MATH 55B SPRING 2017

TAUGHT BY PROF. NOAM ELKIES

Contents

Part 1
Topology

1. Lecture 01 – 23 January 2017

Almost Always: $\mathcal{S}(55\text{b}) \subset \mathcal{S}(55\text{a})$

"I realize we have a sense of 'almost all' that means 'with finitely many exceptions' which is meaningless for finite sets"–NE

1.1. **Announcements.** There will be no class on Wednesday, but there will be a department lunch! Be there for Mexican food. Today there are handouts, including problem sets 1 and 2, as well as a topology handout.

The CAs will be holding office hours tonight, from 8 to 10 PM, though not on Wednesday.

1.2. **Course Summary.** Analysis is the formal study of calculus. We will begin by studying differentiable maps $\mathbb{R} \to \mathbb{R}$, then continue with maps from $\mathbb{R}^m \to \mathbb{R}^n$ (or $V \to W$ $\mathbb{R}$-vector spaces, if you loved 55a), or from some reasonably nice subset thereof. We will finish with a study of complex differentiable maps from $\mathbb{C} \to \mathbb{C}$, which interestingly can also be used to study real analysis. For instance, the simplest way to evaluate $\int_{-\infty}^{\infty} \frac{\cos ax}{x^2+1} dx$ is to pretend that we're doing a complex integral, and then our integral detects a pole at $x = i$, which we can compute separately to find the integral.

1.3. **Introduction to Metric Topology.** Historically calculus was discovered before topology. But though topology came later, it is the natural context for analysis, providing the "correct" notion of continuity. To do this, we want to formalize what we mean by distance, and how this can be used to give structure to the a space as a whole.

Recall when we studied linear algebra, we at first pretended the important thing was the vector spaces themselves. However, as we developed the theory, we found the important thing was the complete category of vector spaces, including the maps between the objects. We will find something similar for the category of metric spaces: maps that preserve a metric are an important thing to study.

Definition 1.1. A metric space $X = (X, d)$ is a set X with a map $d : X \times X \to \mathbb{R}^{\geq 0}$ subject to the following conditions:

- For all x, y, $d(x, y) \geq 0$, $d(x, y) = 0$ if and only if $x = y$ (Positive Definiteness)
- For all x, y, $d(x, y) = d(y, x)$ (Symmetry)
- For all x, y, z, $d(x, y) + d(y, z) \geq d(x, z)$ (Triangle Inequality). Note this implies a similar inequality for $n > 3$ points.

Example 1.2. There are a number of familiar examples of metric spaces:

- $X = \mathbb{R}$, $d(a, b) = |a - b|$
- X is an inner product space $d(v, w) = \sqrt{\langle v - w, v - w \rangle}$
- If X is a metric space and $E \subset X$ then E is a metric space, where the distance function is the restriction

Another natural way we can build metric spaces is to take the product of two metric spaces. For instance, $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$, and we already have a standard metric on $\mathbb{R}$. Therefore, we'd like to get the usual Euclidean metric on $\mathbb{R}^2$ from these. To more generally define $d_{X \times Y}$ in terms of d_X and d_Y, we have several options.

- We could define $d_{X\times Y}((x, y), (x', y')) = \sqrt{d_X(x, x')^2 + d_Y(y, y')^2}$, which will induce the Euclidean norm in the above example. It is not horribly difficult to show that this will satisfy the conditions of Definition 1.1, but it is a bit inconvenient to use in practice.
- More simply, we could set $d_{X\times Y} = d_X + d_Y$, which in our $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ example, would induce the taxicab metric.
- Most frequently, however, we will use the sup metric, d_{sup}, given by
$$d_{\text{sup}}((x, y), (x', y')) = \max(d_X(x, x'), d_Y(y, y')).$$
This is frequently also called the max metric, or max norm.

It's not hard to show that the cartesian product is associative–that is, $(X \times Y) \times Z$ is canonically isomorphic to $X \times (Y \times Z)$ as a metric space. Since we know they're canonically the same as sets, using the sup norm, we simply need to check $\max(a, \max(b, c)) = \max(\max(a, b), c) = \max(a, b, c)$, which is obvious. The arguments for the other product norms are similar. With this in hand, we've defined any finite cartesian product of metric spaces.

We can really talk about distance on any *finite* cartesian product this way. What happens with an infinite product $\prod X_i$? It becomes hard to form the distance function as a sum because we don't allow infinite distances. Even max does not work, because the sequence of distances might not be bounded. For instance, using the max norm, what is the distance between $(0, 0, 0, 0, ...)$ and $(1, 2, 3, 4, ...)$? If we required all of the X_i were bounded by some common number M, however, we could use the max norm. (We would need to replace it with *sup*, which will be covered later in the course, as there might not be a maximum in an infinite set). This motivates another definition:

Definition 1.3. A metric space X is *bounded* if there exists some real N such that $d(x, y) < N$ for all $x, y \in X$.

Example 1.4. Here are some common bounded sets:

- The stupid example: The empty set is a bounded metric space (it can support a negative bound)
- The less stupid example: The singleton set is bounded with any positive bound

Lemma 1.5. *Let X be a metric space, $E \subset X$ is nonempty with the induced metric from X. Then the following are equivalent:*

1. *E is bounded*
2. *There is $p \in E$ and $M \in \mathbb{R}$ such that $d(p, q) < M$ for all $q \in E$*
3. *There is $p \in X$ and $M \in \mathbb{R}$ such that $d(p, q) < M$ for all $q \in E$.*

Proof. (1) implies (2) by taking $M = N$, and (2) implies (3) by using the same point. All that remains is to show (3) implies (1): by the triangle inequality, the distance between any two points $x, y \in E$ is less than the sum $d(x, p) + d(p, y) < 2M$, so taking $N = 2M$ finishes the argument. $\square$

2. Lecture 02 – 27 January 2017

2.1. **Announcements.** Handouts are all now on the website, though hardcopies are also available for now. Prof. Elkies needs to approve requests to add Math 55 to your Study Card, so make sure to do it soon, so you have time to properly enroll. Remember, you have plenty of time for add-drop if you still aren't sure whether or not you want to take Math 55.

2.2. **Metrics on function spaces.** Last time we introduced the notion of a bounded space. If $B \subset X$ then B being bounded is equivalent to there being some point $x \in X$ (not necessarily in B) such that any point in B is at most distance N away from x for a fixed N. This gives us a lot of bounded subsets, just by taking balls around points of X. You don't have to know what E is a subset of to tell whether or not it is bounded, this is an *intrinsic* property.

If S is a set X a metric space, then we want to put a metric space structure on the set of functions from S to X, often called X^S. We would like to have $d(f,g) = \max_{s \in S}(f(s), g(s))$. This satisfies the metric space axioms, but it only works for a finite set S, for an infinite set S max might not be defined. So we must assume S is bounded and we have no problem taking a maximum.

But there is a problem, let $S = (-1, 1)$, and look at a map from $S \to S$ where $f(a) = |a|$. compare the distance between this function and the function $g(a) = 0$. What is $d(f,g)$? It should be $\max[0,1)$, but the maximum is not contained in the set. So we need to use the notion of a suprenum.

Definition 2.1. If $S \subset \mathbb{R}$ (bounded above and nonempty) then $\sup(S)$ is the smallest real number r such that $r \geq s$ for all $s \in S$

Remark 2.2. The real numbers are a rather subtle object–this is what necessitates things like the use of the supremum. While 55b won't cover the construction of the reals, it's a good thing of which to be aware. One reference, for a construction via Dedekind cuts, can be found in chapter 1 of Rudin.

Given this, we can check that $\sup_{s \in S}(f(s), g(s))$ is indeed a metric.
Additionally, if S were not bounded, we could take the sup metric on functions that have bounded image.

2.3. **Open Sets.** One way to visualize the structure of a metric space is to, for a fixed point, envision all the points at most distance r away, or less than distance r away. These are closed and open sets, respectively. The set of points of distance less than r is often denoted

$$B_r(p) = \{s \in S | d(s,p) < r\}$$

. We will also require $r > 0$, and call this construction *the open ball of radius r*.

Definition 2.3. We say p is *in the interior* of E is there is an $r > 0$ such that $B_r(p) \subset E$. Equivalently, we could state that p is in the interior if every point x sufficiently close to p is also in E.

Note that this is *not* intrinsic to E–it very much depends on what space it is embedded in. E, considered as a subset of itself, for instance, always has every point in E in its interior.

Definition 2.4. We say E is *open* in X if E is equal to its interior. That is, E is open if every point of E is contained in a ball $B_r(p) \subset E$.

This is a very important notion, arguably more important than the distance function. But now we need to show our notation makes sense.

Lemma 2.5. *An open ball is an open set.*

Proof. Let $q \in B_r(p)$, then $d(p,q) = r - \epsilon$ for $\epsilon > 0$, so if $d(q,s) < \frac{\epsilon}{2}$ then $d(p,s) \leq d(p,q) + d(q,s) \leq r - \frac{\epsilon}{2}$ so $B_{\epsilon/2}(q) \subset B_r(p)$ □

Theorem 2.6. *The collection of open subsets of X is closed under*

- *finite intersections–that is, if $E_1, ...E_n$ are open, so is $E_1 \cap ... \cap E_n$,*
- *and arbitrary unions–that is, if I is any set, $\{E_i\}_{i \in I}$ are open, then $\cup_{i \in I} E_i$ is also open.*

Example 2.7. We cannot necessarily take even countable intersections. Consider, for instance the countable intersection

$$\bigcap_{n \in \mathbb{Z}^+} \left(-\frac{1}{n}, \frac{1}{n}\right) = \{0\}$$

is not an open subset of $\mathbb{R}$.

Proof. Unions are easy. If $p \in \bigcup E_i$ then $p \in E_i$ for some i so $B_r(p) \subset E_i$ for some r so $B_r(p) \subset \bigcup E_i$.
If $p \in \bigcap E_i$ then for all i $p \subset E_i$ so we can find r_i to make $B_{r_i}(p) \subset E_i$ Then let $r = \min(r_i)$, $B_r(p)$ is contained in all E_i so it is contained in the intersection so the intersection is open. □

Remark 2.8. While we could try to take the infinimum in the latter argument, to get a smallest ball, this wouldn't necessarily work because the infinimum of an infinite set of positive reals is not necessarily positive. (Infinimum is the opposite of the supremum, denoted respectively inf and sup.)

In the completely general sense, a topology on a set X is given by stating which subsets of X one considers open. Formally,

Definition 2.9. A *topological space* is a set X with a collection of distinguished subsets $\mathscr{U} = \{U_i\}_{i \in I}$, called open sets, such that

- $X, \emptyset \in \mathscr{U}$
- If $U, V \in \mathscr{U}$, then $U \cap V \in \mathscr{U}$.
- If $\{U_j\}_{j \in J}$ is a collection of open sets, then $\cup_{j \in J} U_j$ is an open set.

There are lots of topologies that don't come from the metric structure, though we will mostly be concerned with ones that do.

Example 2.10. Here are some non-metric topologies:

- Given any set X, we can define a topology by stating that the only open sets are the empty set and the entire set. (This is a metric topology only if $X = \{\text{pt}\}$.)
- The cofinite topology (which generalizes in Algebraic Topology to the very important Zariski topology) is defined by setting open sets to be the entire set, minus only finitely mainy points, plus the empty set.

Different metrics can give the same topology. We noted in Math 55a that, on a finite dimensional vector space, all norms are equivalent. What this means topologically is that all norms define the same topology on a finite dimensional vector space.

2.4. **Closed Sets.** Go to `https://www.youtube.com/watch?v=SyD4p8_y8Kw`

Definition 2.11. We say a set $E \subset X$ is *closed* if $X - E$ is open.

Note that a closed set can also be open–for instance, both the entire set and the empty set are both closed and open subsets of any topological space X. Most sets are neither open nor closed (consider, for instance, $[0, 1)$ is neither open nor closed in $\mathbb{R}$, since it's clearly not open, nor is $(-\infty, 0) \cup [1, \infty)$).

Remark 2.12. Closed sets give the same data as open sets do, so we could have used them for our definition of a topological space. Our choice to use opens is mostly convention.

Theorem 2.13. *Closed sets are preserved under finite unions and arbitrary intersections*

Proof. Take 2.6 and take the complement of everything. □

We define $\overline{B}_r(p)$, the closed ball centered at p, to be the set

$$\overline{B}_r(p) = \{x \in X \mid d(x, p) \leq r\}.$$

We will generally require $r > 0$ for closed balls, though arguments could be made to allow $r = 0$. Then, for instance, we'd like to show

Lemma 2.14. *The complement of $\overline{B}_r(p)$ is open–that is, $\overline{B}_r(p)$ is closed.*

Proof. If $q \notin \overline{B_r(p)}$ then $d(p, q) = r + \epsilon$ so by triangle inequality $B_{\epsilon/2}(q)$ does not meet $\overline{B_r(p)}$ so the complement of $\overline{B_r(p)}$ is open □

Definition 2.15. The closure of a set $E \subset X$, denoted $\overline{E}$, is the smallest closed set containing E. We can define it to be the intersection of all closed sets that contain E.

3. Lecture 03 – 30 January 2017

3.1. **The Axiom of Choice.**

Speaking about the axiom of choice...I'd rather not

We showed last week that an arbitrary union of open sets is open, and that open balls are open. In fact, any open $E \subset X$ can be written as a union of open balls (this is definitionally the topology associated to a metric space). Given an open set it contains a ball around every point, and the union of a ball around every point in E is E, for instance, proves the result.

We needed the axiom of choice to choose a radius for every point in the set, so we have assumed choice. We will continue to assume it.

You could avoid this by redefining open by giving it a function $E \to \mathbb{R}^+$, where each point is sent to a radius, but making such a function is hard. Much easier is assuming the axiom of choice.

3.2. **Continuity.**

As fundamental as spaces are the natural maps between spaces. We have already seen some maps between metric spaces:

- The identity map
- the distance function from $X \times X$ to $\mathbb{R}$
- Isometries

Intuitively, the appropriate kind of map for these objects are those that are continuous–informally, points that are close together should stay close together.

Definition 3.1. A map of metric spaces $f : X \to Y$ is *continuous at a point* $p \in X$ if for all $\epsilon > 0$ there exists some $\delta >$ such that if $d_X(p, q) < \delta$ then $d_Y(f(p), f(q)) < \epsilon$
We say that a function is *continuous* if it is continuous at every point.

This means that if we want to guarantee a point q near p gets sent to something near $f(p)$ we can do so by making q close enough to p

Example 3.2. We're familiar with many continuous functions–for instance, polynomials are intuitively continuous. Some simple examples of continuous functions are:

- A composition of two continuous functions is continuous (We leave the proof of this until later, when we get a nice lemma)
- If f and g are continuous maps from $X \to \mathbb{R}$ or $\mathbb{C}$ then $f + g, f - g, fg$ are all continuous, and $\frac{f}{g}$ is continuous if g is never 0.
- The constant function is continuous
- Isometries are automatically continuous
- For a fixed p_0 The function $X \to \mathbb{R}$ $x \mapsto d_X(p_0, p)$. Proof of continuity is the triangle inequality. If $d(q, q') < \delta$ then $|d(p_0, q) - d(p_0, q')| < \delta$
- The function $d : X \times X \to \mathbb{R}$ is continuous. Draw a triagnle with one corner (p, q) one corner (p', q') and one corner (p, q'). You wind up choosing $\delta = \frac{\epsilon}{2}$.

We can prove the second example is continuous by using the first: if we have two maps $f, g : X \to \mathbb{R}$ then we get a map $(f, g) : X \to \mathbb{R} \times \mathbb{R}$. But we also have a map $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$ that is addition (or subtraction or multiplication). Then using the first example, it suffices to check the following Lemmas:

Lemma 3.3. *If $f : X \to Y_1$ and $g : X \to Y_2$ then $(f, g) : X \to Y_1 \times Y_2$ is continuous.*

Lemma 3.4. *The maps $(r, s) \mapsto rs$, $(r, s) \mapsto r + s$, $(r, s) \mapsto r - s$, and $(r, s) \mapsto \frac{r}{s}$ are all continuous (where defined).*

These lemmas are not hard, and their proof is left as an exercise. *Hint for division*: prove $x \mapsto \frac{1}{x}$ is a continuous map of $\mathbb{R} - \{0\}$.

So far our choice of δ has been invariant of p, but it need not be. Showing continuity is essentially playing a game where an adversary picks p and ϵ and you must find a δ

Definition 3.5. If the choice of δ can be made without regard to p, the function is *uniformly continuous*.

Multiplication and inversion are *not* uniformly continuous. In inversion, for instance, we must pick a δ for each p so that the ball of radius δ does not contain 0.

Remark 3.6. We usually define a continuous map on a subspace of a metric space. Be careful here; an open ball around a point in a subspace is the intersection of an open ball in X with the subspace. This is what is known as the *subspace topology*. A set U is open (respectively closed) in a subspace $E \subset X$ if $U = E \cap V$ for some open (respectively closed) $V \subset X$.

Though continuity is defined by inequalities, there is a geometric picture. This essentially says that the image of an open ball of radius δ around p is mapped into the open ball of radius ϵ around $f(p)$. So the preimage of $B_\epsilon(f(p))$ contains $B_\delta(p)$. This is equivalent to our earlier definition.

Theorem 3.7. *f is continous if and only if the preimage of every open set is open.*

Proof. See the handout. It is not too hard of an epsilon-delta argument (though perhaps annoying). □

This is actually the topological definiton of continuity, for when we don't have a distance function, just open sets. Note uniform continuity is meaningless in the broader category of topological spaces.

Corollary 3.8. *Composition of continuous functions is continuous*

Proof. Take a preimage twice. The set is still open. □

4. Lecture 04 – 01 February 2017

$$\begin{aligned} 0! - 1! + 2! - 3! + 4! - ... &= \int_0^\infty e^{-x}dx - \int_0^\infty xe^{-x}dx + ... \\ &= \int_0^\infty (1 - x + x^2 - x^3 + ...)e^{-x}dx \\ &= \int_0^\infty \frac{e^{-x}dx}{1+x} \approx .56 \end{aligned}$$

(Quod infernorum!?!) This is one attempt Euler made to compute the above series. However, it's kind of nonsensical.

> "If you saw this online today, you'd probably be tempted to start your reply with three capital letters, the first two of which were 'WT.'"

4.1. **Sequences.**

Definition 4.1. A *Sequence* in a set X is an infinituple of elements or a function $\mathbb{Z}^+ \to X$

Beyond this, we'd sort of like to have a notion of convergence of a sequence. For instance, the sequence $(3, 3.1, 3.14, 3.141, ...)$ converges to π.

Definition 4.2. p is *a limit* of the sequence $\{p_i\}$ if for all $\epsilon > 0$ there is some N such that $d(p, p_n) < \epsilon$ for all $n > N$

A limit of a sequence is in fact *the* limit of the sequence.

Lemma 4.3. *A sequence has at most one limit.*

Proof. Say $\{p_i\}$ has two limits, p and p', then for any ϵ, we can take N such that $d(p_n, p) < \frac{\epsilon}{2}$ and $d(p_n, p') < \frac{\epsilon}{2}$ for $n > N$, but then $d(p, p') < 2\epsilon$. But if $p \neq p'$ then $d(p, p') > 0$, so we get a contradiction because we can take $\epsilon < d(p, p')$ $\square$

Definition 4.4. If there is a point $p = \lim \mu$, we say that the sequence μ *converges*. On the other hand, if there is no such p, we say that μ *diverges*.

Definition 4.5. A *subsequence* of a sequence $\{p_i\}$ is just a subset of that sequence, which we think of as a new sequence. Here is Wyatt's snooty definition:

Let $\rho : \mathbb{Z}^+ \to \mathbb{Z}^+$ be a strictly monotonically increasing map of sets. Then we say $\eta : \mathbb{Z}^+ \to X$ is a subsequence of μ if there is such a ρ such that

$$\begin{tikzcd} \mathbb{Z}^+ \arrow[r, "\eta"] \arrow[d, "\rho"] & X \\ \mathbb{Z}^+ \arrow[ur, "\mu"] & \end{tikzcd}$$

commutes.

We can generalize our notion of convergence of a sequence to an arbitrary topological space.

Definition 4.6. We say that a sequence $\{p_n\}$ converges to a point $p \in X$ if for every open set $U \ni p$, there is an N such that $\{p_n\}_{n \geq N}$ is contained in U. (That is, eventually the sequence lies in any open set that contains p.)

However, in this case we lose uniqueness of the limit point. For instance, in the cofinite topology or in the trivial topology (only open sets are empty and the entire set), we will find that most sequences converge to most points.

Remark 4.7. We can restore uniqueness to the topological limit by requiring our space X be *Hausdorff*. Roughly, this means that given any two points of X, there is a pair of disjoint open sets of $U, V \subset X$ such that $p \in U, q \in V$.

The notion of convergence is *not* intrinsic. Consider the sequence $\{\frac{1}{i}\}$ in $(0,1)$. This is divergent in $(0,1)$ but convergent in $\mathbb{R}$.

