < 2013$.

Fix P. Now there exists a k such that $f(t)$ is unbounded as $x_{k-1} \le t \le x_k$; after all f is unbounded. We will now construct a bad T. Choose t_i arbitrarily for all $i \ne k$. Then pick t_k to be really really big. Now $R(f, P, T)$ is large, and in particular, greater than $I + 2013$. □

Note that improper integrals are something different entirely! For example, define

$$f(x) = \begin{cases} x^{-\frac{1}{2013}} & x > 0 \\ 0 & x = 0. \end{cases}$$

Although $\int_0^1 f(x)\, dx$ can be evaluated as an improper integral, we still consider $f : [0, 1] \to \mathbb{R}$ to be not Riemann integrable.

Proposition 18.5. *The map $\mathcal{R} \to \mathbb{R}$ by $f \mapsto \int_a^b f(x)\, dx$ is bilinear. Furthermore, if $f, g \in \mathcal{R}$ and $f(x) \le g(x)$ for all x, then $\int f \le \int g$. Finally, if $f(x) \equiv c$ then $\int_a^b f = c(b - a)$.*

Proof. This is obvious. □

18.2 Darboux Integrability

Consider $f : [a, b] \to [-M, M]$, and fix a partition P. Define

$$L(f, P) = \sum_{i=1}^{n} m_i \delta x_i$$

and

$$U(f, P) = \sum_{i=1}^{n} M_i \delta x_i$$

where m_i and M_i are the infimum and supremum of $\{f(t) : x_{i-1} \le t \le x_i\}$.

Definition 18.6. We say P' *refines* P if $P' \supset P$.

It is obvious that

$$L(f, P) \le L(f, P') \le U(f, P') \le U(f, P)$$

in this case. (Check it.) This is the refinement principle.

Proposition 18.7. *For* any *partitions* P_1 *and* P_2*, we have* $L(f, P_1) \le U(f, P_2)$*.*

Proof. $L(f, P_1) \le L(f, P_1 \cup P_2) \le U(f, P_1 \cup P_2) \le U(f, P_2)$ and we're done. We refer to $P_1 \cup P_2$ as the *common refinement.* □

Definition 18.8. Over all partitions P,

$$\underline{I} \stackrel{\text{def}}{=} \sup_P L(f, P)$$

is the *lower Darboux interval*, and

$$\overline{I} \stackrel{\text{def}}{=} \inf_P U(f, P)$$

is the *upper Darboux interval.* If $\underline{I} = \overline{I}$ then f is *Darboux integrable.*

Cool. Here is the nice theorem.

Theorem 18.9. *The following are equivalent for* $f : [a, b] \to [-M, M]$*.*

(a) $\underline{I} = \overline{I}$*. (*$\underline{I} \le \overline{I}$ *is always true.)*

(b) $\forall \varepsilon > 0 \exists P$ *such that* $U(f, P) - L(f, P) < \varepsilon$*.*

(c) f *is Riemann integrable.*

In particular, Darboux integrability and Riemann integrability are equivalent.

Proof. First, let us prove (a) implies (b). By (a) we can find P_1 and P_2 such that $\overline{I} - L(f, P_1) < \frac{1}{2}\varepsilon$ and $U(f, P_2) - \overline{I} < \frac{1}{2}\varepsilon$. Now let $P = P_1 \cup P_2$. We find

$$L(f, P_1) \le L(f, P) \le U(f, P) \le U(f, P_2)$$

but $U(f, P_2) - L(f, P_1) < \varepsilon$ so we're done. Also (b) implies (a) is obvious.

The most involved part is showing that (a) or (b) and (c) are equivalent. First, we show that (c) implies (b); let I be the Riemann integrable. Evidently $\forall \varepsilon > 0$, there is

$\delta > 0$ such that if P is a partition with $\|P\| < \delta$, then $|R(f, P, T) - I| < \varepsilon$. Take any such partition P; we can find T such that

$$|R(f, P, T) - L(f, P)| < \varepsilon$$

because $L(f, P) = \sum_{i=1}^{m} m_i \delta x_i$, so we can choose the points T to be arbitrarily close to the m_i. Similarly, we can choose T' with$|R(f, P, T') - U(f, P)| < \varepsilon$. But because of (c), $R(f, P, T)$ and $R(f, P, T')$ differ by at most 2ε and so $U(f, P)$ and $L(f, P)$ differ by at most 4ε. Adjust accordingly.

Finally we will prove (a) implies (c). Set $I = \underline{I} = \overline{I}$. Given $\varepsilon > 0$, we wish to find $\delta > 0$ such that $\|P\| < \delta$ implies that $|R(f, P, T) - I| < \varepsilon$.

We already know we can find partition P such that $L(f, P_1) - \underline{I} < 0.01\varepsilon$ and $U(f, P_1) - \overline{I} < 0.01\varepsilon$. (Take two partitions and their common refinement.) So

$$U(f, P_1) - L(f, P_1) < 0.02\varepsilon.$$

Take $\delta \leq \frac{\varepsilon}{16 n_1 M}$. Consider a partition P such that $\|P\| < \delta$. Define another common refinement

$$P^* = P_1 \cup P.$$

By refinement, $U(f, P^*) - L(f, P^*) < 0.02\varepsilon$.

Anyways, we wish to compare $U(f, P)$ and $U(f, P^*)$. We look at "bad" intervals; that is, of P^* which have a endpoint of P_1 dropped inside it. Because P has a LOT of points, very few intervals (in fact, at most $2n_1$), have differences, two for each endpoint of the irritating P_1. The difference is bounded by $2M$ because that is the bound of the function.

...okay I am not copying the rest of this down, this is just ε-bashing. $\square$

19 October 31, 2013

Last time we showed that Darboux integrability and Riemann integrability are equivalent. However, this begs the question of whether there is a nice way to determine whether a function is integrable. For example, it is totally nonobvious that the product of two integrable functions is integrable.

It turns out a good criteria does exist!

Theorem 19.1 (Riemann-Lebesgue)**.** *A function $f : [a, b] \to \mathbb{R}$ is Riemann integrable if and only if f is bounded and its set of discontinuities is a zero set.*

19.1 Discussion of Zero Sets

Recall that $Z \subset \mathbb{R}$ is a zero set if for each $\varepsilon > 0$ there exists a countable covering of Z by open intervals with total length less than ε.

1. Finite sets
2. Subsets of zero sets
3. Countable sets
4. Countable unions of zero sets
5. The middle-thirds Cantor set

These are mostly trivial. Actually, maybe it's not so obvious how to cover countable sets, so here is how. If $S = \{x_1, x_2, \dots\}$ is countable, then cover each x_n by an interval of length $\frac{\varepsilon}{2^{n+2013}}$. This is a standard "trick".

The proof for countable unions of zero sets is basically the same – just consider zero sets $X_1, X_2, \dots$ instead of points.

We've already seen that the Cantor set is a zero set. Interestingly, the fat Cantor set is not a zero set, despite being homeomorphic to C.

Definition 19.2. We say that *almost every* $x \in \mathbb{R}$ has some property if the set of counterexamples is a zero set.

19.2 Examples of Riemann Integrable Functions (and non-integrable ones)

Monotone functions may be continuous at only countably many points. Project the "jumps" onto the y-axis in the graph; we get a bunch of disjoint open intervals, and in particular, only countably many such interval can exist. Hence, all monotone functions are integrable.

Definition 19.3. For a set S, define the characteristic function χ_S by

$$\chi_S(x) = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{if } x \notin S. \end{cases}$$

Some books also use the notation $\not\Vdash_S$.

One can easily show that $\chi_{\mathbb{Q}}$ is discontinuous everywhere. On the other hand, χ_C has discontinuous at precisely the Cantor set.

On the other hand, the rational ruler function defined by

$$f(x) = \begin{cases} 1/q & x = p/q \text{ in lowest terms} \\ 0 & \text{otherwise.} \end{cases}$$

It turns out that f is integrable because its discontinuity set is $\mathbb{Q}$!

19.3 Proof of the Riemann-Lebesgue Theorem

Proof. First, we show that if f is Riemann integrable, then it is bounded and its discontinuity set D is a zero set.

Let $\varepsilon > 0$ be given. Define

$$\operatorname{osc} x f = \limsup_{t\to x} f(t) - \liminf_{t\to x} f(t)$$

to be the oscillation, and for each $\kappa > 0$ let

$$D_\kappa \stackrel{\text{def}}{=} \{x \mid \operatorname{osc} x(f) > \kappa\}.$$

Then, $D = \bigcup_{k=1}^\infty D_{1/k}$. Therefore, it suffices to prove that each D_κ is a zero set.