Theorem 4.8. *If $E \subset X$ is closed, then if all $p_n \in E$ and $p_n \to p$ in X, then $p \in E$. In fact, E is closed if and only if it contains every limit of every sequence inside itself.*

Proof. Say $p \notin E$ then $p \in E^c$ so we can take an open ball around p that does not meet E. Then the p_n will eventually enter that ball, contradicting the fact that all p_n are in E.
If E is not closed then E^c is not open, then there exists a point $p \in E^c$ such that every open ball around p meets E. If we let p_i be an arbitrary point of $B_{1/n}(p) \cap E$ this sequence will converge to p $\square$

Theorem 4.9. *If X, Y are metric spaces, then a map $f : X \to Y$ is continuous if and only if whenever $\{p_n\} \to p$ in X, $\{f(p_n)\} \to f(p)$ in Y. (Notationally, we will sometimes write $f_*\{p_n\} = \{f(p_n)\}$, and $f_*\mu$ for the general pushforward of a sequence μ.)*

Proof. First, suppose f is continuous, and we have $\{p_n\} \to p$. We want to show $\{f(p_n)\} \to f(p)$. Let $U \subset Y$ be any open set that contains $f(p)$. Then $f^{-1}(U)$ is open and contains p, so there is an N such that $\{p_n\}_{n \geq N} \in f^{-1}(U)$. But then for the same N, we have $\{f(p_n)\}_{n \geq N} \in U$, as desired.

The other direction is left as an exercise. *Hint*: Suppose f is not continuous at a point p, then construct a sequence that converges to p, but $\{f(p_n)\}$ does not converge to $f(p)$. $\square$

4.2. **Function Spaces.** In Euler's argument, quoted above, we noted some convergence problems when $x \geq 1$. However, Euler was clearly thinking of $x, x^2, x^3, ...$ as maps $\mathbb{R} \to \mathbb{R}$. While this doesn't fix the proof, it is interesting to consider. Let X, Y be metric spaces. We want to think of $\text{Maps}(X, Y)$ as a metric space. Recall to make this work earlier we required X finite or Y bounded to put a nice metric on it. However, it turns out there's a nice intermediate thing we can do, which is consider $\mathscr{B}(X, Y) = \{f \in \text{Maps}(X, Y) \mid f(X) \text{ is bounded}\}$.. In fact, for this we don't need to require X be a metric space, only that Y is. But for now, we will make this assumption. Then we define

$$d(f, g) = \sup_{x \in X} d_Y(f(x), g(x)).$$

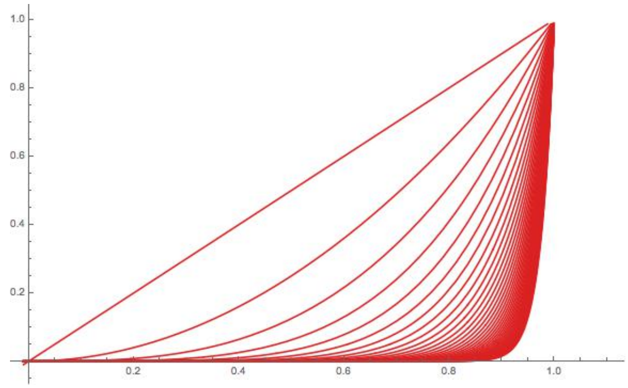
What does convergence mean here? Let $\mu : \mathbb{Z}^+ \to \mathscr{B}(X, Y)$. For all ϵ, we want to find an N such that the distance between any point in the image $\mu|_{n \geq N}$ and p is less than ϵ.

One way we might guess this to work is requiring that the maps converge pointwise–that is, $\{f_n(x)\} \to f(x)$ for all x. However, this isn't quite the same as our definition. We say that this is *pointwise convergence.* On the other hand, our definition required convergence within ϵ by some N for every point $x \in X$. It's quite conceivable that these two notions differ. (Exercise: find an example when the two are different.) In this case, we say that $\{f_n\} \to f$ is a *uniformly convergent* sequence. Thus, we will sometimes say that the metric we defined is called the *uniform metric*.

Note that our definition did not actually require that the f_n are bounded, though it was necessary for us to define our metric. This will be useful later, however.

5. Lecture 05 – 03 February 2017

5.1. **Uniform Limits.** Recall from last class, we considered $f_n : [0,1] \to [0,1]$, given by $f_n(x) = x^n$.



It's clear that the f_n converges pointwise (see definition below)–for every $x \neq 1$, when n is sufficiently large, we get close to 0, and $f_n(1) = 1$ for all n. Moreover, all f_n are continuous–however, it's clear that $\lim f_n$ is not continuous.

Definition 5.1. We say $f_n \to f$ converges pointwise if for every point x for every point ϵ there is an N such that $n \geq N$ implies $d(f_n(x), f(x)) < \epsilon$.

Definition 5.2. We say $f_n \to f$ converges uniformly if for all ϵ there is an N such that for $n \geq N$, we have for all points x that $d(f_n(x), f(x)) < \epsilon$. (In the case of bounded functions, we could simply say $f_n \to f$ in the sup metric.)

Remark 5.3. Because in $\mathscr{B}(X, Y)$, continuous functions are characterized by the fact that sequences of continuous functions converge to something continuous, we actually have that the set of continuous functions $C \subset \mathscr{B}(X, Y)$ is a closed set.

Remark 5.4. With the definition of the limit of a sequence, we can easily define the limit of a series, in topological space where addition makes sense. We calculate $\sum_1^\infty a_n$ by taking the limit of the partial sums $\lim_{N\to\infty} \sum_1^N a_n$. We can extend this to sums of functions. For instance, an important example is the Taylor series

$$\sum_{n=0}^{\infty} a_n x^n = \lim \sum_{n=0}^{N} a_n x^n.$$

Uniform convergence is much stronger than pointwise convergence, and we get nicer results. For instance, we have

Theorem 5.5. *If X is a metric space, $f_n : X \to Y$ are continuous, and $\{f_n\} \to f$ converges uniformly, then $f : X \to Y$ is continuous.*

Proof. Given x, x' "close," we want to make the argument that $f(x), f(x')$ are close. By the triangle inequality, we have for all n

$$d(f(x), f(x')) < d(f(x), f_n(x)) + d(f_n(x), f_n(x')) + d(f_n(x'), f(x')).$$

Since the f_n are uniformly convergent, we can take N such that for all $n > N$, $d(f_n, f) < \frac{\epsilon}{3}$. Pick one such f_n; since this function is continuous we can take a δ such that when $d(x, x') < \delta$, $d(f_n(x), f_n(x')) < \frac{\epsilon}{3}$. Then we have for all x, x' with $d(x, x') < \delta$

$$d(f(x), f(x') \leq d(f(x), f_n(x)) + d(f_n(x), f_n(x')) + d(f_n(x'), f(x)) = \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon$$

as desired. □

Remark 5.6.

Or points out that our requirements were too strong. We only needed uniform convergence on a collection of open sets that cover X. If X admitted only an infinite cover of open sets that worked, our proof would still hold. This is the notion of a *locally uniformly convergent sequence*.

Remark 5.7. Rudin takes the policy of using fractions of ϵ, so we end up with ϵ at the end. This is reasonably common, but utterly unnecessary. You can always delay the choice of what ϵ is, or simply prove that a distance is less than 3ϵ, for instance. Since ϵ is arbitrary, we can just throw away constants. Don't stress about ending with exactly ϵ in your proofs–it's usually much easier if you just start with ϵ and get something else nice as a result.

5.2. **Compactness.** There are a number of notions which, when first defined, can seem arbitrary. Compactness is one of those notions, and we'll be asking you to take on faith, for now, that it's actually a useful and interesting concept.

Anti-Definition 5.8. We will see very soon what compactness means, but let's start with something it does not mean: compactness does *not* mean closed and bounded. It is true that every compact space is closed and bounded, but the converse is not true. By Heine-Borel, the converse holds for subsets of $\mathbb{R}^n$.

Definition 5.9. Let X be a topological space, $K \subset X$ a subspace. We say K is *compact* if every open cover $\{U_i\}$ of K has a finite subcover.

This can be a bit confusing–as a refresher, here's what we mean by cover, subcover, etc:

Definition 5.10. A *cover* of a metric space X is a collection of subsets $\{U_i\}_{i \in I}$ such that $\cup_{i \in I} U_i = K$. We say a cover is *open* if all U_i are open.

A subcover is a subset $J \subset I$ such that $\cup_{i \in J} U_i = K$. We say such a subcover is finite if the index set J is a finite set. (NB: All of these equalities occur in the subspace topology on K, not on the entire set X. In general, it's ok as long as $\cup U_i \supset K$.)

Remark 5.11. Note that this definition requires that *every* open cover has a finite subcover, not just any particular cover. For instance, every set has a finite open cover, just by taking the set itself. However, this is not enough to verify compactness, since we might be able to take more complicated open covers.

In the context of metric spaces, we have

$$K = \cup_{x \in K} B_{\epsilon_x}(x)$$

is an open cover, for instance. This means that, for compact sets, we can always move down to finitely many of these balls. (Note this immediately implies boundedness of a compact space.)

Example 5.12. Though compactness can be really hard to verify, there are some easy cases.

- The empty set $\emptyset$ is compact, since we can always just take the empty subcover.
- K a finite collection of points always works, as at most we need one open set for each of our finite points.

> "It will take us a while before we can show any non-trivial examples...I am calling your examples trivial, sorry." –NE

Theorem 5.13. *If $K \subset X$ is compact, then K is closed and bounded.*

Proof. If K is empty, this is trivial. Otherwise, we can assume there is at least one point in K.

We remarked on the boundedness argument above. A slicker way to make that argument, however, is to fix a point $x \in X$, and consider the open cover

$$\{B_n(x)\}_{n \in \mathbb{Z}^+}.$$

Since K is compact, it has a finite subcover. This necessarily will have a maximal element, hence K is contained in a ball and therefore bounded.

Next, we want to show K is closed. Let $p \in X - K$, and take an open cover $\{G_m\}$, where $G_m = X - \overline{B}_{1/m}(p)$. Then for each $x \in K$, we have $d(x, p) > 0$, so G_m form an open cover of K. Then moving to a finite cover, it follows that there is some m such that $\overline{B}_{1/m}(p) \subset X - K$, which implies $B_{1/m}(p) \subset X - K$, hence $X - K$ is open. Thus by definition K is closed. $\square$

Remark 5.14. While it's not necessarily obvious, compactness is an intrinsic property to K–it doesn't matter how K is embedded into X, or what X is. This contrasts with our remark on being closed, which is not intrinsic.

6. LECTURE 06 – 06 FEBRUARY 2017

6.1. **Compactness Continued.** We want to show that compactness is intrinsic, in the sense that it doesn't depend on the space that we're embedded. This is a nice result, and generally means that maps on compact spaces behave nicely.

Theorem 6.1. *Compactness is an intrinsic notion. i.e., if K is compact, it is compact as a subset of X for all X.*

Proof. Say K is compact. We can consider $K \subset X$, and take an open cover $\{U_i\}$. The restriction of that open cover $\{U_i \cap K$ gives an open cover of K by open subsets of K, which has a finite subcover. THis corresponds to a finite subcover of K in X. □

A corollary of this is if $K \subset X$ is a compact subset, and $X \subset Y$ then K is a compact subset of Y.

Theorem 6.2. *A closed subset of a compact set is compact.*

Proof. Let $F \subset K$ be a closed subset of a compact space X. Let G_α be an open cover of F. Then since F is closed, we have that its complement F^C is open. Thus

$$K \subset F^C \cup \left(\bigcup_\alpha G_\alpha \right)$$

is an open cover of K. By assumption, there is therefore a finite subcover, which necessarily means we are left with finitely many G_α. Moreover, since $F^C \cap F = \emptyset$, we must have the remaining G_α cover F, so our open cover of F has a finite subcover, hence F is compact. □

Theorem 6.3. *The image of a compact set under a continuous map is continuous.*

Proof. Let $f : K \to Y$ be a continuous map with K compact. Then we want to show $\mathrm{Im}(f)$ is compact. Take an open cover of $\mathrm{Im}(f)$, $\{U_i\}$. Then by the continuity of f, $\{f^{-1}(U_i)\}$ is an open cover of K, so we can take a finite subcover. This corresponds to a finite subcover of $\{U_i\}$ on the image of f, so the image is continuous.

This theorem actually gives us a nice proof that compactness is intrinsic, see if you can figure it out. □

Remark 6.4. The above results were all shown without reference to the metric–in fact, our proofs all work for arbitrary topological spaces.

This last theorem is really useful once we prove that $[0, 1]$ is compact. That means that the image of any $f : [0, 1] \to \mathbb{R}$ is a compact (closed and bounded) set in $\mathbb{R}$, so it has a maximum. But first we must prove $[0, 1]$ compact.

6.2. **Sequential Compactness.** Let $\{p_n\}$ be a sequence in X. Recall we defined a *subsequence* as, intuitively, an infinite subset of the points of $\{p_n\}$.

Definition 6.5. A subspace $K \subset X$ is *sequentially compact* if for every sequence $\{p_n\}$ in K, there is a convergent subsequence. That is, there are $n_1, n_2, \ldots$ such that $\{p_{n_i}\} \to p$ converges in K.

Theorem 6.6. *If K is a metric space, then K is compact if and only if K is sequentially compact.*

One part of this is easy, one part requires several steps.

The forward direction (compact implies sequentially compact) is the easy one. For the backwards, we can prove this if we have a countable cover of our space. Without that, we will need to show that a sequentialy compact space is *separable*: for all ϵ a countable union of ϵ-balls cover K or equivalently there is a dense countable subset of K. We can use separability to turn any cover into a countable cover, and then into a finite cover.

Sequential comapctness is nice for lots of things. If K_1 and K_2 are metric spaces, then so is $K_1 \times K_2$. We claim the product is also compact. This is true in a general topological setting, but point-set topology is no fun. It is also hard to show in terms of open covers, but easily proved with sequential compactness.

Lemma 6.7. *A product of compact spaces is compact*

Proof. Let $\{(p_n, q_n)\}$ be a sequence in $K_1 \times K_2$. Then $\{p_n\}$ has a convergent subsequence in K_1, call it $\{p_{n_i}\}$. Also $\{q_{n_i}\}$ has a convergent subsepquence in K_2, call it $\{q_{n_{i_j}}\}$ Then $\{(p_{n_{i_j}}, q_{n_{i_j}})\}$ converges in $K_1 \times K_2$ □

There's another way to think about a subsequence converging at $p \in K$, which can be quite useful for our proof of the theorem.

Lemma 6.8. *Given a sequence $\{p_n\}$ in K, p is the limit of a subsequence if and only if for every $\epsilon > 0$, there are infinitely many n such that $d(p_n, p) < \epsilon$. Equivalently, p is the limit of a subsequence if and only if for every open set $U \ni p$, there are infinitely many n such that $p_n \in U$.*

Corollary 6.9. *p is* not *the limit point of a subsequence of $\{p_n\}$ if and only if there is an open set $U \ni p$ such that $U \cap \{p_n\}$ is finite.*

The proof of these is easy. Now we return to the main theorem

Lemma 6.10. *Compactness implies sequential compactness*

Proof. Assume X is not sequentially compact. Take a sequence $\{p_n\}$ without a convergent subsequence. Construct the following open cover: at every point $p \in X$ if $p \neq p_i$ for some i take an open ball around p that contains no p_k. If $p = p_i$ for some i, take an open ball around p that excludes all other points of $\{p_n\}$. We can do this because it is guaranteed that no subsequence converges to p. This is an open cover that does not admit a finite subcover, because every point p_i in the sequence requires a different open set in the cover, so infinite sets are required. □

Proposition 6.11. *Say K is sequentially compact and equal to a countable union of open G_i, then this countable cover admits a finite subcover*

Proof. Supose not, then construct a sequence as follows. Pick $p_1 \notin G_1$, and $p_2 \not{i}n G_1 \cup G_2$, and so on. If no finite collection covers, we can make an infinite sequence this way. But this sequence has a convergent subsequence which convegres to p. Then $p \notin G_i$ for all i, contradicting the fact that $\{G_i\}$ is a cover □

In order to reduce to the above case, we will now show that K sequentially compact implies K is separable–that is, there is a countable dense subset of K. This is argued in the following proposition. We will then complete this argument.

Proposition 6.12. *If K is sequentially compact, then it is separable.*

Proof. When K is finite, this is clear.

Note sequential compactness implies boundedness, since if not, take $p_0, p_1, ...$ such that $d(p_0, p_n) \geq n$, and then $\{p_n\}$ obvious doesn't have a convergent subsequence. Now, choose p_1, and let $R_1 = \sup_{q \in K} d(p_1, q)$. Now, choose p_2 such that $d(p_1, p_2) \geq \frac{1}{2} R_1$. Let $R_2 = \sup_{q \in K} \min(d(p_1, q), d(p_2, q))$, and choose a point p_3 that is at least $R_2/2$ away from both p_1, p_2. Then we claim this sequence is dense, which we will argue next class. $\square$

7. Lecture 07 – 08 February 2017

7.1. **Sequential Compactness Continued.** We have been trying to prove that a set is sequentially compact if and only if it is compact. Last time we proved Compactness implies sequential compactness, and we are in the middle of proving the other direction. Let's refresh the chain of reasoning.

- If X is sequentially compact, every countable cover has a finite subcover (Proved in last lecture)
- If X is sequentially compact, it is *separable*, i.e. it has a countable dense subset
- If X is separable every open cover has a countable subcover

Upon proving all of these propositions, we will have proven that a sequentially compact set is compact.

Proposition 7.1. *A sequentially compact space is separable.*

Proof. We already argued that a sequentially compact space must be bounded. If our space has only finitely many points, the statement is vacuous. Now consider $p_1, ..., p_n$, constructed as we did last time such that our points are, roughly speaking, "as far away from one another as possible." Specifically, we made the following construction:

Choose p_1, and let $R_1 = \sup_{q \in K} d(p_1, q)$. Now, choose p_2 such that $d(p_1, p_2) \geq \frac{1}{2}R_1$. Let $R_2 = \sup_{q \in K} \min(d(p_1, q), d(p_2, q))$, and choose a point p_3 that is at least $R_2/2$ away from both p_1, p_2. Continue doing this inductively to construct p_n.

Note $R_1 \geq R_2 \geq R_3 \geq ....$ Now, it's sufficient to show that $R_n \to 0$, as this means our points $\{p_n\}$ get arbitrarily close to every point in our metric space. If not, then there's some $\epsilon > 0$ such that $d(p_m, p_n) \geq \frac{\epsilon}{2}$, hence there's no convergent subsequence amongst the p_i, but that's a contradiction. $\square$

Lemma 7.2. *If X is separable, then any open cover of X has a* countable *subcover.*

Proof. Let $\{p_n\}$ be a dense sequence, let $X = \cup_{s \in S} U_s$ be a cover, and Let $r_k = \frac{1}{k}$. Look at $\{B_{r_k}(p_i)\}_{i,k>0}$ This is totally a cover of X. WE can use it to get a subcover of the $\{U_s\}$, for each p_i we can take a small enough r_k so that $B_{r_k}(p_i) \subset U_{s'}$. Form a collection of all balls around p_i with rational radius that are contained in some U_s. This is a countable collection. It remains to show it is a cover. Pick any $p \in X$, we have $p \in U_s$ so $B_r(p) \subset U_s$, but $B_r(p) \ni p_i$ and we can take a ball around p_i with rational radius that contains p and is contained in U_s. This gives that the countable collection of rational balls is a cover. If for each ball we pick a U_s containing it, we get a countable subcover of the original cover. $\square$

This concludes the proof. Now we know lots about sequentially compact sets, but still no examples.

7.2. **Cauchy Sequences.** We have previously seen that the notion of convergence is not intrinsic. Let $E = (0, 1)$ and consider the sequence $\{\frac{1}{i}\}$. This converges in $\mathbb{R}$, but not in E To get around this, we use a related notion from the famous mathematician Cauchy

Definition 7.3. A sequence $\{p_n\}$ is a cauchy sequence if for any $\epsilon > 0$ there is some integer $N > 0$ so that for all $n, m > N$ $d(p_n, p_m) < \epsilon$

This is very similar to convergence, but it is intrinsic. If a sequence in E is *cauchy*, and $E \subset X$, the sequence is trivially cauchy in X.

Corollary 7.4. *A convergent sequence is a cauchy sequence*

Proof. Very Easy! Try it yourself. □

The converse does not necessarily hold–it only holds in a *complete* topological space, to be discussed later. Recall that *totally bounded* means that for every $\epsilon > 0$ we can cover X in finitely many ϵ-balls.

Theorem 7.5. *If X is totally bounded, if and only if every sequence in X has a cauchy subsequence*

Proof. If every sequence in X has a cauchy subsequence, create a sequence of $p_i \in X$ like we did at the start of the lecture, with every p_i a distance of ϵ away from all other points. If this sequence was infinite, it would have a cauchy subsequence, contradicting the fact that our points are spaced out.