We know there exists a partition P given by $a = x_0 < x_1 < \cdots < x_n = b$ such that $U(f,P) - L(f,P) < \frac{\kappa}{2}\varepsilon$, where M is the supremum. We define a disjoint union of $[a,b] = G \sqcup B \sqcup E$, the "good", the "bad", and the "endpoints". Here E is the set of endpoints of the intervals of P. We will let G consist of those intervals (x_i, x_{i+1}) containing no points of D_κ, and B those that do. Evidently

$$\begin{aligned}
\frac{\kappa}{2}\varepsilon &> U(f,P) - L(f,P) \\
&= \sum_{i=1}^{m} (M_i - m_i)\Delta x_i \\
&= \sum_{i \text{ good}} \bullet + \sum_{i \text{ bad}} \bullet \\
&\ge \sum_{i \text{ bad}} \bullet \\
&> \sum_{i \text{ bad}} \kappa \Delta x_i
\end{aligned}$$

because each of the bad intervals has a point with oscillation greater than κ. Then

$$\frac{1}{2}\varepsilon > \sum_{\text{bad}} \Delta x_i$$

So D_κ, which mostly consists of these bad intervals, can be covered with length less than $\frac{1}{2}\varepsilon$. The only points in D_κ that might be missed our endpoints, so we just cover all the endpoints too.

Now for the converse. Assume $f : [a,b] \to \mathbb{R}$ is bounded by M and has D a zero set. For each $\kappa > 0$, we find D_κ is also a zero set. Fix $\varepsilon > 0$. We want to find a partition such that $U(f,P) - L(f,P) < \varepsilon$.

Remark that for all $x \notin D_\kappa$, there is an open interval I_x containing x such that $\sup\{f(t) : t \in I_x\} - \inf\{f(t) : t \in\} < \kappa$. Also, we can cover D_κ by open intervals J_j with total length less than $\frac{1}{4M}\varepsilon$. Let U be the combined covering.

Let λ be a Lebesgue number for the covering U on the compact $[a,b]$. Hence every set of diameter less than λ is contained in either an I_x or J_j. Now just choose any P with mesh less than λ. We claim that we win. Intervals contained in J_j have total length less than $\frac{\varepsilon}{4M}$ and maximum "width" $2M$, so the sum here is less than $\frac{1}{2}\varepsilon$. The sum of everything else is at most $\kappa(b-a)$. So we just need to take $\kappa < \frac{\varepsilon}{2(b-a)}$ and life is good. $\square$

The idea of good versus bad intervals repeats itself.

19.4 Consequences of the Riemann-Lebesgue Theorem

Corollary 19.4. *Continuous implies Riemann integrable.*

Proof. Bounded because the domain is closed, and the discontinuity set is empty.

You can prove this directly. Compactness implies that a continuous $f : [a, b] \to \mathbb{R}$ is uniformly continuous: for each $\varepsilon > 0$ there exists a $\delta > 0$ such that $|x - x'| < \delta \Rightarrow |f(x) - f(x')| < \varepsilon$. So we just pick $\varepsilon = \frac{\varepsilon'}{b-a}$ for any $\varepsilon' > 0$, and take P with mesh less than δ. □

Corollary 19.5. *Monotone implies Riemann integrable.*

Corollary 19.6. $f, g \in \mathcal{R} \Rightarrow f \cdot g \in \mathcal{R}$.

Proof. $D(f) \cup D(g)$ contains $D(fg)$. □

Corollary 19.7. *Let $f : [a, b] \to [-M, M]$ is Riemann integrable, and $h : [-M, M] \to \mathbb{R}$ is continuous, then $h \circ f$ is Riemann integrable.*

Proof. $D(h \circ f) \subseteq D(f)$.because h is continuous. Furthermore, h is bounded because it is continuous. □

20 November 5, 2013

Recall the Riemann-Lebesgue Theorem. It is really useful.

20.1 Corollaries of the Riemann-Lebesgue Theorem

1. f continuous implies f is Riemann integrable.
2. Let $S \subset [a, b]$. Then χ_S, the characteristic function of S, has discontinuity equal to the boundary of S, ∂S. Hence χ_S is Riemann integrable only when ∂S is countable.
3. If each discontinuity of f is a jump discontinuity, and f is bounded, then f is Riemann integrable.[7] In particular, monotone functions are Riemann integrable.
4. The product of Riemann integrable functions is Riemann integrable. So is their ratio if the second function is bounded away from zero.
5. If $f \in \mathcal{R}$ is bounded by M and $h : [-M, M] \to \mathbb{R}$ is continuous. Then $h \circ f$ is Riemann integrable.
6. $|f|$ is Riemann integrable for any $f \in \mathcal{R}$. (Just take $h : x \mapsto |x|$ in the above.)
7. Let $a < c < b$ and $f \in \mathcal{R}$. Then $f|_{[a,c]}$ (that is, f restricted to $[a, c]$) is Riemann integrable on $[a, c]$, as is $f|_{[c,b]}$, and $\int_a^b f(t)\, dt = \int_a^c f(t)\, dt + \int_c^b f(t)\, dt$. One "clean" proof is $f = \chi_{[a,c]}f + \chi_{(c,b]}f$ and $\chi_{[a,c]} + \chi_{(c,b]} \equiv 1$.
8. Suppose $f(x) \ge 0$ for each $x \in [a, b]$ and $f \in \mathcal{R}$. If $\int_a^b f(t)dt = 0$, then f is zero almost everywhere. Indeed, just claim that $f(x) = 0$ whenever f is continuous at x. (This is not completely trivial, consider χ_C.)
9. Suppose $h : [c, d] \to [a, b]$ and $f : [a, b] \to \mathbb{R}$, where $f \in \mathcal{R}$. Suppose further that h is a homeomorphism and h^{-1} satisfies the Lipschitz condition[8]; i.e. there exists L such that $\left|h^{-1}(x) - h^{-1}(y)\right| < L(|x - y|)$. It's basically equivalent to verify that $D(f \circ h) = h^{-1}(D(f))$.

20.2 Diffeomorphisms

Definition 20.1. A *diffeomorphism* $h : [a, b] \to [c, d]$ is a homeomorphism if it is a homeomorphism and of class C^1 and moreover $h^{-1} : [c, d] \to [a, b]$ is also of class C^1.

The fact that h^{-1} is C^1 is actually necessary. Consider the homeomorphism $x \mapsto x^3$. Its inverse $y \mapsto y^{\frac{1}{3}}$ is NOT differentiable at zero.

Proposition 20.2. *If $[c, d] \xrightarrow{h} [a, b] \xrightarrow{f} \mathbb{R}$ has h a diffeomorphism and $f \in \mathcal{R}$, then $f \circ h$ is Riemann integrable.*

Proof. It suffices to show h^{-1} satisfies the Lipschitz condition. We know $(h^{-1})'$ is bounded by some constant L for all x. Now consider x, y. By the Mean Value theorem there exists θ between x and y with

$$\left|h^{-1}(x) - h^{-1}(y)\right| = \left|h^{-1}(\theta)(x - y)\right| \le L\,|x - y|\,.$$

□

[7] Here's a proof. Consider the set of jumps of size ≥ 1. We claim there are only finitely many such jumps. Otherwise, by compactness of $[a, b]$, then there exists $x_0 \in [a, b]$ at which the jump points accumulate. This is clearly bad.

[8] Meaning points are not stretched out greatly

This becomes false when h is a homeomorphism instead of a diffeomorphism. Let $h : [0,1] \to [0,1]$ taking the Cantor set C to the fat Cantor set F. Then $\chi_F \equiv \chi_C \circ h^{-1}$. But $\partial C = C$ is a zero set while $\partial F = F$ is not. Good game!

Remark 20.3. Homeomorphsims are sometimes called "changes of variable". The counterexample above shows that some changes of variables are unacceptable for preserving Riemann integrability.

20.3 Antiderivatives

Definition 20.4. A function $G : [a,b] \to \mathbb{R}$ is called an *antiderivative* of $g : [a,b] \to \mathbb{R}$ if for each $x \in [a,b]$, $G'(x)$ exists and equals $g(x)$.

Proposition 20.5. *If g has a jump discontinuity, then g does NOT have an antiderivative.*

Proof. This follows from the fact that the derivatives have the intermediate value property. □

Example 20.6. The fairly nice function $f : [0,2] \to \mathbb{R}$ by $f = \chi_{[0,1]} + \chi_{(1,2]}$ has no antiderivative. Meanwhile, the much more unfortunate function

$$g(x) = \begin{cases} \sin \frac{1}{x} & x > 0 \\ 0 & x \le 0. \end{cases}$$

does.