In the other direction, let $\epsilon_n = \frac{1}{n}$, each ϵ_i correspond to finitely many points whose ϵ-balls cover the space. Let $\{q_i\}$ be a sequence. There are finitely many balls of radius 1, so there is some ball with infinitely many points of the sequence. So we get a subsequence $\{q_i^{(1)}\}$. If we then look at balls of radius $\frac{1}{2}$ within the 1-ball we picked, we find one must have infinitely many points of $\{q_i^{(1)}\}$, and so we consturct a subsequence of this new sequence, calling it $\{q_i^{(2)}\}$. This gives sequence of subsequences of $\{q_i\}$ for every ϵ_n. We look at the sequence $q_1^{(1)}, q_2^{(2)}, ...$ this is surely infinite, but also cauchy, because all points in the sequence beyond the kth index are of distance at most $\frac{2}{k}$ away from each other. □

Definition 7.6. We say X is *complete* if every Cauchy sequence in X converges in X.

Remark 7.7. We have shown that every totally bounded, complete space is compact. In fact, the converse is also true. We already showed that compact spaces are totally bounded, and completeness follows by having your Cauchy sequence, taking a convergent subsequence (since it's sequentially compact), and it follows immediately that the Cauchy sequence converges to p.

Lemma 7.8. *If X, Y are complete, then $X \times Y$ is complete.*

Proof. The main argument we need to make is that (p_n, q_n) is Cauchy if and only if p_n, q_n are Cauchy in X, Y respectively. Then we've already shown that this implies that if $p_n \to p, q_n \to q$, then $(p_n, q_n) \to (p, q)$. □

8. Lecture 08 – 10 February 2017

"Professor, how much can you bench press?" "Less than has been reported, I am over 50!"

8.1. **Compactness victory lap.** We have proved that the following are equivalent properties for a metric space

$$\text{Compactness} \iff \text{Sequential Compactness} \iff \text{Complete and Totally Bounded.}$$

Now we have the tools necessary to show that some nonfinite set is compact!

Theorem 8.1. $\mathbb{R}$ *is complete.*

Proof. Whether this is definition or theorem depends on how you construct the real numbers. We will address this fully later. □

Theorem 8.2. (Heine-Borel) *A subset* $K \subset \mathbb{R}$ *is compact if and only if it is closed and bounded.*

Proof. We already know that compactness implies being closed and bounded.

Since $\mathbb{R}$ is complete, it follows that K is complete since it's closed in a complete space. Since K is closed and bounded, it is further contained in $B_M(0)$ for some M. Then we can consider $\mathbb{Z}\epsilon \cap B_m(0)$. This will be finite set, and also an ϵ-net, hence K is totally bounded. □

If we use the sup metric, it is not difficult to generalize Heine-Borel to $\mathbb{R}^n$ to get a characterization of all compact sets in $\mathbb{R}^n$ as just closed and bounded sets. Specifically, using the sup metric, we can just take the cartesian product of our ϵ-net from the above theorem, which gives us an ϵ-net for K. Since the Euclidean metric is equivalent to the sup metric, this gives us the argument in our normal Euclidean metric.

Remark 8.3. We have now successfully introduced a notion that seemed abstract and unmotivated, then showed it was equivalent to a notion that is so easily defined as to appear useless.

Remark 8.4. This gives rise to what is sometimes called the Heine-Borel Property: X is Heine-Borel if K closed and bounded in X implies K is compact.

8.2. **Fixing Some Algebra: Loose Ends from Math 55a.** We're now getting to the point we can repair some holes from Math 55a. For instance, last semester we used the fundamental theorem of algebra to conclude that any linear map of complex vector spaces has an eigenvalue.

Theorem 8.5. (Fundamental Theorem of Algebra) *Let* $P \in \mathbb{C}[z]$ *be a polynomial of degree greater than 0. Then* P *has a root.*

Proof. Assume not. We showed last semester that we can find some R so that if $|z| > R$ then $|P(z)| > \alpha|z|^{\deg P}$. The function $z \to |P(z)|$ is continuous, so it attains a minimum on a compact set. Consider $P(0)$; if this is 0 we are done. We can then pick R sufficiently large such that $|z| > R$ implies $|P(z))| > |P(0)|$. Then the function takes a minimum on $\overline{B_R}(z)$. We can assume without loss of generality that the minimum is attained at 0 0: if the minimum occurs at z_0, then we can consider the function $P(z - z_0)$ instead. Now we have a polynomial whose absolute value takes a minimum at 0. Say this minimum is nonzero. We can express our polynomial $P(z) = \alpha_0 + \alpha_1(z - z_0) + ...$ Then consider a small ball around z_0. We have $\alpha_n(z - z_0)^n$ takes all values on a circle of radius ϵ of the form $|\alpha_n|\epsilon^n$. Then we can find an argument that will imply one of these direction will decrease P, contradicting our assumption that the minimum was achieved at z_0.

This is a hard way to go about making the argument. It's far easier once we have complex analysis, in particular, Liouville's theorem. □

We can also recall the spectral theorem. Consider the unit sphere $S^{n-1} \subset \mathbb{R}^n$. Given a Hermitian operator, we showed that once we had the orthonormal eigenbasis,

$$\max_{S^{n-1}} \langle v, T_v \rangle = \lambda_{\max}.$$

We used this to find λ, but we didn't actually know the supremum was actually attained. However, S^{n-1} is closed and bounded (boundedness is clear. To see it's closed, consider $d(0, -) : \mathbb{R}^n \to \mathbb{R}$ is a continuous map, and the inverse image of 1 under this map is S^{n-1}, hence it's the inverse image of a closed set, hence is closed). Since $\mathbb{R}^n$ has the Heine-Borel property, this implies S^{n-1} is compact. Thus $v \mapsto \langle v, Tv \rangle$ is a continuous function on a compact set, so we acheive the supremum.

8.3. **Compactness and function spaces.** We previously discussed the space of bounded functions from X to Y, called $\mathcal{B}(X, Y)$, and the space of continuous bounded functions $\mathcal{C}(X, Y)$. Perhaps surprisingly, $\mathcal{C}(X, Y)$ is complete if Y is.

Lemma 8.6. *$\mathscr{C}(X, Y)$ is complete if Y is.*

Proof. Let f_n be a Cauchy sequence. Then for all x, $f_n(x)$ is Cauchy, hence $f_n(x) \to f(x)$. If f_n are continuous, then f is continuous thanks to the boundedness condition. □

If X is compact, then all continuous maps $f : X \to Y$ are bounded, so we can actually drop the bounded hypothesis. Then $\mathscr{C}(X, Y)$ is just the set of continuous maps $X \to Y$. What's more, we can replace the sup metric with a max metric, because $x \mapsto d_Y(f(x), g(x))$ is continuous, so it attains a maximum.

Finally, our earlier distinction of continuous and uniformly continuous vanishes.

Proposition 8.7. *A function $f : X \to Y$ with X compact is continuous if and only if it is uniformly continuous.*

Proof. We will use the following, which has been assigned as homework

Lemma 8.8. (Lebesgue (le-bayg) Covering Lemma) *Let $\{U_\alpha\}$ be an open cover of a compact set. There is some $r > 0$ such that for all $x \in X$ $B_r(x)$ is contained in some U_α*

If f is continuous then for a fixed $\epsilon > 0$ we can cover X by taking a ball at each point with a radius δ_x satisfying continuity constraints. Then we use the above metric to get a δ that works for every point. In particular, consider all the $\epsilon/2$ balls in Y, which form an open cover. Then $f^{-1}(B_{\epsilon/2}(y))$ for all $y \in Y$ is an open cover of X. Taking a finite subcover, we get a sufficiently small δ to give continuity everywhere, simply by taking the minimum over a finite set, hence uniform continuity. The other direction of the equivalence is obvious. □

8.4. **Completions.** For a number like $\pi = 3.1415926535...$ we can't calculate it exactly, but instead we take approximations.

Remark 8.9. $e^{\pi\sqrt{67}/3}$ is approximately 5280. It takes quite a bit of computation to verify whether it's smaller, larger, or exactly equal.

It is not totally precise to say that π is a decimal expansion, because decimal expansions are nonunique. For instance, we might all be in Math 54.999..., as $54.999... = 55$.

Cauchy sequences of rational numbers do not quite form a metric space, because different cauchy sequences can have the same limit. We can fix this by using the construction from the homework, where we quotient out by an equivalence relation that relates points of distance 0 to each other. In fact, this is one possible way to construct the real numbers, as we shall see soon.

9. Lecture 09 – 13 February 2017

Last lecture we discussed the notion of a complete space and noted that $\mathbb{R}$ was the completion of $\mathbb{Q}$. THere are many ways of showing this, because $\mathbb{R}$ is interesting for lots of reasons beyond being a completion. It also has an order. For this reason we can talk about completions in $\mathbb{R}$ using $\limsup$ and $\liminf$ of a bounded sequence. The precise definiton is $\liminf = \sup \inf_{m \geq n}\{S_m\}$. Note $\inf_{m\geq n}\{S_m\}$ is an increasing sequence, so it makes sense to takes its sup. We can similarly define the $\limsup$.

Definition 9.1. The $\limsup$ of a sequence $\{f_n\}$ is

$$\lim_{n\to\infty} \sup_{m\geq n} f_m = \inf \sup_{m \geq n}\{S_m\}.$$

That is, it's the supremum over the entire set, with finitely many exceptions.

A seqeunce converges if and only if both $\lim inf$ and $\lim sup$ are the same. If you only knew these two things existed, you could show that for a cauchy sequence they both existed, and were withing ϵ of each other

Back in the first problem set, we showed that, given a space and a distance function satisfying everything but that distinct points have distance greater than 0, we can form a metric space by quotienting out by the relation $x \sim y$ if $d(x, y) = 0$.

Let X be a metric space, and X^* its completion–that is, X^* is complete, and there is an isometry $X \hookrightarrow X^*$. In fact, we can define the completion of a space via the universal property

$$\begin{tikzcd} X \arrow[r] \arrow[dr] & X^* \arrow[d, dashed] \\ & A \end{tikzcd}$$

given any map $X \to A$ of metric spaces, with A complete, it factors uniquely through X^*. Note this also implies that X^* is unique, if it exists. However, we do have to show that it exists.

Theorem 9.2. *Let X be any metric space. Then there is a complete metric space X^* with a map $\iota : X \to X^*$ with ι an isometry such that the image of $\iota(X)$ is dense in X^*, satisfying the above universal property.*

Proof. Uniqueness of the map turns out to be easy–an isometry is uniformly continuous, so we need only define it on a dense set, and we know what it must be on X, which is dense. Thus we should really just be looking for X^* that is both complete and has $X \subset X^*$ a dense subset.

The construction is a bit harder. Consider

$$X^* = \{\text{Cauchy Sequences on } X\}/\ ,$$

where $\sim$ relates sequences that are distance 0 from one another. Here we've set the distance function $d(\{p_n\}\{q_n\}) = \lim_{i\to\infty} d_X(p_i, q_i)$. This distance function is well defined because it is a Cauchy sequence in $\mathbb{R}$, and it turns out $\mathbb{R}$ is complete–though we haven't proved it yet. It is not hard to check that we have all the properties of a metric space except positivity. We then quotient out by identifying all cauchy sequences that are distance 0 apart.

We get our embedding of X by $\iota(x) = \{x, x, x, ...\}$ the constant sequence all of whose terms are x. This is dense. To see this, take a Cauchy sequence, and go far enough down that all terms are distance ϵ away from each other. Pick an arbitray term x' in this range, this is an element of X, the distance between the cauchy sequence and a constant sequence of x' is less than ϵ

Now we need to show completeness. Take a sequence in X^*, x_1^*, x_2^*, each "point" is an equivalence class of sequences, and we can take a represenntative element. Then we get the following

$$\begin{gathered} x_1^{(1)}, x_2^{(1)}, x_3^{(1)}, x_4^{(1)}, \ldots \\ x_1^{(2)}, x_2^{(2)}, x_3^{(2)}, x_4^{(2)}, \ldots \\ x_1^{(3)}, x_2^{(3)}, x_3^{(3)}, x_4^{(3)}, \ldots \\ x_1^{(4)}, x_2^{(4)}, x_3^{(4)}, x_4^{(4)}, \ldots \end{gathered}$$

Taking a diagonal almost works, but doesn't quite because we can't control how large our index needs to be in each sequence. For instance, we could in the nth sequence the nth entry to n without affecting convergence of any of the sequences. To get around this set $\epsilon_k = \frac{1}{k}$.For each x_k^*, there is an N such that $m, n \geq N$ implies $d(x_k^m, x_k^n) < \frac{1}{k}$. Then we can consider the sequence $x^* = \{x_1^{N_1}, x_2^{N_2}, x_3^{N_3}, ...\}$ is a Cauchy sequence, hence an element of X^*, and moreover, by construction is equal to $\lim x_i^*$. (Homework: verify the details of this argument. For instance, verify the sequence we used to define x^* is in fact a Cauchy sequence, and is the limit we wanted it to be. Moreover, verify that it satisfies our universal property.) □

Since this is canonical, we know that any complete set containing X as a dense subsapce is X^*. So if we take a subspace E of a complete set Y, the completion E^* is just the closure of E, since we know this is a complete set with E as a dense subset. Indeed, it's not hard to check that the closure either satisfies the universal property we gave above, or aligns with our construction of the completion.

9.1. **The N-adic numbers.**

> "This is one of those cases where people tried to construct a pathological example of a space. Then, many years later, it turns out it's very useful."

In the real numbers we approximate any real number by a sequence of rational numbers– we say two decimal expansions are close if they agree to multiple places to the right of the decimal. For instance, 3.1415 is close to 3.147 because of the shared .14. However, we could also do the reverse. We want to formalize the idea that numbers are "close" if they agree on multiple points to the left of the deicmal when expanded in base N. For instance, in base 10, we might say 625 is close to 231625.

If $x \equiv x' \mod N^k$, and are not equivalent mod N^{k+1}, we set $d(x, x') = N^{-k}$ for the N-adic norm. This is clearly symmetric, and 0 if and only if $x = x'$. In fact, we have something stronger than the triangle inequality, which is $d(x,z) \leq \max(d(x,y), d(y,z))$. We can extend this over the rationals by, for $q, q' \in \mathbb{Q}$, considering $|q - q'| = N^k \frac{r}{s}$, where r, s have no factors of N.

Example 9.3. This gives some strange results–for instance, in the 10-adics, we have

$$1 + 10 + 100 + ... = \frac{1}{1 - 10} = -\frac{1}{9}.$$

Because the metric has weird properties, we end up able to sum series that seem very "wrong." So, for instance, Euler's trick we talked about a few lectures ago could actually work, just by moving to a different metric.

Remember the numbers $0, 1, ..., N - 1$ are all distance 1 from each other. Then N is close to 0, $N + 1$ close to 1, etc. Then N^2 is even closer to 0, $N^2 + 1$ is even closer to 1, and so on. This is actually a sort of analogue of the Cantor set: recall the Cantor set comes from taking the unit interval, throwing out the middle third, then going to your remaining intervals, throw out their middle thirds, and so on. The 2-adic metric looks a lot like this, in that we keep getting smaller subdivisions that correspond to moving to higher powers of 2.

Remark 9.4. The 2-adics are also the relevant metric in music. The correspondences between notes should follow in powers of 2. (Maybe? I'm not a musician, idk.)

Remark 9.5. We will often let $\mathbb{Z}_n$ denote the integers with the n-adic norm. This is why we avoid this notation when talking about the integers mod n, preferring $\mathbb{Z}/n$ or $\mathbb{Z}/n\mathbb{Z}$.

Division in the n-adics is a bit problematic except when n is prime, so we do usually prefer this case.

Part 2
Calculus of a Single Variable

10. Lecture 10 – 15 February 2017

10.1. **A historical aside.** Newton lived 1642-1726/27, and Leibniz lived 1646-1716. George Berkeley lived 1685-1753, and Weierstrass 1815-97. Newton's formulation of calculus set

$$f'(x) = \frac{f(x+\Delta x) - f(x)}{\Delta x}|_{\Delta x = 0}.$$

This evolved eventually to Weierstrass' formulation

$$f'(x) = \lim_{\substack{x' \to x \\ x' \neq x}} \frac{f(x') - f(x)}{x' - x}.$$

Calculus started a couple of centuries before ϵ and δ arguments became standard, which is why we have all of these various formulations of the subject. Regardless of who discovered it, calculus was first formulated in the late 17th century, motivated, at least for Newton, by interest in physics, with questions like "What is the speed of a particle whose position at time x if $f(x)$?"

While Newton's formulation was used by many mathematicians, it was a relatively esoteric art. Weierstrass' limit definition is generally credited with reformulating it such that it can now be taught even in high schools. This followed Barkeley's criticism of taking $\Delta x = 0$ as a "ghost of a departed quantity.

We now say

Definition 10.1. A function f is differentiable at a point x if $\lim_{h \to 0} \frac{f(x+h)-f(x)}{h}$ exists, if it exists, this value is called the derivative. A function is differentiable on a subset if it is differentiable at every point in that subset, in this case its derivative is a function on the subset, which we call f'

By looking at the above definition, we need f to be a function on a place where you can add and multiply, so it must be defined on an open subset of a field, which for the time being will be $\mathbb{R}$. We can also talk about a function being diferentiable on a closed or half closed interval, by taking a one sided limit, but we won't do this for the most part.

In addition to working over the reals, we can work over, say, the p-adics or the complex numbers. We won't say much about the p-adics, but we will cover complex analysis, which turns out to be much nicer than real analysis. The reason for this is that having a complex number as a derivative imposes a major restriction on functrion with domain $\mathbb{C}$ (or domain $\mathbb{R}^2$)

While the domain of f will be $\mathbb{R}$ for now, the target space needs to be somewhere where I can add elements and divide by real numbers and take limits. So f is a map to a vector space with a distance function coming from a norm. Since norms in $\mathbb{R}^n$ are equivalent, they give rise to the same topology, the same notion of limit, and thus the same notion of differentiability.

Another way of thinking of differentiability is that a function has a derivative at p if it has an affine linear aproximation, i.e $f(t) = f(x) + f'(x)(t-x) + E(x)$ where E is some error function satisfying the following: for all $\epsilon > 0$ there is some δ $E(t) < \epsilon|t-x|$ when $|x-t| < \delta$. In this case we say $E(t) = o(|t-x|)$.

The 'o' notation is very useful and we will define it in full generality.

Let $f : X \to \mathbb{R}^{\geq 0}$, and g take X to a normed vector space.

Definition 10.2. We say $g = O(f)$ (read: g is big-oh of f) if there is $C \in \mathbb{R}$ such that, for all $x \in X$, $|g(x)| \leq Cf(x)$.

Note in the case of $\mathbb{R}^n$, C might depend on our choice of norm. However, for any equivalent norms, we can always find appropriate C, should it exist for any of the norms.

Definition 10.3. We say $g = o(f)$ (read: g is little-oh of f) if f, g are as above, and for every $\epsilon > 0$, eventually $|g(x)| \leq \epsilon f(x)$, where "eventually" means as we approach x. (In standard practice, this means as we approach infinity, but we will be using this formulation here.)

Note, of course, that $g = o(f)$ automatically implies $g = O(f)$. Little o is a much stronger condition than big O is. Also the sum of two functions that are little o (or big O) of f remains so, similarly you can multiply by scalars. Finally $O(f_1)O(f_2) = O(f_1 f_2)$.

Example 10.4. When $f = 1$ we see o and O yield familiar concepts

- $O(1)$ means bounded
- $o(1)$ means approaches 0

Theorem 10.5. *When f, g are functions from the functions from $\mathbb{R}$ to a place where you can multiply (one lands in $\mathbb{R}$ one lands in a vector space, for instance), and both f' and g' exist. $(fg)' = f'g + fg'$*

Proof. Consider the error term in our linear approximation $\epsilon(t) = o(|t-x|)$. Then

$$\begin{aligned} fg(t) &= (f(x) + f'(x)(t-x) + o(|t-x|)(g(x) + g'(x)(t-x) + o(|t-x|)) \\ &= f(x)g(x) + (f'(x)g(x) + g'(x)f(x))(t-x) + o(|t-x|) + f'(x)g'(x)(t-x)^2. \end{aligned}$$

Then we just need to show $(t-x)^2$ is $o(t-x)$ as $t \to x$, but this follows just by taking $\delta = \epsilon$. $\square$

Theorem 10.6. *A function being differentiable at x implies it is continuous at x*

Proof. If we look at $f(t) = f(x) + f'(t)(t - x_+ o(|t-x|)$ we see the right hand side approaches 0 as $t \to x$ $\square$

We will mostly jump over proving the rest of the derivative rules from calculus. Look into Rudin 5.3 to see formulas for derivatives of $f+g$, $f-g$ and $\frac{f}{g}$ If you forgot the HI-DE-HO formula for quotient rule, you can use $f = \frac{f}{g}g$.

11. Lecture 11 – 17 February 2017

Recall, last time we defined the derivative, and showed it had the following properties:

- $(f+g)' = f' + g'$
- $(fg)' = f'g + fg'$
- $(f/g)' = (f \cdot 1/g)'$

For the last one, we do still need to compute $(1/g)'$, and of course, must require $g \neq 0$. Let $f : [a,b] \to \mathbb{R}$, $x \in [a,b]$, be differentiable on a neighborhood of x, and let $g : N_r(f(x)) \to \mathbb{R}$ be another map, differentiable at $f(x)$. We'd like to have a way to describe the derivative of $g \circ f$ at x.

Theorem 11.1. *Let $h = g \circ f$ where for f, g functions from $\mathbb{R}$ to itself (techinically, g can be to a vector space V), both differentiable. Then where it is defined, h is differentiable, with derivative*

$$h'(x) = g' \circ f(x) \cdot f'(x)$$

Proof. We know $f(t) = f(x) + f'(x)(t-x) + o(|t-x|)$ which is the same as $f(x) + (f'(x) + u(t))(t-x)$ where $u(t)$ approaches 0 as t approaches x. We can then make u continuous by saying $u(x) = 0$ Then f is differentiable exactly when there is such a u. For t near x we have $f(t) - f(x) = (t-x)(f'(x) + u(t)$. Similarly, to compute $g(s)$ for s near y, we have $g(s) - g(y) = (s-y)(g'(y) + v(s))$. Now we can set $h(t) - h(x) = g(f(t)) - g(f(x)$, letting $y = f(x)$, and get

$$\begin{aligned}
g(f(t)) - g(y) &= (f(t) - y)(g'(y) - v(f(t))) \\
&= (f(t) - y)(g'(y) - v(f(t))) \\
&= (f(t) - y)(g'(y) - v(f(t))) \\
&= (t-x)(f'(x) + u(t))(g'(y) - v(f(t))) \\
&= (t-x)(g'(y)f'(x) - f'(x)v(f(t)) + u(t)g(y) - u(t)v(f(t))
\end{aligned}$$

So we take $u' = f'(x)v(f(t)) + u(t)g(y) - u(t)v(f(t))$ and need to show it approaches 0, which is clear since it does in every term. ☐

11.1. Some Linear Algebra Remarks.

Note that $(f+g)' = f' + g'$ is really half of what we need for differentiation to be a linear map. Indeed, since for k constant, $k' = 0$, it also follows that $(kf)' = kf'$. Thus $d : \mathrm{DiffFunct}_{[a,b]} \to \mathrm{Funct}_{[a,b]}$ is a linear map from differentiable functions on the interval $[a,b]$ to functions on the interval $[a,b]$. The image of this is quite hard to describe, though possible. The kernel is much simpler to describe: it is exactly the constant functions, but proving this is surprisingly nontrivial.

Definition 11.2. A function f from a space X to $\mathbb{R}$ has a global maximum (resp. minimum) at the point where it achieves its maximum (resp. minimum). It has a local maximum (resp. minimum) at a point p if there is an open neighborhood U of p where $f|_U$ takes a global maximum at p. An extremum is a point that is either a local minimum or a local maximum.

Lemma 11.3. *The derivative of a function f at a local extremum (if it exists) is 0.*

Proof. We know $f(t) = f(x) + (t - x)(f'(x) + u(t))$, with $u(x) = 0$. Assume WLOG x is a local maximum. Then $f(t) - f(x) \leq 0$. Then it follows that $(t - x)(f'(x) + u(t)) \leq 0$ for t near x. If $f'(x) \neq 0$, then we can change the sign by changing the sign of $t - x$, as $u(t)$ is small at x, and this is a contradiction. Thus $f'(x) = 0$. $\square$

Indeed, this argument nicely generalizes to all the partial derivatives of a differentiable map of several variables.

Theorem 11.4. (Rolle's Theorem) *If $f : [a, b] \to \mathbb{R}$ is continuous, differentiable on (a, b), and $f(a) = f(b)$, then there is $x \in (a, b)$ such that $f'(x) = 0$.*

Proof. The function is continuous on a compact set, so it has a global maximum and a global minimum. If either of these is an interior point, we're done. Otherwise, the map must be the constant map, hence has 0 derivative everywhere. $\square$

Corollary 11.5. (Mean Value Theorem) *Given a map $f : [a, b] \to \mathbb{R}$, there is a point x on the interior such that*

$$f'(x) = \frac{f(b) - f(a)}{b - a}.$$

Proof. Consider the linear map $g(t) = f(a) + (t - a)\frac{f(b)-f(a)}{b-a}$. Then we can apply Rolle's theorem to $f - g$, and the result follows. $\square$

Now, suppose f has 0 derivative everywhere, but is not the constant function on $[a, b]$. Then it achieves either a maximum or a minimum at a point not equal to a, call it x, then apply the MVT to f on the interval $[a, x]$ and we get a contradiction. Thus a function is constant if and only if its derivative is 0.

Definition 11.6. We say f is increasing if $a < b$ implies $f(a) < f(b)$ for a, b in the domain of f. Note, this is holds if and only if $f'(x) > 0$ where defined.

Remark 11.7. The mean value theorem, as stated, does not hold in higher dimensions. For instance, the map $t \mapsto (\cos t, \sin t)$ on $[0, 2\pi]$ has the same value at the two endpoints, but 0 derivative nowhere.

12. LECTURE 12 – 22 FEBRUARY 2017

12.1. **Linear Operators.** We can combine the notion of norm on a vector space with linear operators to get a notion of a bounded linear operator. A linear operator $T : V \to W$, with both spaces normed norm $||T||$ when it multiplies the norm on v by a bounded factor, i.e. there is some M so that $||Tv|| \leq M||v||$ for all v. We then define $||T|| = \inf M = \sup_{||v||=1} ||Tv||$. In this case we say T is bounded.

Theorem 12.1. *The following are equivalent*

- *T is continuous*
- *T is continuous at 0*
- *T is uniformly continuous*
- *T is bounded*

The equivalence is fairly easy to verify. This gives us a normed vector space $\mathrm{Hom}_{\mathrm{cont}}(V, W)$. Taking $W = \mathbb{R}$ actually gives us the continuous dual, as a special case.

If you have a map $f : I \to V$ then we want to say the function is approxiamted at any point by its derivative. When $V = \mathbb{R}$ we have the mean value theorem, there is some $x \in I$ where $f'(x) = \frac{f(b)-f(a)}{b-a}$. We can no longer say that for arbitrary V, but we can say

$$||f(b) - f(a)|| \leq |b-a| \max_x ||f'(x)||$$

This makes intuitive sense, If a car never drives faster than 60mph it can't be more than 60 miles away from where is started after an hour.

We can observe this by composing f with an element in V^*, where we have $(\eta f)'(x) = \eta(f'(x))$ (this uses uniform continuity, thanks RP!). We now have a map $I \to \mathbb{R}$ and we get the existence of an x where $\frac{|\eta f(b) - \eta f(a)|}{b-a} = \eta f'(x)$ and $|\eta f(b) - \eta f(a)| \leq ||\eta|| \cdot |f(b) - f(a)|$ To make this interesting, we need to find the 'right' η, one where this inequality is an equality so we can cancel $||\eta||$. So for a given v we want an η so $|\eta(v)| = ||\eta|| \cdot ||v||$. This is possible in general, but hard to show, but easy to show for an inner product space, where we can take the dual of v using the inner product.

12.2. **Taylor's Theorem.** If f' is itself differentiable, then we say the derivative of f' is the second derivative of f, denoted f''. In general, for the nth derivative, we will write $f^{(n)}$, though this can be used for other things, so it's worth stating what you mean by this notation, should you use it. We could also write, for instance, $D^n f$, where D denotes the linear map that takes a differentiable function to its derivative.

We'd like to generalize something akin to the mean value theorem to higher derivatives. Recall we used Rolle's Theorem for that–thus we'll consider

Theorem 12.2. (Generalized Rolle's Theorem) *Suppose f is $n-1$ times differentiable on $[a, b]$, and n times differentiable on (a, b). Suppose $f(a) = f'(a) = ... = f^{(n-1)}(a) = 0 = f(b)$, or equivalently $f(a) = 0 = f(b) = f'(b) = ... = f^{(n-1)}(b)$. Then there is a point $x \in (a, b)$ such that $f^{(n)}(x) = 0$.*

Proof. We do this by induction on n: first, we find a point in the interior with first derivative 0. Then restricting to this interval, we can apply Rolle's theorem to the derivative to get a point on the interior of our subinterval with second derivative zero. Continuing in this manner, applying Rolle's theorem, we get a point such that $f^{(n)}(x) = 0$. $\square$

This may not seem useful, because we have imposed lots of conditions on f, but in fact, if we drop all conditions that are not differentiability on the interval (f is $n-1$ times differentiable on the closed interval and n times on the open interval) then we can create a new function that satisfies the hypotheses of generalized Rolle's.

We can construct a polynomial whose derivative match f up to n

$$P(x) = f(a) + f'(a)(x-a) + \frac{1}{2}f''(a)(x-a)^2 + \dots \frac{1}{n-1}f^{(n-1)}(a)(x-a)^{n-1}$$

Then $f - P$ satisfies the conditions we need for generalized Rolle's, becasue by constuction $(f-P)^{(i)} = 0$ for $0 \leq i \leq n-1$. We also need $(f-P)(b) = 0$, and we can acheive this by adding $C(x-a)^n$ to P so that $f(b) - P(b) = 0$, which is possible. Note this new term changes none of the important derivative at a. Then this new function's nth derivative will vanish at a point in (a, b), This gives us information about $f(b)$ by unwinding our construction, the main result is

$$f(b) = \sum_{i=0}^{n-1} \frac{f^{(i)}(a)}{i!}(b-a)^i + \frac{f^{(n)}(c)}{n!}(b-a)^n$$

for some $c \in (a, b)$

WARNING, this does not mean $f(b) = \sum_{i=0}^{\infty} \frac{f^{(i)}(a)}{i!}(b-a)^i$, the sum might not converge and it is rare to have an equality.

13. Lecture 13 – 24 February 2017

There are several sifferent conceptions of the integral.

- The signed area under a curve
- Riemann sums–the sums of rectangles from the x-axis
- Legesgue integrals–taking the sum of rectangles from the y-axis
- Taking the antiderivative of a sufficiently nice function.

These are all the right notions, but connecting the first point to the first two is a major theorem.

Some of these ideas trace back to the ancient Greeks. For instance, consider the classical problem of squaring the circle–finding a square with the same area as a given circle. (This was called finding the quadrature of the circle.) The Greeks developed various techniques for computing the area of various objects. For the circle, they bounded it above and below by using polygons inscribed and excribed on the circle. Using this, Archimedes was able to show

$$3\frac{10}{71} < \pi < 3\frac{10}{70}.$$

A similar method, using rectangles, could be used to compute, for instance, the area under a parabola. Then by taking the limit with this method, you eventually get the area under $y = x^2$ is computed by $\frac{x^3}{3}$, and so on.

This is known as the classical method of exhaustion. You attempt to exhaust a space using shapes of which you know how to compute the area. This of course predates Riemann by around 2000 years, but wasn't formalized until him.

Knowing that we want something that computes area, there are some obvious axioms we'd like to impose. Gillman in AMM 100 #1 16-25 suggests the AB approach:

- Additivity: $\int_a^b + \int_a^c + \int_c^b$
- Betweenness: $m \leq f \leq M$ implies $m(b-a) \leq \int_a^b f \leq M(b-a)$.

In fact, by iterating this procedure over progressively smaller partitions, this is actually sufficient, as it uniquely determines what the integral can be.

Let $f : [a,b] \to \mathbb{R}$ be a bounded (but not necessarily continuous function). A partition is a collection $a = x_0 < x_1 < ... < x_n = b$ and we say $\Delta x_i = x_i - x_{i-1}$. Now let $m_i = \inf_{x_{i-1} \leq x \leq x_i} f(x)$ and $M_i = \sup_{x_{i-1} \leq x \leq x_i} f(x)$ Then by our earlier axioms we want

$$\sum m_i \leq \int_a^b f \leq \sum M_i$$

We call $\sum m_i = L(P,f)$, L for lower, and $\sum M_i = U(P,f)$, U for upper, where P denotes the partition $x_0, ...x_n$. Now we define $\underline{\int_a^b} f = \sup_P L(P,f)$ and $\overline{\int_a^b} f = \inf_P U(P,f)$. We have just defind a lot of things, but we will show that $L(P,f) \leq L(P',f)$, $\underline{\int} \leq \overline{int}$, and when the two are equal we say that f is *integrable*.

The trick here will be that if P is a partition then we can take a refinement, a partition including all points of P and some other points. For example breaking the interval $[0,1]$

into fourths is a refinement of breaking it into halfs. This refinement gives a better estimate, lower bounds go up and upper bounds go down, and for two partitions we can find a mutual refinement.

We will also need a notion of the length of an interval in a way that is additive. To do this we can take a nondecreasing function $\alpha : [a,b] \to \mathbb{R}$ and replace Δx_i with $\Delta\alpha_i = \alpha(x_i) - \alpha(x_{i-1})$. When we use this notion, we compute $\int_a^b f(x)d\alpha(x)$, and we will show this makes sense, i.e. $\int_a^b f(x)d\alpha(x) = \int_a^b f(x)\alpha'(x)dx$ when α' exists.

This generalization (Called The Riemann-Stieltjies Integral) is actually interesting for different α, for instance if α is a step function that goes from 0 to 1 at a point c then $\int_a^b f(x)d\alpha(x) = f(c)$.

We know that $L(P,f,\alpha) \leq U(P,f,\alpha)$, if we take two partitions P, P' then it is still the case that $L(P,f,\alpha) \leq U(P',f,\alpha)$ because we can take a common refinment of both partitions (combine all the points of both partitions). Call the common refinement P^* and we get

$$L(P,f,\alpha) \leq L(P^*,f,\alpha) \leq U(P^*,f,\alpha) \leq U(P',f,\alpha)$$

Now, if $\underline{int} = \overline{\int}$ then we will need that for any ϵ there is a partition P where $U(P,f,\alpha) - L(P,f,\alpha) < \epsilon$. We can do this if f is continuous, just look at the formula for $U(P,f,\alpha) - L(P,f,\alpha)$ it is

$$\sum \Delta\alpha_i \sup_{x_{i-1}\leq x\leq x_i} f(x) - \inf_{x_{i-1}\leq x\leq x_i} f(x)$$

Since f is continuous, it is uniformly continuous, and we can use continuity to figure out how small the partition must be to make the sum less than ϵ

14. LECTURE 14 – 27 FEBRUARY 2017

14.1. **Integration.** Let $[a,b] \xrightarrow{\alpha} \mathbb{R}$ be an increasing function, $[a,b] \xrightarrow{f} V$ a continuous map to a normed vector space. For P a partition $a = x_0 < x_1 < ... < x_n = b$, we defined a Riemann Sum

$$\sum_{i=1}^{n} \Delta\alpha_i f(t_i)$$

for some $t_i \in [x_{i-1}, x_i]$, and $\Delta\alpha_i = \alpha(x_i) - \alpha(x_{i-1})$. Rudin lets m_i, M_i denote the min and max value, respectively, that f takes on the i^{th} interval, which gives us upper and lower bounds for the value of the integral. (Note this only holds if $V = \mathbb{R}$, as in multiple dimensions we don't have a good notion for what maximum or minimum means. When V is not $\mathbb{R}$ we can generalize the difference $M_i - m_i$ by taking $\sup_{x,x' \in [x_{i-1},x_i]} |f(x) - f(x')|$.) Set

$$\Delta(P) = \sum_i \Delta\alpha_i (M - m_i),$$

so $\Delta(P)$ measures "how bad" our partition is, in the sense that it tells us the possible range of values our integral (if it exists) could take. Let $a \leq c < d \leq b$, and set $E(c,d) := \sup_{t,t' \in [c,d]} |f(t) - f(t')| \leq 2K$, where $K = \sup_{[a,b]} |f(t)|$. We could now equivalently define

$$\Delta(P) = \sum_{i=1}^{n} \Delta\alpha_i E(x_{i-1}, x_i),$$

which clearly must be identical to the above definition.

Definition 14.1. We say f is *integrable* (with respect to $d\alpha$) if for every $\epsilon > 0$ there is a partition P such that $\Delta(P) < \epsilon$. We write $\mathscr{R}(\alpha, V)$ for the set of all functions f from $\mathbb{R} \to V$ that are integrable with respect to α.

Note for P^* a refinement of P, $\Delta(P^*) \leq \Delta(P)$, so this condition really is checking that for sufficiently fine partitions, we are converging to a particular value.

It's easy to see $\Delta(P, cf, \alpha) = |c|\Delta(P, f, \alpha)$, $\Delta(P, f_1 + f_2, \alpha) \leq \Delta(P, f_1, \alpha) + \Delta(P, f_2, \alpha)$.

Corollary 14.2. *$\mathscr{R}(\alpha, V)$, the set of Riemann integrable functions to V with respect to α form a vector space (over the base field of V).*

The funny thing about our above presentation is that we've defined what it means to be integrable without actually defining what the integral is. We should fix that. Consider, we have

$$\Delta(P) \geq \sup_{\{t_i, t_i'\}} (R(P, \bar{t}) - R(P, \bar{t}')),$$

where the right hand side is the difference of two Riemann sums with respect to a partition P, and test points $\bar{t}, \bar{t}'$. This gives us the appropriate hint for what the integral should be. If $f \in \mathscr{R}(\alpha, V)$ (that is, if f is Riemann-integrable with respect to α, mapping to V), then

$$\int_a^b f(x) d\alpha(x) = I \in V$$

such that every parititon P, and every Riemann sum $R(P, \bar{t})$ is in a closed $\Delta(P)$-ball around I. We need to show existence and uniqueness of this, but otherwise, it will suffice as a definition.

- If I exists, it's unique: If I, I' both work, then we can take partitions P sufficiently small that $\Delta(P) < \epsilon$, then we must have $d(I, I') < 2\epsilon$ by the triangle inequality on a Riemann sum. But ϵ was arbitrary, so we must have $I = I'$.
- If V is complete, such an I exists. To see this, let $\epsilon_m \to 0$, and choose corresponding partitions P_m such that $\Delta(P_m) \leq \epsilon_m$, which we can do since f is assumed to be integrable. Then choose test points $\overline{t_m}$ to get corresponding Riemann sums. Then setting
$$I = \lim_{m\to\infty} R(P_m, \bar{t}_m)$$
works, if this limit exists. It is easy to verify this, however.

Now that we've defined the integral, we can immediately see some of its properties. It's clear from the limit of Riemann Sums definition that
$$\int_a^b (f_1 + f_2) d\alpha = \int_a^b f_1 d\alpha + \int_a^b f_2 d\alpha,$$
and similarly
$$\int_a^b c f d\alpha = c \int_a^b f d\alpha.$$
Thus, much like the differentiation case, we see that integration is a linear map.

Most of the theorems from Rudin can be immediately generalized to maps valued in normed vector spaces if you use the error function.

Lemma 14.3. (See Ruding 6.11 for a treatment over $\mathbb{R}$) *Let $[a, b] \xrightarrow{f} [m, M] \xrightarrow{\phi} V$ be a map with ϕ continuous, f Riemann integrable with respect to $\alpha : [a, b] \to \mathbb{R}$. Then $h = \phi \circ f$ is Riemann integrable.*

We'll note some important corollaries before giving the proof.

Corollary 14.4. *If f is integrable, then $|f(-)|$ is as well.*

Corollary 14.5. *If f, g are integrable, then so is $f \cdot g$, when f, g are real valued. (We need them to be real valued for multiplication to make sense.)*

To see this corollary, note by the lemma that squaring a function, cubing a function, etc. is a continuous map. Then since $f \cdot g = \frac{1}{4}((f+g)^2 - (f-g)^2)$, we immediately get the result from the Lemma. We'll now prove Lemma 14.3:

Proof. Fix $\epsilon > 0$, and let $K = \sup_{[a,b]} ||h|| \leq \sup_{[m,M]} ||\Phi||$. Since Φ is defined on a compact space, it's actually uniformly continuous, so there's a δ such that $|s - t| < \delta$ implies $||\phi(s) - \phi(t)|| < \epsilon$. We can assume $\delta < \epsilon$, by shrinking it if necessary.

We want to find a partition P of $[a, b]$ such that $\Delta(P, f, \alpha) \leq \delta^2$, which we can do since f is integrable. Why do we want this? Recall $\Delta(P, f, \alpha) = \sum \Delta\alpha_i (M_i - m_i)$. If this is $< \delta^2$, then we can split the sum into parts with $\Delta\alpha_i < \delta$ and parts with $(M_i - m_i) < \delta$.