How do we find this antiderivative? We basically want $G(x) = \int_0^x g(t)\, dt$ but it's not obvious how to do this. The key is that for any $\alpha > 0$ we have

$$t^2 \cos(1/t)|_\alpha^x = \int_\alpha^x \sin \frac{1}{t}\, dt + \int_\alpha^x 2t \cos\left(\frac{1}{t}\right)\, dt.$$

This actually holds for $\alpha = 0$ as well! Just remark that $t^2 \cos(\frac{1}{t}) \to 0$ and $t \cos(\frac{1}{t}) \to 0$ as $t \to 0$. Hence, we derive that

$$\int_0^x \sin\left(\frac{1}{t}\right)\, dt = x^2 \cos \frac{1}{x} - \int_0^x 2t \cos\left(\frac{1}{t^2}\right)\, dt$$

and the RHS is nice.

It remains to check that $G'(0) = 0$ (and exists). Compute

$$\frac{G(h)}{h} = \frac{1}{h} \int_0^h \sin(\frac{1}{t})\, dt = \frac{1}{h}\left[h^2 \cos \frac{1}{h} - \int_0^h 2t \cos \frac{1}{t^2}\, dt\right].$$

The first term tends to zero, so we just need $\frac{1}{h} \int_0^h 2t \cos(\frac{1}{t^2})\, dt$ tends to zero. Now just bound the cosine by one; we conclude

$$\left| \frac{1}{h} \int_0^h 2t \cos\left(\frac{1}{t^2}\right) \right| \le \frac{1}{h} \int_0^h 2t\, dt = h \to 0.$$

Game, set and match.

21 November 7, 2013

We're going to ignore the proofs of the classic calculus theorems. There's other stuff to do. We'll just mention the devil's staircase.

21.1 The Devil's Staircase and Ski Slope

We will construct $f : [0,1] \to [0,1]$ such that $f'(x)$ exists and is zero almost everywhere, yet f is nonconstant.

The construction is based on the Cantor set. For each discarded interval I in $[0,1] \setminus C$, define $f(I)$ to be the midpoint of $\frac{1}{2}$. So, for example, $f(t) = \frac{1}{2}$ for all $t \in \left[\frac{1}{3}, \frac{2}{3}\right]$, and $f(t) = \frac{1}{6}$ for all $t \in \left[\frac{1}{9}, \frac{2}{9}\right]$.

Hence for all $x \notin C$, $f'(x) = 0$. One can also verify that f is continuous. This is the *devil's staircase.*

This can be made worse. The *devil's ski slope* is strictly increasing but has the same properties as the devil's staircase.

21.2 The Improper Integral

We do little more than define this.

Definition 21.1. If $f : [a, \infty) \to \mathbb{R}$ is Riemann integrable when restricted to each $[a, b]$ and the quantity

$$\lim_{b \to \infty} \int_a^b f(t)\, dt$$

exists, then we say that the *improper integral* $\int_a^\infty f(t)\, dt$ exists and equals said limit.

Example 21.2. Consider $f(t) = \frac{1}{t^2}$. We have

$$\int_1^b \frac{1}{t^2}\, dt = 1 - \frac{1}{b}$$

so the improper integral $\int_1^\infty = 1$.

Exercise 21.3. Construct an unbounded function which has an unbounded integral.

Definition 21.4. $\int_{-\infty}^\infty f(t)\, dt$ exists if $\int_{-\infty}^0 f(t)\, dt$ and $\int_0^\infty f(t)\, dt$ both exists, independently.

Remark 21.5. It's not enough for $\lim_{r\to\infty} \int_{-r}^r f(t)\, dt$ to exist. This is strictly weaker.

21.3 Series

Given a sequence of real numbers (a_n), we wish to determine when $\sum_{k=1}^\infty a_k$.

Definition 21.6. Define $A_n = \sum_{k=1}^n a_k$ as the nth partial sum. Then $\sum_{k=1}^\infty$ *converges* if and only if $A = \lim_{n\to\infty} A_n$ exists. We then say it converges to A. Otherwise, it *diverges.*

Because we are dealing with real numbers, we have that $\sum a_k$ converges if and only if it is Cauchy: that is, for all $\varepsilon > 0$, there exists N such that for all $n > m \ge N$ such that

$$\varepsilon > |A_n - A_m| = \left| \sum_{k=m+1}^n a_k \right|.$$

Remark 21.7. A convergent series must have terms tending to zero. The converse is EXTREMELY false.

Definition 21.8. We say $\sum a_k$ *converges absolutely* if $\sum |a_k|$ converges. Otherwise it *converges conditionally.*

Example 21.9. $\sum_{k=1}^{\infty} \frac{1}{k}$ diverges, but $\sum_{k=1}^{\infty} \frac{(-1)^k}{k}$ does. Hence the latter is conditionally convergent.

21.4 Examples of Series

One example is the harmonic series.

The *geometric series*: for any $-1 < \lambda < 1$ the series

$$\sum_{k=0}^{\infty} \lambda^k$$

converges to $(1-\lambda)^{-1}$. This is obvious: the mth partial sum is precisely $\frac{1-\lambda^{m+1}}{1-\lambda}$ tends to $\frac{1}{1-\lambda}$.

On the other hand, if $|\lambda| > 1$ then the geometric series do not tend to zero, and hence the sequence diverges.

A very general series is the alternating series. If $a_n \downarrow 0$ (i.e. a_n decreases monotonically to zero) as $n \to \infty$, then the *alternating series* $\sum_{n=1}^{\infty} (-1)^n a_n$ converges. The series appears as $-a_1 + (a_2 - a_3) + (a_4 - a_5) + \dots$. The numbers $a_2 - a_3$, $a_4 - a_5$, and so on are all nonnegative.

21.5 Tests

Most tests are consequences of the comparison test.

Proposition 21.10 (Comparison Test)**.** *If $\sum b_k$ converges and if $\sum a_k$ obeys $|a_k| \le b_k$ then the a_k-series converges absolutely.*

Definition 21.11. We say that (b_n) dominates (a_n).

Remark 21.12. Conditionally convergent series are strange. Here is an example. We can rearrange the indices with a bijection $\beta : \mathbb{N} \to \mathbb{N}$. If $\sum a_n$ converges absolutely, then $(a_{\beta(n)})$ converges absolutely to the same thing. But if $\sum a_n$ is conditionally convergent, then $(a_{\beta(n)})$ can converge to anything! And we mean anything: for any $\sum a_n$ and $\alpha \in \mathbb{R}$ there exists a bijection β_α forcing $\sum a_{\beta(n)}$ to converge to α.

You can even get conditionally convergent series with the following property: if $A_1, A_2, \dots$ is the sequence of partial sums, some suitable bijection β will cause subsequences of the (A_n) to converge to every point of any closed interval.

21.6 That Homework Exercise that was Nontrivial

Proposition 21.13. *If $\sum a_n$ converges and $b_n \uparrow b$, then $\sum a_n b_n$ converges.*

One way to see this is

$$\begin{aligned}
&a_m b_m + a_{m+1} b_m + \dots + a_{n-1} b_m + a_n b_m \\
&a_{m+1}(b_{m+1} - b_m) + \dots + a_{n-1}(b_{m+1} - b_m) + a_n(b_{m+1} - b_m) \\
&\qquad\qquad \ddots + \vdots \\
&\qquad\qquad\quad + a_{n-1}(b_{n-1} - b_{n-2}) + a_n(b_{n-1} - b_{n-2}) \\
&\qquad\qquad\quad + a_n(b_n - b_{n-1})
\end{aligned}$$

21.7 Exponential Growth Rate

Definition 21.14. For a sequence (a_n), the *geometric growth rate* is defined as $\rho = \limsup_{k\to\infty} |a_k|^{1/k}$.

Proposition 21.15 (Root Test)**.** *If $\rho < 1$ then $\sum a_k$ converges absolutely; if $\rho > 1$ then $\sum a_k$ diverges. The test is inconclusive if $\rho = 1$.*

Proof. If $\limsup_{k\to\infty}|a_k|^{1/k} = \rho$, then for k large we have $|a_k| < \rho^k$ and comparison to the geometric series works. If $\rho > 1$, then there are infinitely many terms with magnitude greater than one, so it cannot possibly converge.

The p-series is an example when the root test is useless; $\rho = 1$ for every p-series. $\square$

Musing: what if you knew $\limsup_{k\to\infty}|a_k|^{1/k}$ approached at a certain rate? Would that help?