Let $A, B \subset \{1, ..., n\}$ be a partition, A containing those i such that $|M_i - m_i| < \delta$, and B contains those i with $|M_i - m_i| \geq \delta$. Then we have

$$\delta \sum_B \Delta\alpha_i \leq \sum_B \Delta\alpha_i (M_i - m_i) \leq \Delta(P) \leq \delta,$$

hence $\sum_B \Delta\alpha_i \leq \delta$. Now we have

$$\Delta(P, h, \alpha) = \sum \Delta\alpha_i E(x_{i-1}, x_i) = \sum_A \Delta\alpha_i E(x_{i-1}, x_i) + \sum_B \Delta\alpha_i E(x_{i-1}, x_i).$$

We know

$$\sum_A \Delta\alpha_i E(x_{i-1}, x_i) \leq \sum_A \epsilon \Delta\alpha_i \leq \epsilon(\alpha(b) - \alpha(a)).$$

Similarly, we have

$$\sum_B \Delta\alpha_i E(x_{i-1}, x_i) \leq 2K\delta \leq 2K\epsilon.$$

Thus, we've shown

$$\Delta(P, h, \alpha) \leq (\alpha(b) - \alpha(a) + 2K)\epsilon = C\epsilon$$

for some constant C, and ϵ is arbitrary, so we're done. $\square$

14.2. **Change of Variables.** We can recall from high school calculus the change of variables formula, which was something like

$$\int_{g(a)}^{g(b)} f(g(x))g'(x)dx = \sum_a^b f(g(y))dy.$$

If you saw a proof of this, it was probably very involved. In the Riemann-Stieltjes context, we'll split this up into two parts. Suppose we have a strictly monotonically increasing map $\varphi : [A, B] \to [a, b]$ (recall this means $s < t \implies \varphi(s) < \varphi(t)$) such that $\varphi(A) = a, \varphi(B) = b$.

$$\begin{tikzcd}
{[A,B]} \arrow[r, "\varphi"] \arrow[rr, bend left, "g"] \arrow[rrd, bend right, "\beta"'] & {[a,b]} \arrow[r, "f"] \arrow[rd, "\alpha"] & V \\
 & & \mathbb{R}
\end{tikzcd}$$

Now, given $[a, b] \xrightarrow{\alpha} \mathbb{R}$, a map $[a, b] \xrightarrow{f} V$, setting $g = f \circ \varphi$ and $\beta = \alpha \circ \varphi$, we claim

$$\int_A^B g d\beta = \int_a^b f d\alpha.$$

Why is this equivalent to the above formulation? Basically just plug things in. However, now that we've phrased it in this manner, the proof becomes more or less trivial. Partitions of $[A, B]$ correspond bijectively with partitions of $[a, b]$, related by the map φ, and then the proof just falls out.

15. Lecture 15 – 01 March 2017

Today we'll talk a bit more about the Riemann-Stieltjes integral, and then discuss the fundamental theorem of calculus. Recall that we said the differential operator D, given by $f \mapsto f'$ is a linear map. Recall one common version of FTC states that for some F with $F' = f$, we have $F(x) = F(a) + \int_a^x f(\xi)d\xi$. This gives some useful corollaries–for instance, by considering $\int_a^b (FG)'$, by comparing the product rule with the fundamental theorem of calculus, we immediately get integration by parts.

> **Example 15.1.** We have $(x^n e^x)' = nx^{n-1}e^x + x^n e^x$, so we can use induction to compute the integrals of $x^k e^x$.

Theorem 15.2. *When f is Reimann Integrable with respect to α, and α has an integrable (w/r/t x) derivative, then $\int_a^b f d\alpha = \int_a^b f\alpha' dx$.*

Proof. If we examine terms in riemann sums, a term in a sum for $\int_a^b f d\alpha$ is $f(t_i)(\alpha(x_i) - \alpha(x_{i-1}))$, but since α is differentiable we can rewrite this as $f(t_i)\alpha'(u_i)(x_i - x_{i-1})$ for $u_i \in [x_{i-1}, x_i]$ but our choice of t_i was arbitrary, so we can set $t_i = u_i$, which gives us the terms in a riemann sum of $\int_a^b f\alpha' dx$ $\square$

Theorem 15.3. (Fundamental Theorem of Calculus) *If f is continuous then the function $F(x) = \int_a^x f(u)du$ is continuous, and differentiable, and has derivative equal to f.*

Proof. For Continuity. Take $x < y$ then $F(y) - F(x) = \int_a^y f(t)dt) - \int_a^x f(t)dt = \int_x^y f(t)dt \leq M(y - x)$ because f can be bounded above by M. You can then make $|y - x|$ very small and make $F(y) - F(x)$ similarly small. `\droptablestudents;` We still have

$$F(y) - F(x) = \int_x^y f(t)dt = \int_x^y f(x_0)dt + \int_x^y f(t) - f(x_0)$$

Taking absolute values we get $|F(y) - F(x)| \leq |f(x_0)|(y - x) + (y - x)\sup|f(t) - f(x_0)|$ we rearrange this to see $|\frac{F(y)-F(x)}{y-x} - f(x_0)| \leq \sup|f(t) - f(x_0)|$ In the limmit as $y \to x$ this means F is differentiable at x, and replacing $x_0 = x$ gives us $F'(x) = f(x)$. $\square$

Remark 15.4. In general, the way that we got the lower case Greek letters is by taking the upper case letters and trying to write them quickly. Hence Z became ζ, and Ξ became ξ.

Theorem 15.5. (Fundamental Theorem of Calculus II, Rudin 6.21) *Suppose $f \in \mathscr{R}[a, b]$, $f = F'$ for F differentiable on $[a, b]$. Then*

$$\int_a^b f(x)dx = F(b) - F(a).$$

Proof. Recall for any function f, we said there were sufficiently fine partitions P such that $\Delta(P) < \epsilon$. Note the resemblance of our equation to the Mean Value Theorem for derivatives. Unfortunately, this worked over $\mathbb{R}$, but didn't generalize very well to arbitrary normed vector spaces.

Take our partition P for F. Then

$$F(b) - F(a) = F(x_n) - F(x_0) = \sum_i F(x_i) - F(x_{i-1}).$$

Then by the mean value theorem, we have the RHS is

$$\sum_i f(t_i)(x_i - x_{i-1}),$$

which is the definition of the integral, so in the real case we're done.

In general, we have

$$F(x_i) - F(x_{i-1}) = \Delta x_i f(x_i) + E_i,$$

where $\Delta x_i = x_i - x_{i-1}$, E_i is an error term. This sum still telescopes, so we get

$$||F(b) - F(a) - \sum (x_i - x_{i-1}) f(x_i)|| \leq \sum_i ||E_i|| \leq \Delta(P) \cdot M,$$

where M is the penalty factor we get for a version of the mean value theorem on a normed vector space. Then by taking $\Delta(P)$ sufficiently small, we can kill the RHS and bound the sum into an integral to get

$$||F(b) - F(a) - \int_a^b f(x) dx|| \leq \epsilon + M\epsilon,$$

where the RHS dies and so we get the general version of the theorem. □

Example 15.6. Consider α a step function, discontinuous at x_0 and nowhere else. Then given f continuous at x_0, it follows by definition of the Riemann-Stieltjes integral that for $x_0 \in [a, b]$,

$$\int_a^b f(x) d\alpha = f(x_0).$$

There are some immediate properties we get about $d\alpha$ from thigns like this. Note

$$\int f d(\alpha + \beta) = \int f d\alpha + \int f d\beta,$$

as well as

$$\int f d(c\alpha) = c \int f d\alpha.$$

This is almost linear–but not quite? What's going wrong? Remember α has to be increasing. If we multiple α by a nonnegative constant, we're fine, but the latter equality doesn't really make sense for $c < 0$. We can resolve this fairly easily by using the first equality, and then splitting up $\alpha = \alpha_+ - \alpha_-$, though we do have to modify a few proofs. We end up with a requirement of bounded variations on α, for this to be sensible.

16. Lecture 16 – 03 March 2017

16.1. **Limits of Integrals.** Recall when we say $F = \sum_{n=1}^{\infty} f_n$, we mean

$$F = \lim_{N\to\infty} \sum_{n=1}^{N} f_n.$$

For this to make sense, we want a group structure (a vector space structure), and (in the case that this sum is actually an integral) a metric on our space. This gives us a rather subtle question to consider, however: how do derivatives and integrals interact with limits?

> **Example 16.1.** It certainly seems that the function
>
> $$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$
>
> should be its own derivative. However, we don't necessarily know how to show this, as we're taking the derivative of infinitely many terms.

In general, in class we see very settled math, where we understand the results and what's important. However, when the subject was being developed, this was not necessarily the case. There are many interesting counterexamples to seemingly intuitive results. You can see quite a few in Rudin, or even more in Gelbaum and Olsted's *Counterexamples in Analysis*. Before proving theorems that make this make sense, we should see some cautionary tales about why this process is not always so easy.

We'll just present a single notable counterexample here. Consider

$$T(x) := 1 - 2d(x, \mathbb{Z}).$$

This looks like a triangle-sine function, or something similar. It is $\mathbb{Z}$-periodic, and so completely determined by how it acts on $[0,1]$. Now powers of T, $T^n(x)$, are still functions, they just spend more time near 0. We can also consider $T^n(mx)$, which still is integrable. Indeed, taking the limit as $n \to \infty$, we get the characteristic function of $\frac{1}{m}\mathbb{Z}$, which has integral 0. However, note that $\chi_{\mathbb{Q}}$, the characteristic function of the rationals, is not Riemann Integrable, since $\mathbb{Q}$ is dense in $\mathbb{R}$, so we can't get sufficiently fine partitions.

Then taking

$$T^n(m!x),$$

when m gets sufficiently large this will vanish at all the rationals, and n sufficiently large it will vanish at all the irrationals. Thus

$$\lim_{n\to\infty} \lim_{m\to\infty} T^n(m!x) \xrightarrow{\text{pointwise}} \chi_{\mathbb{Q}}.$$

Thus we have a function that it really seems like we should be able to talk about, as we've just done normal things with it, converge to something that isn't integrable. As is hinted by the above, the important condition to add is uniform convergence of functions. (NB: If you consider our functions, it's nonzero only at the rationals, which are a countable subset. It therefore really feels as though the integral should be 0. Going about formalizing this starts you on the road to the Lebesgue Integral.

We already have a theorem which states that the map $f \mapsto \int_a^b f(x)d\alpha(x)$ has $||\int|| = \alpha(b) - \alpha(a)$. That is, if $||f|| \leq M$, then

$$||\int_a^b f d\alpha|| \leq (\alpha(b) - \alpha(a))M.$$

This is uniformly continuous, which makes it plausible that if $f_n \to f$ uniformly, we expect

$$\int_a^b f_n d\alpha \to \int_a^b f d\alpha,$$

as $n \to \infty$. The difficult part of this argument is that for $f_n \to f$ uniformly convergent and f_n integrable, then f is integrable. That is, $\mathscr{R}([a,b], d\alpha, \mathbb{R})$ is a closed space.

We'll fix α, and we need to show $\Delta(P, f) = \sum_{i=1}^n \Delta\alpha_i \sup_{[x_{i-1},x_i]} ||f(t) - f(t')||$ can be made arbitrarily small. Consider $\Delta(Pf_n)$. If $||f - f_n|| < \epsilon$ in the uniform norm, then

$$\begin{aligned} ||f(t) - f(t')|| &\leq ||f_n(t) - f_n(t')|| + ||f_n(t) - f(t)|| + ||f_n(t') - f(t')|| \\ &\leq ||f_n(t) - f_n(t')|| + 2\epsilon. \end{aligned}$$

Then it follows that $\Delta(P, f) \leq \Delta(P, f_n) + 2\epsilon(\alpha(b) - \alpha(a))$, so by taking ϵ sufficiently small, we can squeeze down the second term. Refining the partition also squeezes down $\Delta(P, f_n)$ by hypothesis, so it follows that $\Delta(P, f)$ can become arbirarily small, by refining P.

16.2. **Limits of Derivatives.** In contrast to the integral case, uniform limits do *not* commute with derivatives. Consider some nice periodic function, such as sin and cos even though we haven't formally defined either yet. We have

$$\frac{d}{dx} f_0(Nx) = N f_0'(Nx).$$

Then we have that the function

$$f_0(x) + \frac{1}{3} f_0(4x) + \frac{1}{3^2} f_0(4^2 x) + ... = \sum \frac{1}{3^n} f_0(4^n x)$$

is clearly well defined and converges, since the partial sums have difference $F_N - F_M = \frac{1}{3^M}(\text{something bounded})$. However, if we take the termwise derivative, we see that we get something that very definitely does not converge.

Luckily, however, whenever the derivative of a sequence does make sense, then the limit of the derivative is equal to the derivative of the limit. For reference, see Rudin on the Weierstrass M test.

Lemma 16.2. *If f_n' converges uniformly, and there is x_0 such that $f_n(x_0)$ converges to a point, then f_n converges uniformly to some function f.*

Proof. Consider we have $f_m(x) - f_n(x) = (f_m(x) - f_m(x_0)) - (f_n(x) - f_n(x_0)) + (f_n(x_0) - f_n(x_0))$. The last term converges by hypothesis. For the first two terms, we can either use the MVT, or if f_n' is integrable, then it follows by the Fundamental Theorem of Calculus. ∎

Lemma 16.3. *Let $f, g : [a,b] \to \mathbb{R}$ both be nondecreasing. We want to show*

$$\int_a^b f dg + \int_a^b g df = fg(b) - fg(a)$$

without assuming the product is integrable.

Note we have

$$fg(b) - fg(a) = \sum_{i=1}^{n} fg(x_i) - fg(x_{i-1}),$$

which we can rewrite as

$$\sum f(x_i)(g(x_i) - g(x_{i-1}) + g(x_{i-1})(f(x_i) - f(x_{i-1})) = \sum f(x_i)\Delta g_i + g(x_i)\Delta f_i.$$

Then for sufficiently fine partitions, this approaches $\int f dg + \int g df$

17. Lecture 17 – 06 March 2017

The Weierstrass (1815-1897) approximation theorem was proved in 1895, when Weierstrass himself was 80! He was still doing great work at the end of his life. This is in fact pretty early, considering Galois only proved things in the year before he died.

Theorem 17.1. *(Putnam Exam 1958, B7) Let $f : [a,b] \to \mathbb{R}$ be continuous. If for every nonnegative integer k the integral*

$$\int_a^b x^k f(x)dx = 0,$$

then f is identically 0.

Proof. We saw on a recent problem set that $\int_a^b (f(x))^2 dx = 0$ implies $f \equiv 0$. Then we want to show $\langle f, f \rangle = 0$, and we know $\langle x^k, f \rangle = 0$. It follows that $\langle p, f \rangle = 0$ for every $p \in \mathbb{R}[x]$. Then if f were a polynomial, we'd be done immediately. In general, of course, f isn't. Thus, instead, we'd like to show there are $f_1, f_2, ... \in \mathbb{R}[x]$ such that $f_n \to f$ uniformly. This, however, is the statement of the Weierstrass Approximation Theorem, and gives at least one reason to care about the theorem–it lets us solve Putnam problems. $\square$

Theorem 17.2. (Weierstrass Approximation Theorem) *For every $\in \mathcal{C}([a,b], \mathbb{R})$, there is a sequence of functions*

$$f_1, f_2, ..., f_n, ... \in \mathbb{R}[x]$$

that converge to f uniformly.

NB: $f_n \to f$ uniformly implies f is continuous, so the above theorem is actually if and only if. An equivalent statement is that $\mathbb{R}[x]$ is *dense* in $\mathcal{C}([a,b], \mathbb{R})$.

Proof. There are a number of different ways to approach this theorem. We will not be using the proof that can be found in Rudin, though you can check there for an alternative. Given f, we will approximate it by a piecewise linear map $g : [a,b] \to V$: that is, there is a partition P such that g is linear on each component of the partition. Note that the set of piecewise linear maps forms a vector space, which we'll denote $\mathrm{PL}([a,b], \mathbb{R})$.

One neat thing about $\mathrm{PL}([a,b], \mathbb{R})$ is that it has basis $\{|x - x_0| \mid a < x_0 < b\}$. Then it's enough to show that PL is dense in $\mathcal{C}([a,b], \mathbb{R})$, and $\mathbb{R}[x]$ uniformly approximates PL. Then it's enough to show that $\mathbb{R}[x]$ uniformly approximates $|x - x_0|$, since every PL function is a finite linear combination of these.

To see PL is dense in $\mathcal{C}([a,b], \mathbb{R})$, note that $[a,b]$ is compact, so f is uniformly continuous, then on an interval of length $< \delta$, the function doesn't change by more than ϵ, so we can just have our linear approximation connect the two endpoints of the subinterval, and we won't be off by more than 2ϵ.

For uniform approximation, we'll change variables such that $x_0 = 0$. Then it's sufficient to approximate $|x|$ on the interval $[-M, M]$, for M arbitrarily large. But then we can change variables to $\frac{x}{M}$. Thus, we've reduced to checking that $|x|$ on the interval $[-1, 1]$ can be uniformly approximated by polynomials.

Note that $|x| = \sqrt{x^2}$, where we take the convention of using the positive square root. Then $|x| = (1 - (1 - x^2))^{1/2}$. Now since $|x| \leq 1$, we have $|1 - x^2| \leq 1$. Then $x = 0 \iff 1 - x^2 = 1$. Now consider the Taylor expanstion of $(1 - t)^{1/2}$, with $|t| \leq 1$. We can compute derivatives to get a Taylor expansion using the chain rule, since we know what derivatives of x^2 look like. However, we do need to verify that $\sqrt{x}$ is differentiable, which we can do by considering

$$\sqrt{x} - \sqrt{x'} = \frac{x - x'}{\sqrt{x} + \sqrt{x'}},$$

and then we're only relying on continuity of $\sqrt{x} - \sqrt{x'}$, which is clear. From this, we pull out a taylor expansion

$$(1-t)^{\frac{1}{2}} = 1 - \frac{t}{2} - \frac{t^2}{8} - \frac{t^3}{16} - \frac{5t^4}{128} - ...$$

for $|t| \leq 1$. Note other than the first term, all of these signs are negative. This gives us a useful tool.

Claim: If $c_n \geq 0$, $M > 0$, and $\sum c_n t^n < M$ for all $t \in [0,1]$, then $\sum c_n \leq M$, and the sum $\sum c_n t^n$ converges uniformly on $[0,1]$. To see this, note that if $\sum c_n > M$, then there's an N sufficiently large that $\sum_{n=1}^{N} c_n > M$, but then we have a polynomial $\sum_{n=1}^{N} c_n t^n$, and there's some t sufficiently close to 1 such that this is $> M$, which is a contradiction. Then we can apply the Weierstrass M-test to get the value at 1.

Now the function $1 - (1-t)^{\frac{1}{2}}$ is of this form, so notably, the convergence is uniform. Then chopping off the first N terms, we get a sequence of polynomials which converge uniformly to $1 - (1-t)^{\frac{1}{2}}$. $\square$

Note that any sequence $1, x^n, x^{2n}, x^{3n}, ...$ would actually work for this approximation. In general, we might ask what conditions we must put on $1, x^{n_1}, x^{n_2}, ...$ for this vector space to be dense in the set of continuous functions. Knows as the Muntz-Szósz, these polyonimials are dense if and only if

$$\sum \frac{1}{n_i} = \infty.$$

In fact, polynomials and continous functions aren't just vector spaces–we can take products in a nice way, so they're actually normed algebras. Then we're saying that the algebra of polynomials is dense in the algebra of continuous functions on the interval $[0,1]$. This gives a vast generalization we might consider: what is a sufficient condition for $\mathcal{A} \subset \mathcal{C}(X, \mathbb{R})$, with X a metric space, to be dense? Clearly, $\mathcal{A}$ must separate points, and it can't vanish identically at any point. The incredible thing is that, if X is compact, these conditions are actually also sufficient.

18. Lecture 18 – 08 March 2017

Show up to class, you hooligans!

The Stone-Weierstrass Theorem is a vast generalization of the Weierstrass approximation theorem. It's actually due to Stone–Weierstrass provided a key lemma, but it was Stone who showed the theorem.

Theorem 18.1. *Let X be a compact topological space, consider the algebra $C(X, \mathbb{R})$. If $\mathcal{B} \subset C(X, \mathbb{R})$ is a closed (topologically) subalgebra then it is equal to $C(X, \mathbb{R})$ provided it meets the following requirements:*

(i) There is no point $x \in X$ so that $f(x) = 0$ for all $f \in \mathcal{B}$
(ii) For all points x, y there is a function $f \in \mathcal{B}$ so that $f(x) \neq f(y)$

Remark 18.2. It's actually important that this is over the reals, We will be taking maximums and minimums. While most theorems actually work slightly better over $\mathbb{C}$, here we actually would need to add some additional hypotheses.

Theorem 18.3. (Rudin 7.31) *For any field F and any algebra of functions$A : E \to K$ satisfying conditions 1 and 2 then for any $e_1, e_2 \in E$ and any $k_1, k_2 \in K$ there is $f \in A$ so that $f(e_i) = k_i$ A restatement is that if A is a subalgebra of F^2 so that neither the first and second coordinates is always 0, and A is not contianed in the diagonal (pairs with identical first and second coordinate), then $A = F$*

Proof. The only subalgebras of F^2 are the ones we just excluded and the whole Algebra. □

This proves Stone-Weierstrauss theorem in the case that X is 2 points. We iterate the process to prove it for any finite set.

(0) Given $P \in \mathbb{R}[x], f \in \mathcal{B}$, we must have $P(f) \in \mathcal{B}$. This is just the definition of an algebra
(1) Given $f \in \mathcal{B}$, we have $|f(\cdot)| \in \mathcal{B}$. This follows from (0) and the argument we used to show the Weierstrass approximation theorem last class.
(2) We have $\max(f, g) = \frac{f+g}{2} - \frac{1}{2}|f - g|$, thus the max, and similarly the min of two functions, is contained in the algebra. By induction, this follows for the max/min of any finite set of functions.
(3) Fix $f \in \mathcal{C}(X, \mathbb{R})$, and $x \in X$, $\epsilon > 0$. We want to approximate it with functions in $\mathcal{B}$. We know for each x there is $g_x \in \mathcal{B}$ such that $g_x(x) = f(x)$, and $\min_y(g_x(y) - f(y)) > -\epsilon$.