22 November 12, 2013

We now begin the study of function spaces. Instead of adding numbers, we will be adding functions.

22.1 Sequences of Functions and Uniform Convergence

First, let us consider a sequence of functions (f_n), and another function f, each defined from $[a,b] \to \mathbb{R}$.

Definition 22.1. We say f_n *converges pointwise* to f if $\forall x \in [a,b]$, $f_n(x) \to f(x)$ as $n \to \infty$. We write $f_n \to f$ or $\lim_{n\to\infty} f_n = f$.

Definition 22.2. We say f_n *converges uniformly* if $\forall \varepsilon > 0$, $\exists N \in \mathbb{N}$ such that $\forall n \ge N$ and $\forall x \in [a,b]$ we have $|f_n(x) - f(x)| < \varepsilon$. We write $f_n \rightrightarrows f$.

The "correct" way to view this is to construct a "tube" of width ε around each function f.

Example 22.3. Let $f_n(x) = x^n$ for each n, where $f_n : [0,1] \to \mathbb{R}$. Let $f(x) = 0$ for each $0 \le x < 1$ and $f(1) = 1$. It's easy to see that f_n converges to f pointwise.
However, this is far from uniform convergence.

Uniform convergence is "good". For example.

Theorem 22.4. *Suppose $f_n \rightrightarrows f$ and $x_0 \in [a,b]$. If f_n is continuous at x_0 for infinitely many n, then so is f.*

Remark 22.5. This is false if $f_n \rightrightarrows f$ is replaced by $f_n \to f$; just consider the example from before.

Proof. Let $\varepsilon > 0$ be given. We wish to show that there is a $\delta > 0$ such that $|x - x_0| < \delta \Rightarrow |f(x) - f(x_0)| < \varepsilon$.
By uniform convergence, there is some large N such that $|f_N(x) - f(x)| < \frac{1}{3}\varepsilon$ (actually for all $n \ge N$). We may also assume that f_N is continuous at x_0. Now by continuity of f_N there exists a $\delta > 0$ such that $|x - x_0| < \delta \Rightarrow |f_N(x) - f_N(x_0)| < \frac{1}{3}\varepsilon$.
In that case, for all x within δ of this x_0,

$$\begin{aligned}
|f(x) - f(x_0)| &\le |f(x) - f_N(x)| + |f_N(x) - f_N(x_0)| + |f_N(x_0) - f(x_0)| \\
&< \frac{1}{3}\varepsilon + \frac{1}{3}\varepsilon + \frac{1}{3}\varepsilon \\
&= \varepsilon
\end{aligned}$$

$\square$

Corollary 22.6. *If f_n is continuous at all $x \in [a,b]$ and $f_n \rightrightarrows f$, then f is continuous at all $x \in [a,b]$.*

Example 22.7. Construct the *growing steeple* as follows: define

$$f_n = \begin{cases} n^2 x & 0 \le x \le \frac{1}{n} \\ n - n^2\left(x - \frac{1}{n}\right) & \frac{1}{n} \le x \le \frac{2}{n} \\ 0 & \text{otherwise} \end{cases}.$$

22.2 Function Spaces

Definition 22.8. Let C_b denote the set of bounded functions from $[a,b] \to \mathbb{R}$. Equip it with a metric $d : C_b^2 \to \mathbb{R}$ by

$$d(f,g) = \sup \{ |f(x) - g(x)| : x \in [a,b] \} .$$

One can check that this distance is in fact a metric! We also have the following result.

Proposition 22.9. *$f_n \to f$ in C^b if and only if $f_n \rightrightarrows f$.*

Proof. Trivial. ☐

We may also define the norm of f by $\|f\| = d(f, 0)$, where 0 denotes the zero function. In fact, we have the following theorem.

Theorem 22.10. *C_b is a complete metric space.*

The intuition is that this follows from completeness of $\mathbb{R}$.

Proof. Let (f_n) be a Cauchy sequence. Evidently the sequence $(f_n(x))_{n \in \mathbb{N}}$ is Cauchy for each x, because $|f_n(x) - f_m(x)| \le d(f_n, f_m)$. Since this occurs in the reals, we deduce $\lim_{n\to\infty} f_n(x)$ exists for all x. Define $f(x)$ to be this limit.

We now have a target.

Remark that $f_n \to f$ with respect to d if and only if f_n converges uniformly to f. We wish to show that for each $\varepsilon > 0$ we can find an N with $n \ge N \Rightarrow |f_n(x) - f(x)| < \varepsilon$. Evidently there exists N such that for all $n, m \ge N$ and for all x, we have

$$|f_n(x) - f_m(x)| < \frac{1}{2}\varepsilon.$$

Now for all $x_0 \in [a,b]$ there exists an $m(x_0) \ge N$ such that $\left| f_{m(x_0)}(x_0) - f(x_0) \right| < \frac{1}{2}\varepsilon$. Then for each $n \ge N$ we have the inequality

$$|f_n(x) - f(x)| \le \left| f_n(x) - f_{m(x)}(x) \right| + \left| f_{m(x)}(x) - f(x) \right| < \varepsilon.$$

Finally we just have to show f is bounded. But this follows rather readily from uniform convergence now that we have $f_n \rightrightarrows f$. ☐

Notice how when applying the triangle inequality, we had to pick the m dependent on x. This is apparently a standard trick for pointwise convergence.

22.3 More properties of uniform continuity

Corollary 22.11. *C_0 is complete. So is $\mathcal{R}$.*

Proof. We claim they are closed subsets of the complete space C_b; in fact $C^0 \subset \mathcal{R} \subset C_b$. That C^0 is closed follows from our earlier theorem.

Hence, we only consider the second statement, showing that $\mathcal{R}$ is closed. Suppose $f_n \in \mathcal{R}$ and $f_n \rightrightarrows f$. We already know f is bounded. Now by our continuity theorem the discontinuity set of f is within

$$\bigcup_{n=1}^{\infty} D(f_n)$$

which is a zero set. Hence Riemann-Lebesgue implies the conclusion. ☐

Proposition 22.12. *If $f_n \rightrightarrows f$ and each f_n is Riemann integrable, then*

$$\int_a^b f_n(t)\ dt \to \int_a^b f(t)\ dt.$$

Proof. Just write the inequalities

$$\left| \int_a^b (f_n(t) - f(t))\ dt \right| \le \int_a^b |f_n(t) - f(t)|\ dt \le \int_a^b d(f_n, f)\ dt = (b-a)d(f_n, f) \to 0. \qquad \square$$

So the integrals work out nicely. Derivatives are less pleasant, as the following examples show.

Example 22.13. Suppose (f_n) is a sequence of functions in C^1 and $f_n \rightrightarrows f$. We would like f to approach f, but this is false. Define

$$f_n(x) = \sqrt{x^2 + \frac{1}{n}}.$$

We observe $f_n(x) \rightrightarrows f$, where $f(x) = |x|$. But the derivatives at $x = 0$ do not agree, because f does not even have a derivative at that point!

The following extra condition is necessary.

Theorem 22.14. *Suppose $f_n : [a,b] \to \mathbb{R}$ is continuously differentiable and $f_n \rightrightarrows f$. Suppose further $f_n' \rightrightarrows g$ for some function g. Then f is differentiable and $f' = g$.*

Proof. Uses the Fundamental Theorem of Calculus. We know that

$$f_n(x) = f_n(a) + \int_a^x + \int_a^x f_n'(t)\ dt.$$

Considering the behavior as $n \to \infty$, we obtain

$$f(x) = f(a) + \int_a^x g(t)\ dt.$$

We know $g(n)$ is continuous since f_n' are all continuous and converge uniformly to g. Then by the Fundamental Theorem of Calculus, we obtain $f'(x) = g(x)$ as desired. $\square$

One can actually strengthen this to "differentiable" instead of "continuously differentiable".

22.4 Series of Functions

Again we have $f_n : [a,b] \to \mathbb{R}$. We wish to consider partial sums

$$\sum_{k=0}^{n} f_k(x) = F_n(x).$$

Definition 22.15. The series of functions converges uniformly if and only if the sequence of partial sums (F_n) converges uniformly.

As for the reals, we may consider the Cauchy condition instead: for each $\varepsilon > 0$ there is a large N such that for all $n > m \ge N$ we have

$$\varepsilon > |F_n(x) - F_m(x)| = \left| \sum_{k=m+1}^{n} f_k(x) \right|$$

for all x.

Here is a nice fact.

Theorem 22.16 (Weierstrass M-Test)**.** *Let $\sum_{k=0}^{\infty} M_k$ be a convergent sequence of non-negative reals. If (f_n) is a sequence of functions with $\|f_k\| \le M_k$ for all k, then $\sum f_k$ converges uniformly.*

Proof. Just apply the Cauchy condition. □

22.5 Power Series

These are series of functions $\sum_{k=0}^{\infty} f_k(x)$ where $f_k(x) = c_k x^k$.

Definition 22.17. Define

$$R = \frac{1}{\limsup_{k\to\infty} |c_k|^{\frac{1}{k}}}.$$

This is called the *radius of convergence.*

Example 22.18. If $c_k = 1$ then $R = 1$. If $c_k = 2^k$ then $R = \frac{1}{2}$. If $c_k = k^k$ (very fast!) then $R = 0$. If c_k is very small, then $R = \infty$. And so on. In short, we can achieve all spectrums of R.

Theorem 22.19. *If $|x| < R$ then the series converges to a function $f : (-R, R) \to \mathbb{R}$. Furthermore, f is C^1, and converges uniformly on closed intervals. Moreover, f' is also given by a canonical power series*

$$\sum_{k=1}^{\infty} k c_k x^{k-1}.$$

Needless to say, f'' is also given by a power series of the same radius, and so on. In particular, if f is given by a power series, then f is smooth.

23 November 14, 2013

Today's topic is equicontinuity.

23.1 Equicontinuity

Intuitively, $y = x$ is "more continuous" than $\sin(1000x)$. Here is the definition.

Definition 23.1. A sequence of functions (f_n) from $[a, b]$ to $\mathbb{R}$ is *equicontinuous* if $\forall \varepsilon > 0$, there is a $\delta > 0$ such that for all n and $x, y \in [a, b]$ we have

$$|x - y| < \delta \Rightarrow |f_n(x) - f_n(y)| < \varepsilon.$$

The important thing here is that δ depends only on ε.

Definition 23.2. (f_n) is *pointwise equicontinuous* if for every $x_0 \in [a, b]$ and $\forall \varepsilon > 0$ there exists $\delta > 0$ such that if $x \in [a, b]$ and $|x - x_0| < \delta$, then

$$|f_n(x) - f_n(x_0)| < \varepsilon.$$

We will be dealing with the first definition.

Theorem 23.3. *Suppose $f_n : [a, b] \to \mathbb{R}$ has the property that each f_n is differentiable and there exists L such that $|f_n'(x)| \le L$. Then (f_n) is equicontinuous.*

Proof. Let $\varepsilon > 0$ be given. Choose $\delta < \frac{\varepsilon}{L}$. Now cite the Mean Value Theorem and win:

$$|f_n(x) - f_n(y)| = f'(\theta)\,|x - y| \le L\delta \le \varepsilon. \qquad \square$$

The converse is not true. The function given by

$$f(x) = \begin{cases} x^{3/2} \sin x^{-1} & x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

has unbounded derivative but is nonetheless continuous. On a compact, it is uniformly continuous. Now take $f_n \equiv f$ for all f.

23.2 The Azela-Ascoli Theorem

Theorem 23.4 (Arzela-Ascoli Theorem)**.** *If (f_n) is equicontinuous and uniformly bounded*[9] *then there is a uniformly convergent subsequence, or equivalently, a convergent in C^0.*

So this is essentially a compactness result.

Proof. By compactness, $[a, b]$ has a sequence $d_1, d_2, \dots,$ such that the set of values of d_i is dense on $[a, b]$. That means for any $\delta > 0$ we can find $J \in \mathbb{N}$ such that every $x \in [a, b]$ is within δ of some d_j where $j \le J$.

In that case, for each δ, the set

$$\{(d_j - \delta, d_j + \delta) \mid j \in \mathbb{N}\}$$

is an open covering of $[a, b]$ and hence has a finite subcovering.

Consider $d_1 \in [a, b]$. Because $(f_n(d_1))_{n\in\mathbb{N}}$ is a sequence in $[-M, M]$ there exists a subsequence $(f_{1,n})_{n\in\mathbb{N}}$ of $(f_n)_{n\in\mathbb{N}}$ such that $f_{1,n}(d_1)$ converges to some y_1.

[9] i.e. for some M, $|f_n(x)| \le M$ for all n, x.

Then $(f_{1,n}(d_2))_{n\in\mathbb{N}}$ is again in $[-M, M]$ so it has some subsequence $(f_{2,n})$ such that $f_{2,n}(d_2) \to y_2 \in [-M, M]$. On the other hand $f_{(2,n)}(d_1)$ still approaches y_1 as it is a subsequence of a continuous sequence.

Hence as $k = 1, 2, \dots$ we construct $f_{k,n}$ such that $f_{k,n}(d_i) \to y_i$ for each $i < k$.

Remark that $f_{1,1}, f_{2,2}, f_{3,3}, \dots$ is indeed a subsequence of the f_i, because if $m < n$ then $f_{n,n}$ may not precede $f_{m,m}$ in any sequences. As such, we define $g_n = f_{n,n}$ and observe that $g_n(d_j) \to y_j$ as $n \to \infty$; just discard all the n with $n < j$.

Now we claim that the g_n converges uniformly. We already now g_n converges pointwise at each point of a dense subset. Now we apply equicontinuity, and we wish to show that (g_n) is uniformly Cauchy; that is for all $\varepsilon > 0$ there is an N such that $|g_n(x) - g_M(x)| < \varepsilon$ for all $x \in [a, b]$ and $n, m \geq N$.

By equicontinuity, there exists a $\delta > 0$ such that for all x, y within δ of each other, $|g_n(x) - g_m(y)| < \frac{1}{3}\varepsilon$ for each n. For this specific δ, we can consider J as in the very first paragraph so that $d_1, d_2, \dots, d_J$ are δ-dense. Thus there exists N_j such that if $m, n \geq N_j$ wit h$|g_n(d_j) - g_m(d_j)| < \frac{1}{3}\varepsilon$.

Now we take $N = \max\{N_1, \dots, N_J\}$. Then $m, n \geq N$ implies

$$|g_n(x) - g_m(x)| < |g_n(x) - g_n(d_j)| + |g_n(d_j) - g_m(d_j)| + |g_m(d_j) - g_m(x)|$$

for some d_j within δ of x. The end! □

Summary: get the sequence to converge to a sequence of dense points. Then use equicontinuity to get all of $[a, b]$.

Remark 23.5. This holds for any compact metric space.

23.3 Sets of Equicontinuous Functions

We can instead look at sets of functions $\mathcal{F}$ rather than sequences. We can thus modify our theorem to read the following.

Theorem 23.6. *A set $\mathcal{F} \subseteq C^0$ is compact if and only if $\mathcal{F}$ is closed, bounded, and equicontinuous.*

Proof. One direction is trivial. The other direction is 2/3 trivial; we only need to prove $\mathcal{F}$ is continuous given compactness. Consider an open subcover

$$M_{\varepsilon/3}(f) : f \in \mathcal{F}$$

where $M_{\varepsilon/3}$ is the $\varepsilon/3$ neighborhood. Hnece we obtain a finite subcover $M_{\varepsilon/3}(f_1), \dots, M_{\varepsilon/3}(f_N)$. Hence there exists δ_i with uniform continuity of f_i for $i = 1, 2, \dots, N$. Let $\delta = \min\{\delta_1, \dots, \delta_N\}$.

Just note now that for all $f \in \mathcal{F}$, we have $|x - y| < \delta$ implying that

$$|f(x) - f(y)| < |f(x) - f_n(x)| + |f_n(x) - f_n(y)| + |f_n(y) - f(y)|$$

for some f_n within $\frac{\varepsilon}{3}$ of f. This is thus less than ε. □

24 November 21, 2013

Sports injuries from math contests? I guess.

25 November 21, 2013

25.1 The Stone-Weierstrass Theorem

Definition 25.1. A *function algebra* is an algebra which is closed under addition, scalar multiplication, and function multiplication.

Let $\mathcal{A} \in C^0(M, \mathbb{R})$ be a function algebra, where M is a compact metric space which *separates* every pair of points; that is, $\forall p, q \in M \exists f \in \mathcal{A}$ such that $f(p) \neq f(q)$. Suppose that $\forall p \in M, \exists f \in \mathcal{A}$ such that $f(p) \neq 0$.

We will show the closure $\overline{\mathcal{A}}$ is the entire C^0.

25.2 First Lemmas

Lemma 25.2. *Assume the two properties above. For all $p_1, p_2 \in M$ and $c_1, c_2 \in \mathbb{R}$, we can find a function $f \in \mathcal{A}$ with $p_1 \mapsto c_1$ and $p_2 \mapsto c_2$.*

Proof. This is a matter of cooking up functions.