    To see this, by 7.31 we have $h_y \in \mathcal{B}$ such that $h_y(x) = f(x), h_y(y) = f(y)$. Then there's an open neighborhhod such that $h_y(z) - f(z) > -\epsilon$ by continuity for all $z \in J_y$. Now we can do this for every point, then use compactness to move down to a finite subcover. Then take the pointwise max, using (2), to get

$$\max(h_{y_1}, ..., h_{y_n}) = g_x.$$

    We have an analagous result for bounding on the other side.
(4) We have a neighborhood

$$N_x := \{x' \mid |g_x(x') - f(x')| < \epsilon\}.$$

By construction, $x \in N_x$. Then by compactness, X is the finite union of such neighborhoods, which we can take the min over and the result follows.

Remark 18.4. Recall we stated that the theorem didn't hold over $\mathbb{C}$. Consider the map $\mathbb{R}/\mathbb{Z} \cong \mathbb{C}_1$, given by $x \mapsto e^{2\pi i x}$. Then consider the algebra $\mathcal{A} = \mathbb{C}[z]$. This is certainly an algebra, so let $\mathcal{B} = \overline{\mathcal{A}}$ be its uniform closure. Then $\mathcal{B} \not\ni z^{-1}$: one easy way to see this is that all of the z^k are pairwise orthogonal under the inner product

$$\frac{1}{2\pi}\int_0^{2\pi} f(x)\overline{g(x)}dx.$$

Then they're also orthogonal to z^{-1}, and hence so is everything in $\mathcal{B}$.

If we add the criterion that $\mathcal{B} = \overline{\mathcal{B}}$, that is, that it's closed under complex conjugation, then we will actually get $\mathcal{B} = \mathcal{C}(X, \mathbb{C})$.

One important thing we learned from algebra is that the properties of an object are often encoded in some function space. Indeed, we can reconstruct X (as a topological space) from $\mathcal{C}(X, \mathbb{R})$. It's often profitable, therefore, to study this function space instead of X.

A point $x \in X$ corresponds to an evaluation map $i_x : C(X, \mathbb{R}) \to \mathbb{R}$ that sends $f \to f(x)$. i_x is a surjective, continuous algebra homomorphism.

Proposition 18.5. *The map $x \to i_x$ is a bijection from X to the continuous, surjective algebra homomorphisms $C(X, \mathbb{R}) \to \mathbb{R}$*

Proof. The map is well defined and injective, because we can separate points. We just need to show surjectivity. Let i be a continuous, surjective algebra homomorphisms $C(X, \mathbb{R}) \to \mathbb{R}$, and let $M = \ker i$, then there exists an$x \in X$ so that $M|_x = 0$

Assume there is no x. Then M is a closed subalgebra of $C(X, \mathbb{R})$, and it is not everything if i is not the zero map. But then M vanishes at some x or there are some x, y that are not separted by M. But if M did not separate x, y it could not be an ideal in $C(X, \mathbb{R})$, so it muct vanish at some point. $\square$

Note that the closed sets of X correspond to closed ideals of $\mathcal{C}(X, \mathbb{R})$, by looking at where they vanish. This actually lets us pull out the topology of X as well.

If we considered $\mathcal{C}((0, 1], \mathbb{R})$, then we have closed ideals that vanish on things that aren't closed sets in $\mathbb{R}$–for instance, the ideal of all functions that vanish on all but finitely many of $\frac{1}{n}$. Then Zorn's Lemma gives us a maximal ideal. Doing this gives us something called the Stone-Čech compactification. This has the very strange property that it contains $(0, 1]$, and has the property that all bounded functions on it are continuous.

Remark 18.6. If instead of all continuous functions, we'd restricted to polynomials, we'd get something called the Zariski topology, which is the topology used for algebraic geometry (Warning: it's *very* different from the usual topology). So if you liked this style of argument, make sure to check out Math 137 next spring!

Definition 19.1. An analytic function (near 0–we can translate by a should we choose) is a function of the form

$$\sum_{n=0}^{\infty} c_n x^n$$

which converges at some $x_0 \neq 0$. This can be done over any complete, valued field. That is, any complete field with $d(x,y) = |x-y|$ and $|xy| = |x||y|$. For instance $\mathbb{R}, \mathbb{C}, \mathbb{Q}_p$ are all complete, valued fields.

Theorem 19.2. (Rudin 8.1+) *If $|x_0| = R > 0$, then*

$$f(x) = \sum_{n=0}^{\infty} c_n x^n$$

converges for all x such that $x < R$. For every $\epsilon > 0$, converges is uniform in $\{x \in \mathbb{F} \mid |x| < R - \epsilon\}$, f is continuous and differentiable on $|x| < R$, with derivative

$$f'(x) = \sum_{n=1}^{\infty} n c_n x^{n-1}.$$

Remark 19.3. We can't necessarily expect convergence on the boundary of this neighborhood–for instance, consider the alternating harmonic series,

$$f(x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + ...$$

which converges at 1, but diverges at -1. On the interval $(-1,1)$ however, we get nice things from the theorem, like

$$f'(x) = 1 - x + x^2 - x^3 + ... = \frac{1}{1+x},$$

which eventually will give us one way we can define the logarithm.

Example 19.4. Applying the theorem to the function

$$\exp(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + ...,$$

we see this converges everywhere, since factorials are big, so exp is continuous and differentiable everywhere, with derivative $\exp'(x) = \exp(x)$.
What about in the p-adics? We have $|n!|_p \sim p^{n/(p-1)}$, so this certainly isn't going to converge everywhere–we will need $x \equiv 0 \mod p$ for p an odd prime.

Proof. We clearly must have $c_n x_0^n \to 0$ as n grows large. Taking the absolute value of this, it follows that $|c_n|R^n \to 0$, so the $|c_n|R^n < C$ for all n. Then

$$|c_n x^n| = |c_n| \left(\frac{x}{R}\right)^n R^n < c \left(\frac{x}{R}\right)^n.$$

Then $|x| < R$ implies that the series converges, and the M-test gives uniform converges whence $|x| \le R - \epsilon$.

For the derivative, consider

$$\lim_{y \to x} \frac{f(y) - f(x)}{y - x}.$$

We'd like to show that this is equal to

$$\lim_{N \to \infty} \sum_{n=1}^{N} n c_n x^{n-1}.$$

The first thing to note is that this converges in the same region, as when $(x/R)^n \to 0$, we also have $n(x/R)^n \to 0$. For y close to x, we have $|y| < R$, hence $|y| \le R - \frac{\epsilon}{2}$, and we can choose y close enough to x that $|y - x| < frac\epsilon 2$. Then

$$f(y) = \lim_{N \to \infty} \sum_{0}^{N} c_n y^n \implies \frac{f(y) - f(x)}{y - x} = \lim_{N \to \infty} \sum_{0}^{N} c_n \frac{y^n - x^n}{y - x},$$

and we can expand our the right side, as $y - x$ divides $y^n - x^n$, and get a bound, which we can check converges, and indeed, converges to $n c_n x^{n-1}$. The remainder terms have uniform bound that goes to 0, which gives us differentiability. ☐

We can of course iterate the above argument to compute higher derivatives. It follows from this that $f^{(n)}(0) = n! c_n$, so given all of the derivatives at 0, we can actually construct the function, assuming that it's analytic.

Warning 19.5. If we aren't assuming that the function is analytic, this is *not true* over $\mathbb{R}$. (It is true over $\mathbb{C}$, however, which is one thing that makes complex analysis so much nicer than real analysis.) It's in fact very useful, for construction of a thing called "partitions of unity," that we can construct a nonconstant infinitely differentiable function with $f^{(n)}(0) = 0$ for all $n \ge 0$. As an exercise, construct such a function.

Let's return to the example

$$x - \frac{x^2}{2} + \frac{x^3}{3} - \ldots,$$

which we mentioned earlier converges for $|x| < 1$, and gives a way to define $\log(1 + x)$.

Theorem 19.6. (Abel) *If $s = \sum_{n=0}^{\infty} c_n$ is a convergent sum, then*

$$f(x) := \sum_{n=0}^{\infty} c_n x^n$$

extends to a continuous function on $(-1, 1]$, with $f(1) = s$.

Proof. We already knew convergence on $(-1, 1)$, so the interesting thing about this theorem is that we can extend the covnergence to 1. We already know $s = \lim s_n$, for $s_n := \sum_{m=0}^{n} c_n$. Then we have

$$f(x) = \sum_{n=0}^{\infty} (s_n - s_{n-1}) x^n.$$

Collecting like terms, we can rewrite this as
$$\lim_{N\to\infty} s_0(1-x)+s_1(x-x^2)+...+s_{n-1}(x^{n-1}-x^n)+s_nx^n.$$
Given $\epsilon > 0$, we can find an N such that $|s_n - s| < \epsilon$ for $n > N$. Then we can split up our sum as
$$f(x)-s=(1-x)\sum_0^N(s_n-s)x^n+(1-x)\sum_{N+1}^{\infty}(s_n-s)x^n.$$
Then the right hand side is by assumption less than □ andy's notes

Consider
$$\left(\sum_{n=0}^{\infty}a_nx^n\right)\left(\sum_{n=0}^{\infty}b_nx^n\right)=\left(\sum_{n=0}^{\infty}c_nx^n\right).$$
Because each of these converges somewhere, the terms are bounded by some constant over R^n. Then we have
$$(n+1)\left(\frac{\alpha}{R^m}\frac{\beta}{R^{n-m}}\right)=(n+1)\frac{\alpha\beta}{R^n},$$
so we get $c = a \cdot b$, and
$$c_n = \sum_{m+m'=n} a_m b_m.$$
Then we can apply Abel's theorem to the a, b, c functions, and if they all converge, then the product really is the true product.

20. Lecture 20 – 20 March 2017

> "π-day happened over the break–I believe sometime this week, we'll even be able to prove that π-day is supposed to happen in March."

We were talking before the break about analytic functions–that is, functions of the form

$$f(x) = \sum_{n=0}^{\infty} c_n x^n,$$

which converge in some circle $|x| < R$. We remarked this worked over $\mathbb{R}, \mathbb{C}, \mathbb{Q}_p$, and really any complete valued field.

We observed that there's an equivalent notion of a function being analytic around a point a by considering $f(x-a)$. However, we can simply change variables to move back to the origin–in particular, it doesn't change any derivatives. There is an exception to this, if we want to talk about a function at multiple points. Suppose we have a point a in our circle of points, so $|a| < R$. This gives us a circle of radius r around a, completely contained in our larger circle. Then it seems like a natural guess that f should be analytic around a with radius r. We already know that $f'(a), f''(a)$, ... exist and are the expected thing at a. Moreover,

$$f(x) = \sum_m = 0^\infty f^{(m)}(a)\frac{(x-a)^m}{m!}$$

if $|x-a| < R - |a|$. The proof of this is more or less a computation. We can rewrite

$$\begin{aligned} f(x) &= \sum c_n((x-a)+a)^n \\ &= \sum_n \left(\sum_{m=0}^{n} c_n \binom{n}{m} a^{n-m}(x-a)^m \right) \\ &= \sum_{m=0}^{\infty} \sum_{n=m}^{\infty} c_n \binom{n}{m} a^{n-m}(x-a)^m. \end{aligned}$$

If this converges absolutely, then we can rearrange the terms arbitrarily. (See Rudin 8.3, or better yet, prove it yourself.) To see this in our case, consider

$$|c_n| \sum \binom{n}{m} |a|^{n-m}|x-a|^m = |c_n|(|a| + |x-a|)^n < |c_n|R^n,$$

and this converges. Thus we can rewrite our original above sum as

$$\frac{1}{m!}\sum_{n=m}^{\infty} c_n \frac{n!}{(n-m)!}a^{n-m} = \frac{1}{m!}f^{(m)}(a),$$

as desired.

20.1. Special Functions.

We already gave the example of

$$E(x) := \sum_{n=0}^{\infty} \frac{x^n}{n!} = E'(x),$$

the exponential function, which converges everywhere. The manipulation above therefore works everywhere, so

$$E(x) = \sum_{m=0}^{\infty} \frac{E^{(m)}(a)}{m!}(x-a)^m = \sum_{m=0}^{\infty} E(a)\frac{(x-a)^m}{m!} = E(a)E(x-a).$$

then over $\mathbb{R}$ or $\mathbb{C}$, this proves for every x, a, that $E(x) = E(a)E(x-a)$. (Note this isn't true e.g. over $\mathbb{Q}_p$, where factorials get small instead of large, so the radius of convergence is not infinite.) Let $F = \mathbb{R}$ or $\mathbb{C}$. Then the above suggests very strongly that E gives us a map $(F, +) \to F^\times$, which is a group homomorphism. It's easy to check inverses and that the identity is preserved, so the map is indeed a homomorphism.

Over $\mathbb{R}$, we will show

$$(\mathbb{R}, +) \xrightarrow{E} \mathbb{R}^\times$$

has trivial kernel, and image the positive reals. Over $\mathbb{C}$, the picture is more interesting. We actually get a short exact sequence

$$\{0\} \to 2\pi\mathbb{Z} \to (\mathbb{C}, +) \xrightarrow{E} \mathbb{C}^\times \to \{1\}.$$

First, we'll do the argument over $\mathbb{R}$: for $x > 0$, we have $e^x > x$ clearly. This means that as $x \to \infty, e^x \to \infty$ as well.

Remark 20.1. We've switched to the notation $E(x) = e^x$, though we haven't actually proved this is the case yet. We will make gestures at the argument in the following remark.

We know $e^0 = 1$, so we know that $E : [0, \infty) \to [1, \infty)$ is surjective. But then since we already verified that E is a group homomorphism, it follows immediately that as $x \to -\infty$, $E(x) \to 0$, but E must remain strictly positive. Thus $E : (-\infty, 0] \to (0, 1]$ is surjective. To see that the kernel is trivial, note $e^x > x + 1$ for x positive, so no positive number is sent to 1, hence no negative number can be.

Remark 20.2. Justifying the notation e^x: set $e = E(1)$. Then by induction, it follows that $e^n = E(n)$ for all $n \in \mathbb{Z}$. Then $(e^{a/b})^b = e^a$, so we get the result for the rationals, and then $\mathbb{Q}$ is dense in $\mathbb{R}$, so this is good enough.

Some properties of e^x:

- Continuous and differentiable follows from analytic
- $\frac{d}{dx}e^x = e^x$ is clear
- e^x is strictly increasing since the derivative is $e^x > 0$ for all $x \in \mathbb{R}$.
- $e^{x+y} = e^x e^y$ is something we showed earlier
- $x^{-n}e^x \to \infty$ as x^∞ is obvious just by using the power series: $e^x > \frac{x^{n+1}}{(n+1)!}$, so $x^{-n}e^x > \frac{x}{(n+1)!} \to \infty$.

From the above, we've shown that e^x is a group isomorphism between $(\mathbb{R}, +)$ and $(\mathbb{R}^{>0}, \times)$. Thus there's an inverse map, which we'll call L, because secretly it's the logarithm.

Remark 20.3. Whenever we write log, we mean the natural log, or the log with base e. In CS, they will likely take the convention that all logs are base 2, but this is a math course.

To get information about/a construction of L, we'll pretend it exists and find some properties. Suppose L is differentiable. Then

$$E \circ L = L \circ E = \mathrm{id},$$

so if the derivative of L exists, we can use the chain rule to compute $\frac{d}{dy}L(y)$ as follows:

$$L(e^x) = x \implies L'(e^x)e^x = 1 \implies L'(y) = \frac{1}{y}.$$

Now using the fundamental theorem of calculus, we can construct L using the integral. We already know $L(1) = 0$, so this should be a complete definition.

Definition 20.4. The logarithm function L is given by

$$L(y) = \int_1^y \frac{dx}{x}$$

for $y > 0$.

In case $y < 1$, we should note that using $\int_a^c + \int_c^b = \int_a^b$ gives us a formal rule that $\int_1^y = -\int_y^1$, so we don't need to worry about what backwards intervals mean for integrals.

Thus $L'(y) = \frac{1}{y}$. Then by the chain rule, we should have

$$\frac{d}{dy}E(L(y)) = 1, \frac{d}{dy}L(E(y)) = 1,$$

and in the first map $1 \mapsto 1$, and in the second $0 \mapsto 0$. Thus, it follows that $E(L(y)) = y$, $L(E(x)) = x$.

We can also define various things now, for instance, we can define the function a^x for any a by

$$a^x = E(xL(a)).$$

Indeed, we can do weird things with this like compute $\frac{d}{dx}a^x = (\log a)a^x$.

Then we can try to work out a Taylor series, and get

$$L'(1+x) = \frac{1}{1+x} = 1 - x + x^2 - x^3 + ...,$$

so

$$L(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - ...$$

for $|x| < 1$.

21. Lecture 21 – 22 March 2017

21.1. **Fun with Exponential Definitions.** Recall from last time, Euler gave exponential versions of the definition for sine and cosine. Specifically,

$$\sin(x) = \frac{1}{2i}(e^{ix} - e^{-ix}),$$
$$\cos(x) = \frac{1}{2}(e^{ix} + e^{-ix}).$$

It's an immediate consequence of this definition that $\sin^2(x) + \cos^2(x) = 1$, from which it's immediate that $|\sin|, |\cos| \leq 1$ when working over $\mathbb{R}$. It's clear, using the Taylor series if necessary, that

$$\frac{d}{dx}e^{ax} = ae^{ax}$$

for every $a \in \mathbb{C}$, from which it follows $\sin'(x) = \cos(x), \cos'(x) = -\sin(x)$ just by checking our definitions.

By checking properties with the first few terms of the power series expansion, we can verify cosine vanishes first somewhere between $\sqrt{2}$ and 2. Then since we define $\frac{\pi}{2}$ to be the first positive 0 of cos, we have some bounds on the value of π. Continuing in this manner gives us a method for computing π, and not even a bad one, should we desire.

Consider $x \mapsto e^{ix}$ is a map $(\mathbb{R}, +) \to \{z \in \mathbb{C}^* \mid |z| = 1\}$. Then the kernel of this map is $2\pi\mathbb{Z}$, and it follows that the image is surjective on S^1 if you try hard enough. (Excercise: Do this.) In fact, the map $(\mathbb{C}, +) \xrightarrow{\exp} \mathbb{C}^*$ is surjective, with kernel $2\pi i\mathbb{Z}$. This actually follows immediately by writing

$$\exp(x + iy) = e^x e^{iy}$$

which can give us any positive number and any direction from the origin. That is, for $w \in \mathbb{C}^*$, we can find x with $|e^x| = |w|$, and y such that $e^{iy} = \frac{w}{|w|}$. Then $\exp(x + iy) = w$, which shows surjectivity.

Note the ambiguity that results from the fact that there's a copy of $\mathbb{Z}$ in the kernel of the exponential. This means there are infinitely many choices for the logarithm. Indeed, if you do a full circle around the origin, you'll find your logarithm has changed. We'll cover this in more detail when we get to the complex analysis.

21.2. **Inverse Trigonometric Functions.** We've done derivatives of inverse functions, so we can directly compute

$$\frac{d}{dx}\cos^{-1}(x) = -\frac{1}{\sqrt{1 - x^2}}$$
$$\frac{d}{dx}\sin^{-1}(x) = \frac{1}{\sqrt{1 - x^2}}$$

These results tells us we can use integral definitions locally, for instance,

$$\sin^{-1}(y) = \int_0^y \frac{dx}{\sqrt{1 - x^2}} \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right],$$

for $|y| \leq 1$.

21.3. **The Tangent.** We can define a function called the *tangent* as $\tan(x) = \frac{\sin x}{\cos x}$ where $\cos x \neq 0$, from which it's elementary to compute $\tan'(x) = 1 + \tan^2(x)$. Then it follows

$$\tan^{-1}(y) = \int_0^y \frac{dx}{1+x^2}.$$

Note that the power series

$$\frac{1}{1+x^2} = 1 - x^2 + x^4 - ...$$

converges for $|x| < 1$. Then we know

$$\tan^{-1} y = y - \frac{y^3}{3} + \frac{y^5}{5} - ...,$$

which converges for $|y| < 1$, but also at $y = 1$, where we can compute $\tan^{-1} 1 = \frac{\pi}{4}$. This is a consequence of Abel's Theorem, which we mentioned a few times ago. This gives the famous definition of $\frac{\pi}{4}$, though it's not a very efficient way of going about computing π.

Remark 21.1. There are ways to make this series converge more quickly by moving to the averages of alternating terms instead of just summing the terms, which goes much faster. In fact, repeating this procedure we continue to get faster approximations. Thus, it is possible to get an efficient way to compute π using this method.

There are some related ways to compute this, for instance, $\tan^{-1}(\frac{1}{2}) + \tan^{-1}(\frac{1}{3})$ and $4\tan^{-1}(\frac{1}{5}) + \tan^{-1}(\frac{1}{239})$ are both equal to $\frac{\pi}{4}$.