Let $g_i \in \mathcal{A}$ not vanish at p_i for $i = 1, 2$ and let $h \in \mathcal{A}$ separate p_1 and p_2. For convenience, define $g = g_1^2 + g_2^2$ and note that $g(p_1), g(p_2) > 0$. We wish to find coefficients $\xi, \eta \in \mathbb{R}$ such that

$$\psi g(p_i) + \eta g(p_i) h(p_i) = 0$$

for $i = 1, 2$. Verify that the determinant is nonzero. □

We also state the following two facts.

Fact 25.3. If $f, g : M \to \mathbb{R}$ are continuous, and $f(p) < g(p)$ then there exists U a neighborhood of p such that $f(u) < g(u)$ for each $u \in U$.

Proof. Clear enough. □

Fact 25.4. The closure of a function algebra is also a function algebra.

Proof. Just use properties of convergence. □

25.3 Getting absolute values and maximums

Lemma 25.5. *If $\mathcal{A}$ is a function algebra and f is in $\overline{\mathcal{A}}$, then so is $|f|$.*

Proof. There exists a large $B \in \mathbb{R}$ such that f takes values in $[-B, B]$ (because M is compact). The Weierstrass Approximation Theorem produces a polynomial

$$P(y) = a_0 + a_1 y + a_2 y^2 + \cdots + a_n y^n$$

for which $|P(y) - |y|| < \varepsilon$ for each $y \in [-B, B]$. We may also assume $a_0 = 0$ since $|0| = 0$. (Specifically, tighten $P(y)$ to be within $\frac{1}{2}\varepsilon$ and then consider $Q = P(y) - y$.)

Substitute $y = f(x)$ to produce

$$P(y) = a_1 f(x) + a_2 f(x)^2 + a_3 f(x)^3 + \cdots + a_n f_n(x)^n \in \overline{\mathcal{A}}.$$

Evidently for all x in M, we get that

$$|Q(f(x)) - |f(x)|| < \varepsilon.$$

This implies we approach $|f|$ with functions in $\overline{\mathcal{A}}$, namely $Q \circ f$. Hence $|f|$ lies in the closure. □

Note that because we have used the Weierstrass Approximation Theorem, the general Stone-Weierstrass Theorem cannot be used to deduce the Weierstrass Approximation Theorem. This is a case of the specific theorem being used to derive the general theorem.

We now write
$$\max\{a,b\} = \frac{a+b}{2} + \frac{|a-b|}{2}$$
and similarly for the minimum. To this end we can take
$$\max(f,g) = \frac{f+g}{2} + \frac{|f-g|}{2} \in \overline{\mathcal{A}}.$$

Then $\max(f,g,h) = \max(\max(f,g),h)$, and accordingly we can show that
$$\max(f_1,\dots,f_n) \in \mathcal{A}.$$

25.4 Proof of the Stone-Weierstrass Theorem

We wish to show that given $F \in C^0$ and $\varepsilon > 0$, we can find $G \in \overline{\mathcal{A}}$ such that $\|G - F\| < \varepsilon$; that is,
$$F(x) - \varepsilon < G(x) < F(x) + \varepsilon \quad \forall x \in M.$$

First, we fix $p \in M$. Then for all $q \in M$, we can find $H_{p,q} \in \overline{\mathcal{A}}$ sending $p \mapsto F(p)$ and $q \mapsto F(q)$ (by the first lemma).

We may think of q and x as varying. Remark that we may find U_q a neighborhood of q such that $\forall x \in U_q$, $F(x) - \varepsilon < H_{pq}(x)$ (again this is with our given ε). But now
$$\bigcup_{q\in M} U_q$$
is a open covering! So we can find a finite subcover
$$\{U_{q_1},\dots,U_{q_m}\}.$$

Now we define
$$G_p = \max\{H_{pq_1},\dots,H_{pq_m}\} \in \overline{\mathcal{A}}.$$
By construction $G_p(x) > F(x) - \varepsilon$.

So, what now? Now we unfix $p \in M$! Then we can obtain a neighborhood V_p with $G_p(v) < F(v) + \varepsilon$ for each $v \in V_p$. Again we can obtain a finite subcover! We obtain
$$G_{p_1}, G_{p_2}, \dots, G_{p_n}$$
and now consider
$$G = \min\{G_{p_1},\dots,G_{p_n}\} \in \overline{\mathcal{A}}.$$
By construction, $G(x) < F(x) + \varepsilon$ for all x. And yet $G(x) > F(x) - \varepsilon$ for all x because the G_{p_k} each have that property.

Hence $\|F - G\| < \varepsilon$ and we are done.

Problem-solving tactic: the G_p *supersolve* the F we are trying to approximate. So rather than trying to hit F immediately, we get something bigger than F, and then push it downwards.

26 November 26, 2013

26.1 Contractions

Definition 26.1. A *contraction* is a map $f : M \to M$ such that for some $k < 1$ we have

$$d(fx, fy) \le kd(x, y)$$

for all $x, y \in M$.

Theorem 26.2 (Banach Contraction Theorem)**.** *If M is a complete metric space, there exists a unique point $p \in M$ such that $fp = p$; that is, there exists a unique fixed point.*

Proof. Choose any $x_0 \in M$ arbitrarily and let $x_n = f^n(x_0)$ for every positive integer n. We claim (x_n) is Cauchy. After all, $d(x_n, x_{n+1}) \le k^n(x_0, x_1)$ by a simple induction, and hence $d(x_n, x_m)$ can be bounded by (for $n < m$)

$$(k^n + k^{n+1} + \dots k^{m-1})d(x_0, x_1) < \frac{k^n}{1-k}d(x_0, x_1)$$

as desired.

Hence $x_n \to p$ for some p and it is easy to show that $d(p, fp) < \varepsilon$ for any $\varepsilon > 0$, implying $d(p, fp) = 0$. Uniqueness is immediate. □

Second Proof. Pick some large R and $x_0 \in M$. Let $A = \{x : d(x, x_0) \le R\}$. We wish to show $f(A) \subseteq A$. Note that

$$d(x_0, fx) \le d(x_0, fx_0) + d(fx_0, fx) \le r + kd(x_0, x) = r + kR.$$

where $r = d(x_0, fx_0)$. This is less than R for sufficiently large R. Because $A \supset f(A)$, we obtain $f(A) \supset f(f(A))$ (this is set theory, just take f of both sides). We conclude that

$$A \supset f(A) \supset f^2(A) \supset \dots$$

and the diameters tend to zero. But A is closed. Unfortunately, $f(A)$ is not necessarily closed. So, we take the closure, and notice the diameter doesn't change. That is

$$A \supset \overline{f(A)} \supset \overline{f^2(A)} \supset \dots$$

and the diameters still tend to zero. Hence the intersection $\bigcap \overline{f^n(A)}$ is a single point, the fixed point. □

26.2 Ordinary Differential Equations

The following result is the standard application of the Bananch Contraction Theorem.

Recall that an ordinary differential equation has the form

$$x' = f(x) \qquad x(0) = x_0.$$

Here x is a function of one variable t.

Given f a real-valued function of the real variable x (ugh), we wish to find a function $x(t)$ such that

$$\frac{dx(t)}{dt} \equiv f(x(t)), \quad x(0) = x_0.$$

Example 26.3. If $f(x) = ax$, the solution to $x' = ax$ with $x(0) = 1$ is $x(t) = e^{at}$.

If $f(x) = x^2$, the solution is $x(t) = \frac{-1}{t-c^{-1}}$. Now $x(0) = c$.

These are quite specific single-variable ODE's. Other ODE's are more general:

$$\begin{aligned} x' &= f_1(x,y) \\ y' &= f_2(x,y) \end{aligned}$$

and with some initial condition $x(0) = x_0$ and $y(0) = y_0$.

To solve this means to find a curve

$$(x(t), y(t))$$

such that $\frac{dx}{dt} \equiv f_1(x(t), y(t))$ and $\frac{dy}{dt} \equiv f_2(x(t), y(t))$.

Of course, we can generalize this further to multiple variables. But the picture is this; for any open $U \in \mathbb{R}^2$, consider the field of vectors at each point in U by

$$(x,y) \mapsto (f_1(x,y), f_2(x,y)).$$

Then a solution is a curve tangent to the vectors at every point.

Theorem 26.4 (Picard)**.** *Let $U \subset \mathbb{R}^m$ be open. Suppose that $f : U \to \mathbb{R}^m$ satisfies a Lipschitz condition: there exists $L \in \mathbb{R}$ with $|f(x) - f(y)| \le L|x-y|$ for every $x, y \in U$. Then the ODE $x' = f(x)$, $x(0) = x_0$ has a unique solution $\gamma(t)$; we have*

$$\frac{d\gamma(t)}{dt} \equiv f(\gamma(t)), \quad \gamma(0) = x_0.$$

By unique we mean that any two solutions agree on their common time interval. That is, uniqueness up to extending or restricting time intervals.

Proof. We can instead search for a continuous curve $\gamma(t)$ (not necessarily differentiable) with

$$\gamma(t) = x_0 + \int_0^t f(\gamma(s))\, ds.$$

for all t. This is sufficient, since $\gamma(0) = 0$ and $\gamma'(t) = f(\gamma(t))$ by the Fundamental Theorem of Calculus.

The upshot is that we don't care about differentiability now, because if γ is continuous as above then it is automatically differentiable.

Here's what we do. Consider a compact ball around $x_0 \in U$, say

$$N = \{x \in U : |x - x_0| \le r\} \subset U.$$

By Heine-Borel, N is compact. Hence there exists M such that for all $x \in N$, we have

$$|f(x)| \le M.$$

We can select a time $\tau > 0$ with $\tau L < 1$ (where L is from the Lipschitz condition) and $\tau M < r$.

Consider the function space

$$\mathcal{C} = \{\gamma : [-\tau, \tau] \to N \mid \gamma(0) = x_0,\ \gamma \text{ continuous}\}.$$

equipped with the sup norm. We know that $\mathcal{C}$ is complete relative to this sup metric. Now consider a map T from $\mathcal{C}$ to itself by

$$T : \gamma \mapsto x_0 + \int_0^t f(\gamma(s))\, ds.$$

By virtue of and $\tau M < r$, one can show that the right-hand side is contained in N for any t. Hence $T(\gamma)$ indeed lies in $\mathcal{C}$.

Now consider $\sigma, \gamma \in \mathcal{C}$. We wish to show $d(T\sigma, T\gamma) \le kd(\sigma, \gamma)$. Compute

$$d(T\sigma, T\gamma) = \sup \left| \int_0^t \left(f(\sigma(s)) - f(\gamma(s)) \right) \, ds \right|.$$

By Lipschitz this is less than

$$\int_0^t L \left| \gamma(s) - \sigma(s) \right| \, ds < \tau L d(\sigma, \tau)$$

Hence take $\tau L < 1$ and we win! We get a fixed point of T which is the desired path. $\square$

> We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at any given moment knew all of the forces that animate nature and the mutual positions of the beings that compose it, if this intellect were vast enough to submit the data to analysis, could condense into a single formula the movement of the greatest bodies of the universe and that of the lightest atom; for such an intellect nothing could be uncertain and the future just like the past would be present before its eyes.
> – Laplace

27 November 27, 2013

We will construct a uniformly continuous function which is differentiable nowhere.

Theorem 27.1. *There exists $f : \mathbb{R} \to \mathbb{R}$ continuous and for all $x \in \mathbb{R}$, f is not differentiable at x.*

It gets worse.

Theorem 27.2. *The "generic" continuous function $f : [a, b] \to \mathbb{R}$ is nowhere differentiable.*

Here, generic applies in the sense that e.g. the generic 2×2 matrix is nonzero, the generic real number is irrational.

27.1 Constructing the Beast

We begin with the sawtooth function $\sigma_0(x)$ as follows. For each $n \in \mathbb{Z}$, set $\sigma_0(x) = x - 2n$ for $2n \le x \le 2n+1$, and $\sigma_0(x) = 2n + 2 - x$ for $2n + 1 \le x \le 2n + 2$. In effect, this is a 2-periodic mountain.

Now for each $k = 1, 2, \dots$ we define σ_k by

$$\sigma_k(x) = \left(\frac{3}{4}\right)^{\sigma_0(4^k x)}.$$

Note that the slopes in σ_k are quite steep, albeit uniformly continuous. The period of σ_k is $2 \cdot 4^{-k}$.

Now we define

$$f(x) = \sum_{k=0}^{\infty} \sigma_k(x).$$

This is continuous by the Weierstrass M-test. Now we claim f is non-differentiable. Here is why. Let $\delta_n = 4^{-n}$. We claim that for all x, for some suitable choice of $\pm$, we have

$$\left| \frac{f(x \pm \delta_n) - f(x)}{\delta_n} \right| \to \infty.$$

Remark that the quantity in the absolute values may be written in the form

$$\frac{1}{\delta_n}\left(\sum_{k=0}^{n-1} \left(\sigma_k(x \pm \delta_n) - \sigma_k(x)\right) + \left(\sigma_n(x \pm \delta_n) - \sigma_n(x)\right) + \sum_{k=n+1} \infty \left(\sigma_k(x \pm \delta_n) - \sigma_k(x)\right) \right).$$

But now, by construction, $\sigma_k(x \pm \delta_n) - \sigma_k(x) = 0$ for each $k \le n+1$. After all, δ_n divides the period of the function.

How about $\delta_n(x \pm \delta_n) - \delta_n(x)$? One of the sides is monotone by our construction, and has slope 3^n in magnitude. Pick the sign that corresponds to that. That means

$$\left| \frac{\delta_n(x \pm \delta_n) - \delta_n(x)}{\delta_n} \right| = 3^n$$

for some choice of $\pm$.

How about the rest of the terms? When $k < n$, we have that $\frac{1}{\delta_n}\left(\sigma_k(x \pm \delta_n) - \sigma_k(x)\right) < 3^k$, but $\left|1 + 3 + \dots + 3^{n-1}\right| < \frac{1}{2} \cdot 3^n$. Hence,

$$\left| \frac{f(x \pm \delta_n) - f(x)}{\delta_n} \right| > \frac{1}{2} \cdot 3^n.$$

27.2 Generic?

Definition 27.3. Let M be a metric space. A set $S \subset M$ is *thick* if it is a countable intersection of open and dense subsets of M.

Example 27.4. $\mathbb{R} \setminus \mathbb{Q}$ is thick in $\mathbb{R}$. For each rational number α define

$$G(\alpha) = \mathbb{R} \setminus \{\alpha\}.$$

Each $G(\alpha)$ is open and dense in $\mathbb{R}$, and the intersection of these countably many G's is precisely $\mathbb{R} \setminus \mathbb{Q}$.

Definition 27.5. A set $S \subset M$ is *thin* if its complement is thick.

Theorem 27.6 (Baire's Theorem)**.** *If M is a complete metric space and $S \subset M$ is thick, then S is dense in M.*

Completeness is necessary. Otherwise, by replacing $\mathbb{R}$ with $\mathbb{Q}$ in our prior example, we find that $\varnothing$ is thick in $\mathbb{Q}$.

Proof. Given $p_0 \in M$ and $\varepsilon_0 > 0$, we wish to show that $S \cap M_{\varepsilon_0}(p_0) \neq \varnothing$. Set $S = \cap_{n=1}^{\infty} G_n$, where each n is open and dense.

By density of G_1, we can furnish an $\varepsilon_1 > 0$ and $p_1 \in G_1$ such that the closure of $M_{\varepsilon_1}(p_1)$ is contained in the open set $M_{\varepsilon_0}(p_0) \cap G_1 \neq \varnothing$. Now we just repeat, constructing a $p_1, p_2, \dots$. We can also dictate that $\varepsilon_n < \frac{1}{n}$ when we select our ε_n.

As a result, we have constructed

$$M_{\varepsilon_0}(p_0) \supset \overline{M_{\varepsilon_1}(p_1)} \supset \overline{M_{\varepsilon_2}(p_2)} \supset \dots$$

By completeness, these converge to a unique point q in the intersection of all these sets. Then $q \in G_n$ for each n, since we dictated that $M_{\varepsilon_n}(p_n) supset M_{\varepsilon_{n-1}}(p_{n-1}) \cap G_n$. Therefore, $q \in S$. $\square$

27.3 Usually Nowhere Differentiable

Here is the precise statement.

Theorem 27.7. *There exists a sequence R_n of open dense sets in $C^0([a, b], \mathbb{R})$ such that $\forall F \in \bigcap_{n=1}^{\infty} R_n$, F is nowhere differentiable.*

Proof. This proof will use the Weierstrass Approximation Theorem.

Define

$$R(n) \in \left\{ f \in C^0 : \forall x \in [a, b - \frac{1}{n}] \exists h > 0 \text{ with } \left| \frac{f(x+h) - f(x)}{h} \right| > n \right\}.$$

The point is that as n gets big, we have functions with terrible right slopes. We have to use $[a, b - \frac{1}{n}]$, though, because $x + h$ needs to be defined.