21.4. **Hyperbolic Trig.** Starting witht he definitions for $\sin, \cos$ from above, it's natural to consider these functions without the i. Then we get what are called the *hyperbolic sine* and *hyperbolic cosine*, given by

$$\sinh(x) = \frac{1}{2}(e^x - e^{-x})$$
$$\cosh(x) = \frac{1}{2}(e^x + e^{-x}).$$

These have a number of properties similar to those of sine and cosine, and do show up on occasion.

21.5. **Fourier Analysis in Four-ish Minutes.** Considering the functions $\sin nx, \cos nx$ for $n \in \mathbb{Z}^{\geq 0}$, these form an orthogonal basis of maps $\mathbb{R}/2\pi\mathbb{Z} \to \mathbb{C}$, with respect to the norm

$$\langle f, g \rangle = \frac{1}{2\pi} \int_0^{2\pi} f(x)\overline{g(x)} dx.$$

Then expressing functions in terms of this basis can be very useful. As the above hints, what we really want to do is move then to the expression of this as $\{e^{inx} \mid n \in \mathbb{Z}\}$, which is an orthonormal system. Then for any $f : \mathbb{R}/2\pi\mathbb{Z} \to \mathbb{C}$, we define

$$a_n = \langle f, e^{inx} \rangle,$$

with respect to the norm given above. Then assuming this converges (it turns out it does in the L_2 norm, but whatevs), we get the expression

$$f(x) = \sum a_n e^{inx}.$$

It results from this

$$\sum |a_n|^2 < \langle f, f \rangle.$$

We'd like to say that the expression for f as $\sum a_n e^{inx}$ converges absolutely, which requires the stronger constraing $\sum |a_n| < \infty$. If this holds, then the function

$$\varphi(x) = \sum_{n \in \mathbb{Z}} a_n e^{inx}$$

is absolutely uniformly convergent, and hence is continuous.

Theorem 21.2. *Given a continuous function f, with a_n as constructed above with $\sum |a_n| < \infty$, then $\varphi(x) = f(x)$.*

This is a much weaker version of the fundamental theorem of Fourier analysis, which gives the statement under much weaker conditions. However, it's nevertheless still a very useful theorem.

Example 21.3. Finite linear combinations

$$\left\{ \sum_{-N}^{N} a_n e^{inx} \right\} = \text{“trig functions”},$$

form an algebra, as they're closed under addition, subtraction, multiplication by scalars and multiplication. Moreover, it is closed under complex conjugation and contain $\mathbb{C}$. What's more, this algebra separates points. (To see this, note sin, cos separate points mod 2π.) Then let

$$f = \lim_{n \to \infty} P_n.$$

Then to see f, φ are the saem, you consider

$$\langle f - \varphi, f - \varphi \rangle,$$

and a uniform trigonometric approximation tells us the inner product is 0, which implies that φ converges to f.

22. LECTURE 22 – 24 MARCH 2017

22.1. **Derivatives in Multiple Variables.** Today we're moving on to multivariable calculus. In the single variable case, we had $f(x+h) = f(x) + f'(x)h + o(|h|)$. What would the corresponding thing be for the multivariable case? Let $E \subset U$ be an open subset of $\mathbb{R}^n$, or more generally any normed vector space, and take $f : E \to V$ be a map valued in some normed vector space V. Now, using the notation from above, we will have $x + h \in E$, $x \in E$, $h \in U$, and $f'(x) : U \to V$ a linear transformation.

Remark 22.1. We were able to avoid explicitly talking about $f'(x)$ being a linear transform in the single variable case, since $L(\mathbb{R}, V)$ is canonically isomorphic to V, with the isomorphism given by $L \mapsto L(1)$. (Equivalently–$\mathbb{R}$ represents the forgetful functor from FDVECT $\to$ SET.)

Picking a basis of U, we get $e_1 \mapsto v_1, ..., e_n \mapsto v_n$, and so $|T(\sum a_i e_i)| = |\sum a_i v_i| \leq \sum |a_i| ||v_i||$ is clearly bounded above for finite dimensional vector spaces, giving continuity of linear maps. Note we must be much more careful in the infinite dimensional case.

For each $x \in U$, $f'(x)$ is the linear transformation with takes $h \mapsto f(h)$. This is not saying it's an exponential–consider $f' : E \to \mathrm{Hom}(U, V)$, so it doesn't really make sense to ask if it's an exponential. It does, however, make sense to ask whether f' is itself continuous, differentiable, etc.

We should check that the derivative is unique: suppose A, B can be bet $f'(x)$, for $A, B \in \mathrm{Hom}(U, V)$. Then plugging this into our formula and taking their difference, we get

$$0 = (A - B)h + o(||h||).$$

From this, we want to deduce $A - B = 0$. Why is this true? Taking h nonzero, it follows that $(A-B)\frac{h}{||h||}$ is $o(1)$. We have $c(A-B)h = (A-B)(ch) = o(||ch||)$, so $|c|||A - Bh|| = o(|c|)$, by fixing h, and this is only possible if $||(A-B)h|| = 0$.

22.2. **Building Functions.** Given two differentiable functions, it's more or less clear that their sum will also be differentiable. We can't really say the same about the product, as we don't know how to multiply vectors.

Another clear method is composition. Given $f : E \to V, g : V \to W$, then we can ask whether the composite $g \circ f : E \to W$ is differentiable. We might expect to get some kind of chain rule, as we did in the 1 dimensional case. Indeed, there is a chain rule:

Proposition 22.2. *If g maps an open set containing $f(x)$ to W, g is differentiable at $f(x)$, and f is differentiable at x, then $g \circ f$ is differentiable at x, and*

$$(g \circ f)'(x) = g'(f(x))f'(x),$$

with $(g \circ f)'(x) : U \to W$.

Indeed, with this set-up, the proof is identical to the single variable case.

22.3. **Wrapping up some Loose Ends.**

Remark 22.3. Let U be a normed vector space, and U' be the same vector space considered under a different, equivalent norm. Then the "identity" map is continuous–and indeed, this statement is equivalent to the fact that the norms on U, U' are equivalent. For this reason, it doesn't matter much if we wish to change norms on our vector space.

Given a map $\mathbb{R}^m \to \mathbb{R}^n$, its derivative is valued in linear transforms. But linear transforms are just matrices, so we might ask what matrices they are, and what the numbers represent. We will study this next time, by introducing the notion of *partial derivatives*.

Remark 22.4. When defining $(fg)'$ in the single derivative case, we glossed over the fact that multiplication really is a map $\mathbb{R}^2 \to \mathbb{R}$. Thus, to verify that $(fg)'$ is differentiable, we need to verify this statement, in addition to f, g being differentiable. Luckily, we now have the machinery to deal with this.

22.4. **Inverse Function Theorem.** We've run a few times into the inverse function theorem–in the one variable case, this was relatively simple. In the multiple-variable case, it's clear at the very least that if f exists and is invertible, with its inverse differentiable, then f^{-1} should still have the inverse derivative of f. The converse is not quite true–but is close. We need to addd some conditions–specifically, that f is continuous differentiable, with invertible derivative on some neighborhood of x.

We might ask if this is actually very important–in the one-dimensional case, we were always able to get around this by defining our functions as integrals, if we wanted to. The answer is that we're eventually going to want to be able to do integrals as well. Doing a change of variables in the single-variable case was no different than the chain rule, when we looked closely. But in the mulitiple variable case, our rectangular partitions stop being rectangular under continuous maps. The solution is really nonobvious–we actually need the inverse function theorem to tell us that the image of our partition is actually "close enough" to rectangles that we can approximate it with parallelograms.

Definition 22.5. Given a map $f : X \to X$, we say $x \in X$ is *fixed* if $f(x) = x$.

One nice thing about fixed points is that, in the context of a complete metric space with a decreasing function, then f has a fixed point. (In fact, you proved this on the homework, showing there's a unique fixed point given a funciton f with $d(f(x), f(y)) < d(x, y)$, under a number of extra conditions.) We end up with

Theorem 22.6. (Contraction Mapping Theorem) *If there is $c < 1$ such that for every $x, y \in X$, $d(f(x), f(y)) \leq cd(x, y)$, then*

(1) there is at most 1 fixed point, and
(2) if X is complete, there is a fixed point x.

Proof. (1) is obvious. For (2), take arbitrary x_0, and $x_1 = f(x_0), x_n = f(x_{n-1})$. This gives us a sequence. Moreover, $d(x_n, x_{n+1}) \leq cd(x_{n-1}, x_n)$ just by definition, which implies that $\{x_n\}$ is a Cauchy sequence with limit x. Then since continuous functions preserve limits, we have x is a fixed point of f. □

23. LECTURE 23 – 27 MARCH 2017

23.1. **What's New in Multivariable?** One of the main differences now that we're working in multiple variables is that we have the inverse and implicit function theorem. These turn out to be essentially the same, luckily, which is why they're grouped together. We'll get to this Wednesday.

We also have partial derivatives, which commute with one another, under sufficient hypotheses. (e.g. $D_i D_j = D_j D_i$ on maps that are twice differentiable) This gives an obvious way that Taylor series will generalize.

For $E \subset \mathbb{R}^n$, $f : E \to \mathbb{R}$, we might want to find the maximum of f. We have $f'(x) \in V^*$, which is one way we can start testing. However, the very simple second derivative test in 1 dimension becomes a quadratic form $V \to \mathbb{R}$ in the multiple variable case. Now being positive definite corresponds to a maximum, and negative definite corresponds to a minimum.

23.2. **The Inverse Function Theorem.** Let X be a complete inner product space, $E \subset X$ open.

Remark 23.1. If $V \cong \mathbb{R}^n$, you can assume it's an inner product space by just picking your favorite inner product, since all inner products on $\mathbb{R}^n$ are equivalent.

Let f be a differentiable function $f : E \to V$. Fix $a \in E$, and set $b = f(a)$. Clearly we must insist $f'(a)$ is invertible, and the map $x \mapsto f'(x)$ is continuous. We have f' is a map $E \to \mathrm{Hom}(X, X)$. Assume that $A := f'(x)$ has continuous inverse A^{-1}, for x near A. Then there is $U \ni a$ open such that $f|_U$ is $1:1$.

Theorem 23.2. *Witht the set-up above, and letting $V = f(U)$, there is $g = f^{-1} : V \to U$ with $g(b) = f^{-1}(b) = a$, and g is continuous differentiable on V.*

Remark 23.3. By the chain rule $g'(f(x)) = (f'(x))^{-1}$.

By precomposing f by the translation $+a$, and postcomposing by the translation $-b$, followed by postcomposing by A^{-1}, we can assume that f sends 0 to 0, and the derivative at 0 is the identity. We aren't doing this, but it would make our lives easier if we did.

For y close enough to b, we need to construct $g(y)$. We have $g(y) \simeq a + A^{-1}(y - b)$ for y sufficiently close to b. Moreover, if x is sufficiently close to a, then $g(y) \simeq x + A^{-1}(y - f(x))$. (Here we're using $\simeq$ to denote "is close to.")

Remark 23.4. The above is a function, which should converge to the desired value for $g(y)$. Essentially, we want to mimic the proof of the contraction principle from last class.

Let $\varphi(x) = x + A^{-1}(y - f(x))$. We want to find U such that $\varphi|_U : U \to U$ is a contraction. The note a fixed point is equivalent to $0 = A^{-1}(y - f(x))$, which is equivalent to saying we've found x with $y = f(x)$, as desired. To show φ is a contraction mapping, we need to approximate $\varphi(x_2) - \varphi(x_1)$. We can compute directly

$$\varphi(x_2) - \varphi(x_1) = x_2 - x_1 - A^{-1}(f(x_2) - f(x_1)).$$

Theorem 23.5. (c.f. 9.19) *Suppose for $E \subset X$ and open set, $f : E \to Y$ differentiable, and $E \supset L'' = \{tx_1 + (1-t)x_2 \mid 0 \le t \le 1\}$, then we get a composite map $[0,1] \xrightarrow{F} Y$ by including $[0,1]$ as the line connnecting x_1, x_2. Then when Y has an inner product, F is bounded in the sense*

$$||F(1) - F(0)|| \le \max_{t \in (0,1)} ||F'(t)||.$$

Of course, we have $F(1) - F(0) = f(x_2) - f(x-1)$ by construction. Then it follows

$$||f(x_2) - f(x_1)|| \le \sup_{x \in L} ||f'(x)|| ||x_2 - x_1||.$$

This is the MVT for vector valued functions. Apply this not to $f(x)$, but to $f(x) - Ax$, where A is the linear approximation at a. Then we get

$$||f(x_2) - f(x_1) - A(x_2 - x_1)|| \le \sup ||f'(x) - A|| ||x_2 - x_1||.$$

This tells us that $A^{-1}(f(x_2) - f(x_1)) = A(x_2 - x_1) + E$, where $|E| \le \sup_{x \in L} ||f'(x) - A|| ||x_2 - x_1||$. Then if we're in a small enough neighborhood of a, $\sup_{x \in L} |f'(x) - A| < \epsilon$. Then we have $x_2 - x_1 - (x_2 - x_1) - A^{-1}E$ in our formula, so

$$||\varphi(x_2) - \varphi(x_1)|| < ||A^{-1}|| \epsilon ||x_2 - x_1||.$$

Then since f' is continuous, we can apply this on some sufficiently small neighborhood to get φ a contraction.

We do still need to check that $\varphi : U \to U$. Note that our bound is true for all x_2, x_1, so in particular, we can take $x_1 = a$. Then we get $||\varphi(x) - \varphi(a)|| \le c||x - a||$. If x is close to a, then $||\varphi(x) - a|| \le c||x - a|| + ||a - \varphi(a)||$. But the second term is equal to $A^{-1}||y - b||$. We can assume $c||x - a|| \le \delta$, and want to show $||\varphi(x) - a|| \le \delta$. By assuming $||y - b|| < \frac{\delta - c\delta}{||A^{-1}||}$, which we can do by shrinking if necessary, this follows. Thus, we get a neighborhood on which our contraction works. Now the preimage of an open set is open, so we've found a neighborhood of a on which our map is 1:1.

23.3. **Implicit Function Theorem.** Suppose you want to do calculus on, say, the sphere. We'd like to take some patch on the n-sphere, and identify it with a subset of $\mathbb{R}^n$. This is cut out a function, and we'd like to describe polynomials on this vanishing locus. Under sufficient conditions on our map to $\mathbb{R}^n$, we get something nice, which is the implicit function theorem.

If we had $V \times V \to V$ which takes $(x, y) \mapsto f(x) - y$, then the zero set of this is equivalent to just $f : V \to V$, so the inverse function theorem just falls out of the implicit function theorem.

24. LECTURE 24 – 27 MARCH 2017

We'll continue from last time–recall X is a finite dimensional inner product space, $E \subset X$ is open, $f : E \to V$ is continuously differentiable, $A = f'(a)$ is invertible (and in the infinite dimensional case, that A^{-1} is bounded). Last time we found a construction of a suitable inverse g of f such that given $y \in V$, V some neighborhood of $f(a)$, that $\varphi_y(x) = x + A^{-1}(y - f(x))$ is a contraction of U.

This shows that there is a unique fixed point, which gave our construction of g. To show that this is differentiable isn't that hard once we know it's continuous. The problem is that it's difficult to show that our map is continuous. Let $U \ni a$ be the small neighborhood of A we shrunk to.

Lemma 24.1. *$f(U)$ is open.*

Proof. Take $y_0 \in V$, then there's x_0 with $f(x_0) = y_0$ by the construction above. Take y close to y_0, then

$$||\varphi(x_0) - x|| = ||A^{-1}(y - y_0)|| \leq ||A^{-1}||||y - y_0||.$$

Since U is open, there's an open ball around x_0 contained in U, and in fact, we can assume $\overline{B_r}(x_0) \subset U$. Then we can ensure $||A^{-1}||||y - y_0|| \leq \frac{r}{2}$ by making $||y - y_0||$ sufficiently small. Then we also know

$$||\varphi(x_0) - \varphi(x)|| \leq \frac{1}{2}||x - x_0|| \leq \frac{r}{2},$$

i.e. we have $||\varphi(x) - x_0|| < r$. In particular, this means we've found a preimage of y inside $B_r(x_0) \subset U$, so we get an open ball around y_0 by using the ball of radius $\frac{r}{2||A^{-1}||}$. □

Remark 24.2. If f is a function such that for every $U \subset X$ is open, then $f(U)$ is open, we say that f is an *open function*. Thus, the above argument is saying that if $f'(x)$ is invertible for every $x \in E$, then f is an open map.

Lemma 24.3. *g is differentiable.*

Proof. As $y \to b$, we want to show $g(y) - g(b) \stackrel{?}{=} A^{-1}(y - b)$. We know $g(b) = a$, so we can rewrite this as $A(g(y) - a) \stackrel{?}{=} y - b + o(||y - b||)$. We know $y - b = f(g(y)) - f(a)$, and as $y \to b$, we know $g(y) \to g(b) = a$, so $y - b = A(g(y) - a) + o(||g(y) - a||)$. Then $||g(y) - a|| << ||y - b||$, since we can make our proportional bounds arbitrarily small, which is the proof. □

24.1. **Implicit Function Theorem.** We'll begin with a 2-variable example, then generalize to when x, y are vectors.

Example 24.4. Suppose you wanted to plot the function $x^3 + y^3 - 3xy = 0$. Near most points, y is a function of x. There are some problems near the origin, and it has a vertical derivative in a couple of places, but basically everywhere else, we don't have this problem. We can take the derivative of our function with respect to x to get $x^2 - y = y'(x - y^2)$, so

$$\frac{dy}{dx} = \frac{x^2 - y}{x - y^2},$$

with the exception of a couple points where this doesn't make sense.

In general, we have X, Y finite dimensional inner product spaces, $E \subset X \oplus Y$, and a map $f : E \to X$. Identify X with $\mathbb{R}^n$, Y with $\mathbb{R}^m$. Given a point in $X = \mathbb{R}^n$, we have n simultaneous functions which give us the vanishing of f, if we want to look at $f(a, b) = 0$ for some $(a, b) \subset E$.

If our function is differentiable at (a, b), we should have

$$f(a + h, b + k) = f(a, b) + (f'(a, b))(h, k) + o(||h|| + ||k||),$$

and by hypothesis, $f(a, b) = 0$. Then let $A = f'(a, b) \in \mathcal{L}(X \oplus Y, X) = \mathcal{L}(X, X) \oplus \mathcal{L}(Y, X)$, so we can decompose $A = A_x + A_y$. If we want our new point to vanish again, then we should be insisting $A_x h + A_y k = 0$. This rearranges to $h = A_x^{-1}(-A_y k)$. Then we get that, provided A_x is invertible, there's an adjustment we can make in x to compensate for the change in y.

Theorem 24.5. (Implicit Function Theorem) *If f' is continuous at (a, b), and the Jacobian is invertible, then there is an open neighborhood of (a, b) which is a product $U \times V$, and function from U_a to V_b such that $f|_{U_a \times V_b} = 0$ if and only if $g(x) = y$, where g is the inverse function.*

Note the inverse function theorem is a special case, where we take $X = Y$. Moreover, The implicit funciton theorem is a special case of the inverse function theorem applied to $X \oplus Y$, where we take $F(x, y) = (f(x, y), y)$, which is a map $E \to X \oplus Y$, and $f(x, y) = 0$ if and only if $F(x, y) = (0, y)$. Thus if we can invert $F(x, y)$ near (a, b), then we can apply the inverse function theorem to $(0, y)$, and we get some point (x, y), given y, where our function vanishes, which is exactly what we wanted.

It just remains to check the derivative of F. $F' : X \oplus Y \to X \oplus Y$, and we can decompose this into X and Y parts as a "block" matrix, by giving data in $\mathcal{L}(X, X), \mathcal{L}(Y, X)$, namely as

$$\begin{pmatrix} A_x & A_y \\ 0 & I \end{pmatrix}.$$

This is invertible if and only if A_x is invertible, so we can apply the inverse function theorem exactly in this cae, which is what we wanted. In fact, we can also compute the inverse matrix is

$$\begin{pmatrix} A_x^{-1} & T \\ 0 & I \end{pmatrix},$$

where $A_x T + A_y I = 0$, so $T = -A_x^{-1} A_y$.

24.2. **Directional Derivatives.** Another way to think of this is so-called directional derivatives. Given $f : X \to Y$, $f(a) = b$, we know that

$$f(a+h) + f(a) + (f'(a))h + o(||h||),$$

as $h \to 0$, if f is differentiable. But it's also reasonable to ask what this is doing specifically on a line that goes through a. Picking a vector v, and setting $h = tv$, we get

$$f(a+tv) = b + (f'(a)v)t + o(||tv||).$$

Then the middle term gives us the so-called direcitonal derivate, $D_v f$. If f is differentiable at a, then all of its directional derivatives exist. Moreover, $D_{v+w} = D_v + D_w$, and similarly $D_{\lambda v} = \lambda D_v$. However, you can also construction a function with basically anything you want as directional derivatives–$D_{v+w} = D_v + D_w$ doesn't hold at all in this case.

Example 24.6. Consider the function $f(r\cos\theta, r\sin\theta) = r\delta(\theta)$, where δ is some continuous function. Then all the derivatives exist and will correspond to $\delta(\theta)$. However, our function will be continuous, but very not differentiable at the origin. By choosing our function δ to take the kind of values we want it to take, we can make something silly with this.