We claim that R_n is open and dense. First we show density. The Weierstrass Approximation Theorem implies that the set of all polynomials $\mathcal{P}$ is dense in C^0. It suffices to show that $\mathcal{P} \subset \overline{R_n}$. Given a polynomial P we wish to find $f \in \mathbb{R}^n$ such that $\|f - P\| < \varepsilon$. There exists B a bound for $P'(x)$. Now we just consider $P + \sigma_m$ for large m. It's a terrible function with terrible slopes, and in particular belongs to R_n if m is large enough, but nonetheless $\|(P + \sigma_m) - P\| = \|\sigma_m\| \to 0$.

Proving R_n is open is trickier. Given $f \in R_n$ we need to find some $\varepsilon > 0$ such that $\|g - f\| < \varepsilon \Rightarrow g \in R_n$. We need to loosen the condition slightly. Declare that

$$\left| \frac{f(x+h) - f(x)}{h} \right| > n + \nu(x)$$

for some $\nu(x) > 0$. Now we use propagation. For every x there exists T_x an open subinterval of $[a, b]$ such that

$$t \in T_x \Rightarrow \left| \frac{f(t + h(x)) - f(t)}{h(x)} \right| > n + \nu(x).$$

By compactness of $[a, b - \frac{1}{n}]$ so we can find finitely many T_x which actually cover the entire interval. We can find a suitable ε_i for each of the T_{x_i}, and then we can simply take $\varepsilon = \min\{\varepsilon_i\}$.

Great. That means $\bigcap R_n$ is indeed thick. We just need to show that its members are wicked, wicked functions.

For all $f \in \bigcap R_n$ and $x \in [a, b)$, we know that there is some sequence of positive (h_n) such that

$$\left| \frac{f(x+h_n) - f(x)}{h_n} \right| \to \infty.$$

Because $f(x+h_n) - f(x)$ is bounded, this means $h_n \to 0$ and hence f is not differentiable at any $x \in [a, b)$. $\square$

27.4 Extended Doom

The generic function is also monotone on no interval – it happens to often be a fractal.

28 December 5, 2013

Final exam: 234 Hearst Gym, from 7PM - 10PM on Friday, December 20. (Ergh. . .)

Office hours on the 17th and 19th of December from 4PM-5:30PM in 807 Evans.

The test will emphasize the second part of the course in the ratio 5:3 or 2:1. There will be about six questions.

28.1 Review of Chapter 3

In this chapter we discuss functions $f : [a, b] \to \mathbb{R}$.

Differential calculus. What is f'? We have some properties.

- Sums, differences of differentiable functions are differentiable.
- Chain Rule
- Mean Value Theorem
- Ratio Mean Value Theorem. Looking at $\frac{f(x)}{g(x)} - \frac{f(y)}{g(y)}$, and finding a θ with $\frac{f'(\theta)}{g'(\theta)}$.
- RMVT implies L'Hospital's Rule. The cases are $\frac{0}{0}$ and $\frac{\infty}{\infty}$.
- Intermediate Value Property for $f'(x)$. As long as $f'(x)$ exists, even if it is not continuous, it has no jump discontinuities. More formally, if $f'(x)$ exists at all $x \in (a, b)$ and $f'(\alpha) < k < f'(\beta)$ then there is some θ between α and β for which $f'(\theta) = k$.

  Note that there are differentiable functions f, but $f'(x)$ is not even bounded.
- Higher derivatives. Recall C^0 is the set of continuous functions $[a, b] \to \mathbb{R}$, and C^1 is the set of all functions which are continuously differentiable[10]. Define $C^2, C^3, \dots$ analogously. Define $C^\infty = \bigcap C^r$, called *smooth*.
- A function is *analytic* if at every point it is given by a power series. The set is denoted C^ω. All such functions are smooth. Remember the bump function

  $$x \mapsto \begin{cases} e^{-1/x} & x > 0 \\ 0 & x \le 0. \end{cases}$$

  It is smooth, but cannot be expressed as a power series at zero.

Fortunately, (Riemann) integral calculus is much nicer. Consider once more an arbitrary function $f : [a, b] \to \mathbb{R}$.

- We consider meshes. Select

  $$P : x_0 = a < x_1 < x_2 < \cdots < x_n = b$$

  and

  $$T : t_1 \le t_2 \le \cdots \le t_n \text{ with } t_i \in [x_{i-1}, x_i] \text{ for all } i.$$

  Then we define

  $$\mathcal{R}(f, P, T) = \sum f(t_i)\Delta x_i$$

  where $\Delta x_i - x_{i-1}$. Finally, the mesh of P, sometimes denoted $\|P\|$, is given by $\max_{1 \le i \le n} \Delta x_i$.

[10] So, their derivatives exist and are continuous

- We say f is Riemann integrable if $\lim R(f, P, T)$ exists as the mesh of P approaches zero.
- It is easy to check that any Riemann integrable function must be bounded. This is far from sufficient; consider $\chi_{\mathbb{Q}}$.
- Darboux integrability is used in the proof of the Riemann-Lebesgue Theorem. Consider a bounded function f and a partition $P : a = x_0 < x_1 < \cdots < x_n = b$. We define
$$L(f, P) = \sum_{k=1}^{n} m_k \Delta x_k \text{ where } m_k = \inf_{x \in [x_{k-1}, x_k]} f(x)$$
$$U(f, P) = \sum_{k=1}^{n} M_k \Delta x_k \text{ where } M_k = \sup_{x \in [x_{k-1}, x_k]} f(x)$$
One can show $L(f, P_1) \le U(f, P_2)$. The proof uses common refinement. Now we define
$$\overline{I} = \inf \{U(f, P) : P \text{ a partition}\}$$
and
$$\underline{I} = \sup \{L(f, P) : P \text{ a partition}\}.$$
We say f is Darboux integrable when $\overline{I} = \underline{I}$.
- Darboux integrability is equivalent to Riemann integrability, and naturally, the integrals match.
- Riemann-Lebesgue Theorem: A very happy theorem. A function f is Riemann integrable if and only if f is bounded and its set of discontinuity points, $D(f)$, is a zero set.
  - The function given by $x \mapsto \sin x^{-1}$ for $x \ge 0$ and $x \mapsto 0$ otherwise is Riemann integrable.
  - χ_C is Riemann integrable. In general, $D(\chi_A) = \overline{A} \cap \overline{A^c}$.
  - Monotone increasing functions are bounded and have all jump discontinuities, so they have countably many discontinuities.
- Series and the tests (ratio, root, etc.)
- Conditional convergence of $\sum a_n$. Rearrangements to anything.

28.2 Review of Chapter 4

- The metric $d(f, g) = \|f - g\| = \sup_{x \in [a,b]} |f(x) - g(x)|$ is a metric on C_b, the set of bounded functions.
- $\|f_n - g\| \to 0$ if and only if $f_n \rightrightarrows g$.
- C_b is complete metric space under this metric.
- $C^0 \subset \mathcal{R} \subset C_b$, the inclusions are strict, and C^0 and $\mathcal{R}$ are both closed in C_b.
- If $f_n \in \mathcal{R}$ and $f_n \rightrightarrows f$, then $f \in \mathcal{R}$ and $\int_a^b f_n(x)\, dx \to \int_a^b f(x)\, dx$.
- If each f_n is differentiable and $f_n \rightrightarrows f$, it is still conceivable that f is not differentiable. For example, take $f_n : [-1, 1] \to \mathbb{R}$ by $x \mapsto \sqrt[2n+1]{x^{2n+2}}$. If we add the hypothesis that f_n' converges uniformly to something, then that is okay.

- Equicontinuity: quite strong. Consider a set $\mathcal{E}$. Suppose that for every $\varepsilon > 0$ there is a single $\delta > 0$ such that if s and t differ by less than δ, then for any $f \in \mathcal{E}$ we have $|f(s) - f(t)| < \varepsilon$.
- Arzela-Ascoli Theorem. Suppose (f_n) is equicontinuous and a universal bound M. Then there is a uniformly convergent subsequence. The converse is also true.
- Weierstrass Approximation Theorem: Polynomials are dense in C^0.
- Stone-Weierstrass Theorem: Any function algebra $A \in C^0$ which separates points and does not vanish anywhere.
- Banach Contraction Mapping Theorem, ODE's, and Picard's Theorem.
- The generic continuous function is nowhere differentiable. The Weierstrass function is continuous everywhere and differentiable nowhere.
- Lipschitz condition: suppose there is a global L with $|f(x) - f(y)| < L\,|x - y|$, valid for all f and x, y. Then the functions are equicontinuous.