Remark 24.7. We can do something similar to the above, but dependent on which parabola through the origin we're going through instead of through lines. This lets us do a thing.

If our function is defined on $\mathbb{R}^n$, where we start with a basis $\{e_i\}$, we will for notational convenience write $D_i f = D_{e_i} f = \frac{\partial}{\partial x_i} f$.

25. LECTURE 25 – 31 MARCH 2017

Beware the Kalends of April!

Take $E \subset \mathbb{R}^n$, and $f : E \to \mathbb{R}$. For the moment, take $n = 2$. Pick $(a, b) \in \mathbb{R}^2$, and suppose the rectangle with vertices $(a, b), (a + h, b), (a, b + k), (a + h, b + k)$ is contained in E. Let Q denote this entire rectangle. Then set

$$\Delta(f, Q) := f(a + h, b + k) - f(a + h, b) - f(a, b + k) + f(a, b).$$

Define a function $u(t)$, $u : [a, a + h] \to \mathbb{R}$ by

$$u(t) := f(t, b + k) - f(t, b).$$

Note by construction

$$\begin{aligned}\Delta(f, Q) &= u(a + h) - u(a)\\ &= h \cdot u'(t)\end{aligned}$$

by the mean value theorem, for some $t \in [a, a + h]$. Note by definition we have

$$u'(t) = D_1 f(t, b + k) - D_1 f(t, b),$$

which we can apply the mean value theorem to in the second variable if $D_2 D_1$ exists and is continuous, which we can assume if $f \in C^2$. Thus we can write

$$u'(t) = k D_2 D_1 f(t, u)$$

for some $u \in [b, b + k]$. Applying this to our earlier result, we get

$$\Delta(f, Q) = hk D_{21} f(t, u),$$

where $(t, u) \in Q$. And identical argument tells us this is also equal to $kh D_{12} f(t', u')$. It follows, then, for all h, k, there are pairs $(t, u), (t', u')$ suhc that $D_{12} f(t', u') = D_{21}(t, u)$. Then if $f \in C^2$, we can let $h, k \to 0$ and conclude $D_{12} f(a, b) = D_{21} f(a, b)$.

25.1. **Taylor in Multiple Variables.** Let's imitate what we did in one variable. Recall now $f'(0) : \mathbb{R}^2 \to \mathbb{R}$, it isn't just a real number. Then we get something that looks like

$$f(0) + (D_1 f(0) x_1 + D_2 f(0) x_2) + ...,$$

in order to match all of the first derivatives. To match the second derivatives, we have $D_1 D_1, D_1 D_2, D_2 D_1, D_2 D_2$. This is possible only if our function is C^2: if we let our quadratic term be $A_{11} x_1^2 + A_{12} x_1 x_2 + A_{22} x_2^2$, then we find $D_{11} = 2A_{11}, D_{12} = D_{21} = A_{12}, D_{22} = 2A_{22}$. Luckily $D_{12} = D_{21}$, by the above, so this is possible. Then we can plug this kind of stuff ina nd continue out the Taylor series in an intuitive manner.

$$f(t) + (D_1 f(0) x_1 + D_2 f(0) x_2) + \left(\frac{D_{11}}{2} f(0) x_1^2 + D_{12} f(0) x_1 x_2 + \frac{D_{22}}{2} f(0) x_2^2\right) + ...$$

Now in general, let $E \subset \mathbb{R}^n$, $f : E \to \mathbb{R}$. If $x \in E$, f has a *local maximum* if there is a neighborhook $U \ni x$ such that f achieves its maximum on U at the point x. In the one variable case, we found this just corresponds to $f'(x) = 0$, which remains what we want in the multiple variable case.

Recall the one variable case: Suppose $f'(0) = 0$ and f is C^2 near 0. Then $f(h) = f(0) + \frac{f''(0)}{2}h^2 + o(||h||^2)$, so $f''(0) > 0$ implies this is is a local min, and less than 0 it is a local max. We run into a problem in generalizing this, however: there are quadratic forms that are neither positive nor negative definite. For this reason, we have far more possibilities. In general, we have

$$f(h) = f(0) + Q(h) + o(||h||^2),$$

where Q is a quadratic form, given by the matrix $\frac{1}{2}(D_{ij})_{i,j}$. If Q is positive definite, then we have a local minimum, and negative definite a local maximum. If Q is singular, then the test has already failed. If $Q = 0$ exactly, then we can actually just go check the third derivative, and continue.

26. LECTURE 26 – 5 APRIL 2017

26.1. **Multivariate Integration.** Let R denote the rectangle

$$R = \{(x,y) \mid x \in [a_1,a_2], y \in [b_1,b_2]\}.$$

Given a continuous map $f : R \to \mathbb{R}$ (or in general, $f : R \to \mathbb{R}^n$, or really to any normed vector space), we will define

$$\iint_R f dxdy = \int_{y=b_1}^{b_2} \left(\int_{x=a_1}^{a_2} f(x,y)dx \right) dy.$$

We'd like to show that our inner integral is continuous, for this to make sense.

Note that our rectangle is a compact space, so continuity of f implies uniform continuity. Thus, given $\epsilon > 0$, there is δ such that $|x - x'| < \delta, |y - y'| < \delta \implies |f(x,y) - f(x',y')| < \epsilon$. Thus changing y by δ changes the value of f by at most ϵ, thus in our integral we get $\int f(x,y) - \int f(x,y') < \epsilon(a_2 - a_1)$, and ϵ is arbitrary.

Theorem 26.1. (Fubini's Theorem) *With the above set up, we have*

$$\iint_R f(x,y)dxdy = \int_{x=a_1}^{a_2} \left(\int_{y=b_1}^{b_2} f(x,y)dy \right) dx.$$

The result is fairly intuitive–they both should be calculating area, and we can just check that they do work with an arbitrary partition of our rectangle. However, there's a rather slick proof that avoids a lot of epsilon-delta work, which we give here. Let $\mathscr{A} = \{\text{cont. } f : R \to \mathbb{R}\}$. Then we've defined a linear map

$$f \mapsto \iint_R f(x,y)dxdy,$$

Linearity of this map is clear. Moreover, if $||f|| < \epsilon$, then we can check

$$|\int\int f(x,y)dxdy| \leq (b_2 - b_1) \sup_y |\int_{a_1}^{a_2} f(x,y)dx| \leq (b_2 - b_1)(a_2 - a_1)\epsilon,$$

hence our map is continuous. In the same way, we can check that the map

$$f \mapsto \int_x \left(\int_y f dy \right) dx.$$

Then to check that these are the same, it suffices to check that they're equal on a dense subset. However, $\mathscr{A}$ is an algebra, so we can check on a subalgebra $\mathscr{A}_0$ of functions of the form

$$\sum_1^k f_i(x) g_j(y),$$

with f, g continuous. That this is a subalgebra is quite immediate. This separates points and doesn't always vanish at any given point, so by Stone-Weierstrass, we have $\overline{\mathscr{A}_0} = \mathscr{A}$.

Then by linearity, it suffices to check elements of $\mathscr{A}_0$ that are single terms. But now we have the integral of $f(x)$ does not depend on y, nor $g(y)$ on x, so it's not hard to see

$$\int\int f(x)g(y)dxdy = \int f(x)dx \int g(y)dy.$$

This is clearly equal to the other order of integration as well, hence the result.

26.2. **Special Functions.** The beta function, denoted by a B rather than a β generally, was what Euler defined first,

$$B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt,$$

for $x, y > 0$. This turns out to be equal to

$$\frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)},$$

where Γ is the much more commonly used Γ function he found later. The Γ function is a bit weirder, though–it's only got one extra variable, but includes integrating to infinity. Specifically, we define

$$\Gamma(x) = \Pi(x-1) = \int_0^\infty t^{x-1}e^{-t}dt,$$

for $x > 0$. We can compute directly

$$\Gamma(1) = \int_0^\infty e^{-t}dt = \lim_{M\to\infty} e^0 - e^{-M} = 1.$$

Then

$$\Gamma(x+1) = \int_0^\infty t^x e^{-t}dt = -\int_0^\infty t^n d(e^{-t}) + \int_0^\infty e^{-t}d(t^x) = x\int_0^\infty t^{x-1}e^{-t}dt.$$

w to integrate
s fix above
on

Thenit follows that $\Gamma(x) = (x-1)!$ for $x \in \mathbb{Z}^{>0}$. Then we can do weird things like compute $\Gamma(\frac{1}{2}) = (-\frac{1}{2})!$ by plugging this into our equation. We get

$$\Gamma\left(\frac{1}{2}\right) = \int_0^\infty t^{-1/2}e^{-t}dt.$$

Substituting $t = u^2$, we can do a bit of rearranging to get $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

We can compute $\Gamma(x)\Gamma(y)$, and by Fubini's Theorem, this is equal to

$$\lim_{M,N\to\infty} \int_0^M \int_0^N t^{x-1}u^{y-1}e^{-t-u}dtdu.$$

Then instead of integrating one first instead of the other, we'd like to try integrating with respect to $t+u$. We could imagine first doing a linear transformation which changes the lines given by $t+u = k$ so that these become vertical lines. However, we're then no longer in a rectangle. We can surround our parallelogram by a rectangle, but then our function is no longer continuous. We can fix this by noting that there's some δ which bounds the area in the parallelogram within the total area, and then taking the limit as our rectangels get very small, we're able to do the integral. We get

$$\int_0^\infty e^{-t'} \left(\int_0^{t'} (t' - u) u^{y-1} du \right) dt'.$$

If we do a linear change of variables $u = t'v$, then do some arithmetic, we find

$$\Gamma(x)\Gamma(y) = B(x,y)\Gamma(x+y),$$

which shows the relation we stated above.

27. Lecture 27 – 07 April 2017

Let $E \subset V$ be a subspace of a vector space over $\mathbb{R}$.

Definition 27.1. We say E is *convex* if $x, y \in E$, $t \in [0,1]$ implies $(1-t)x + ty \in E$.

We say the *convex hull* of a set of points $x_1, ..., x_n$ is the set of points of the fom $\sum t_i x_i$, where $t_i \geq 0, \sum t_i = 1$. It's immediate that given any set of points $\{y_1, ..., y_m\} \subset E$ in a convex space E, their convex hull is also contained in E.

Definition 27.2. If E is convex, we say $f : E \to \mathbb{R}$ is a convex function if

$$\{(v, y) \in E \times \mathbb{R} \mid y \geq f(v)\}$$

is convex. (We could also say $y > f(v)$ if we so chose.) Equivalently, for $a, b \in E$, this means

$$(1-t)f(a) + tf(a) \geq f((1-t)a + ty).$$

Remark 27.3. If E is nice–e.g. open or closed, V normed–and f is continuous, then it is equivalent to only insist on $t = \frac{1}{2}$ working, as this gives us every linear combination of $\frac{1}{2^k}$, which gives us a dense subset of the unit interval. We will say E is *midpoint convex* if it satisfies this condition.

Example 27.4. Here are some convex functions:

- Any constant or linear function
- Any upward-facing parabola–e.g. $y = x^2$. In arbitrary dimension, any positive semidefinite quadratic form will be convex.
- Given any collection $\{f_i\}$ of convex functions, any positive linear combination of the $\{f_i\}$ will also be convex
- On the positive reals, it certainly looks like $-\log x$ should be as well. It's continuous, so it suffices to check at midpoints. Then we have

$$\log\left(\frac{x+y}{2}\right) \overset{?}{\geq} \frac{1}{2}\log x + \frac{1}{2}\log y.$$

  Rearranging and using our log properties, this reduces to $\frac{x+y}{2} \geq \sqrt{xy}$, which is true by the AM-GM inequality.

For $f > 0$, we will sometimes say that f is *logarithmically convex* if $\log f$ is convext, or equivalently, if

$$f((1-t)x + ty) \leq f(x)^{1-t} f(y)^t,$$

which is a stronger condition than convexity, again by AM-GM.

Recall from last class the Gamma function

$$\Gamma(x) = \int_0^\infty t^x e^{-t} \frac{dt}{t},$$

for $x > 0$. Then it's not hard to check that Γ is midpoint logarithmically convex: indeed, for any function F of the form

$$F(x) = \int_{t_1}^{t_2} ((A(t))^x B(t) dt,$$

with $A, B > 0$, we want to show

$$F(x+y)F(x-y) \geq ((F(x))^2,$$

in order to get F logarithmically midpoint convex. We can express this as

$$\int A^{x+y} B \int A^{x-y} B \geq \left(\int A^x B \right)^2.$$

We can arrange this to be a special case of the Cauchy-Schwarz identity. On the LHS we have

$$||A^{\frac{1}{2}(x+y)} B^{1/2}||^2 ||A^{\frac{1}{2}(x-y)} B^{1/2}||^2 = \left(A^{\frac{1}{2}(x+y)} B^{1/2}, A^{\frac{1}{2}(x-y)} B^{1/2} \right),$$

and then we can just apply Cauchy-Schwarz for integrals. (How do we prove Cauchy-Schwarz for integrals? Basically the same way we did in the finite dimensional case. Left as an exercise.)

From this, we can immediately deduce the following Lemma, using continuity of log, Γ:

Lemma 27.5. *$\Gamma(x)$ is logarithically convex, and hence convex.*

As a result, we have $\Gamma(x)\Gamma(x+2) = (x-1)!(x+1)!$, and $\Gamma(x+1)^2 = (x!)!$, and the lemma tells us that the first expression is $\geq$ the second, which of course is true, since $x+1 > x$.

Theorem 27.6. *Given a function $f : (0, \infty) \to \mathbb{R}^{>0}$ such that*
- $f(x+1) = xf(x)$
- $f(1) = 1$
- $\log f$ *ix convex*

then $f(x) = \Gamma(x)$.

Intuitively: as we noted above, we have $x+1 > x$, and this is related to the convexity of Γ. Then as x grows arbitrarily large, the ratio of this difference goes to 1. In particular, for every ϵ there is an n such that $x > n$ implies

$$|f(x) = \Gamma(x)| < \epsilon.$$

But then they continue differing by at most ϵ everywhere, so we must have that they're the same function.

Proposition 27.7. *We have what amounts to a product formula:*

$$\Gamma(x) = \lim_{n\to\infty} \frac{n!n^x}{x(x+1)...(x+n)}.$$

To see this, we can check that the limit satisfies the conditions of Therem 27.6, which is not hard to do.

Corollary 27.8. *We have*

$$\Gamma(x)\Gamma(1-x) = \lim \frac{(n!)^2 n}{x(x+1)...(x+n)(1-x)(2-x)...(n+1-x)},$$

which we computed on the problem set is $\frac{\pi}{\sin \pi x}$. *In particular, we get* $\Gamma(\frac{1}{2})^2 = \pi$, *which we already know.*

Note $\Gamma(x)\Gamma(1-x) = B(x, 1-x)$, where B is the Beta function we defined last class.

Corollary 27.9. *The so called duplication formula gives us*

$$\Gamma\left(\frac{x}{2}\right)\Gamma\left(\frac{x+1}{2}\right) = \frac{2^{x-1}}{\sqrt{\pi}}\Gamma(x).$$

Let $F(y) = \int_{-\infty}^{\infty} e^{-x^2} \cos xy dx$. We know $F(0) = \sqrt{\pi}$, and we claim $F(y) = e^{-y^2/4}\sqrt{\pi}$. We'll give a quasi-proof here, where we take some derivatives that we don't actually know we can do, and justify next class: taking $\frac{d}{dy}F(y)$, we get $\frac{1}{2}\int_{-\infty}^{\infty} \sin xy d(e^{-x^2}) = \frac{1}{2}\int_{\infty}^{\infty} e^{-x^2} d(\sin(xy)) = -\frac{y}{2}F(y)$.This leaves us with a differential equation

I got this alge- ong idk

$$\frac{dF}{dy} = -\frac{y}{2}F(y).$$

Then by uniqueness of solutions of differential equations, we can just check that our proposed function satisfies this differential equation, which it does.

28. LECTURE 28 – 10 APRIL 2017

28.1. **Commuting Derivatives and Integrals.** At the end of last clas, we said that we could interchange an integral and derivative in order to do a computation. We haven't justified that yet, though, so we'll do it now: Suppose we have some box $B \subset \mathbb{R}^2$, and $f : B \to \mathbb{R}$ continuous. Then Fubini's theorem tells us it doesn't matter whether we integrate with respect to x first, or y first. Consider

$$F(y_2) = \iint_B f(x,y)dxdy = \int_{y_1}^{y_2} \int_{x_1}^{x_2} f(x,y)dxdy,$$

where the box B along the y axis goes from $[y_1, y_2]$ and along the x axis goes along $[x_1, x_2]$. Then by the Fundamental Theorem of Calculus, we have

$$F'(y_2) = \int_{x_1}^{x_2} f(x, y_2)dx.$$

Again by the Fundamental Theorem of calculus, if we let $\varphi(x, y_2) = \int_{y_1}^{y_2} f(x,y)dy$, then we get $\frac{\partial}{\partial y}\varphi(x, y_2) = f(x, y_2)$. Now we can write

$$F(y_2) = \int_{x_1}^{x_2} \varphi(x, y_2)dx$$

by Fubini's theorem. Then the derivative of this is exactly what you'd expect: it's equal to the integral of the derivative with respect to y of φ.

28.2. **Change of Variables: When the moon hits your eye like the $\sqrt{\pi}$, that's a polar(é).** We've already shown a couple of times that

$$\int_{-\infty}^{\infty} e^{-x^2}dx = \sqrt{\pi}.$$

- The first way we showed this was via a computation of $\int_0^{\pi/2} \cos^2\theta d\theta$.
- The second was by noting that $(\Gamma(\frac{1}{2}))^2 = B(\frac{1}{2}, \frac{1}{2})$ was related, and equal to π.

There's also a very classical proof of this, which goes as follows: consider

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-x^2-y^2}dxdy = \int_{r=0}^{\infty}\int_{\theta=0}^{2\pi} e^{-r^2}rdrd\theta.$$

Letting $u = r^2$, we get this is equal to

$$\int_0^{2\pi} \frac{1}{2}\int_0^{\infty} e^{-u}dud\theta = 2\pi \cdot \frac{1}{2} = \pi.$$

We didn't argue this way because we haven't actually justified integrating in polar coordinates. We can set $x = r\cos\theta, y = r\sin\theta$, then find $dx = \cos\theta dr - r\sin\theta d\theta, dy = \sin\theta dr + r\cos\theta d\theta$, then we can compute $|dx \wedge dy| = |-rdr \wedge d\theta| = rdrd\theta$. However, we should really justify this kind of change of variables works in general. (This was done with pictures, I couldn't really translate it.)

28.3. **Partitions of Unity.** (You should actually learn Stokes', folks–it's important and I'm not sure what class it's supposed to be taught in.)

Let K be compact in $\mathbb{R}^n$ with open cover $\cup_\alpha V_\alpha$. We want to find a finite subset of functions $\psi_1, ..., \psi_s$ in $\mathcal{C}^\infty(\mathbb{R}^n)$ which are supported on V_{α_i}: that is to say, $x \notin V_{\alpha_i} \implies \psi_i(x) = 0$. Moreover, we want to require $0 \leq \psi_i \leq 1$, and $\sum \psi_i(x) = 1$ for every $x \in K$.

Remark 28.1. What does this have to do with integration? Given a partition of unity, and a function f defined on K, we can write $f = \sum f\psi_i$. Now, however, each term is supported only on V_{α_i}, so we can just work on small neighborhoods to talk about our function/integrate it.

Construction: for each $x \in K$, choose α such that $x \in V_\alpha$. Recall we have functions that look like

$$f(x) = \begin{cases} e^{-\frac{1}{x^2}} & x > 0 \\ 0 & x \leq 0, \end{cases}$$

which are smooth, and quickly interpolate between 0 and 1. There are balls B, W centered on x such that $\overline{B} \subset W, \overline{W} \subset V_\alpha$. Note $\{B_x\}_{x \in K}$ is an open cover of K, so we get a finite subcover. For each B_i in the subcover, let φ_i be 1 on B_i, 0 outside W_i, and smoothly interpolated between these by using a function that looks like f above (note f doesn't quite work–see the website for some actual examples of $\mathcal{C}^\infty$ bump functions), and rotating it around the ball. Then note φ_i is not supported outside V_{α_i}.

This is almost what we want, but not quite–it sums to something, but not to 1 necessarily. On K we could just divide out by the sum, but then we're not sure what to do outside of K then. Therefore, we take a slightly different approach:

- $\psi_1 = \varphi_1$
- $\psi_2 = (1 - \varphi_1)\varphi_2$
- $\psi_3 = (1 - \varphi_1)(1 - \varphi_2)\varphi_3$
- ...
- $\psi_s = (1 - \varphi_1)...(1 - \varphi_{s-1})\varphi_s$.

Clearly ψ_i remains supported only on V_{α_i}, and moreover, the sum is always going to be 1, just by checking the algebra on each B_i.

29. Lecture 29 – 12 April 2017

Suppose we have a continuous function f, with compact support on $\mathbb{R}^k$. Suppose $T : E \to \mathbb{R}^k$, where $E \subset \mathbb{R}^k$, and the image contains the supported area of f. Then we claim

$$\int_{\mathbb{R}^k} f(T(\lambda)|\det T'(x)|dx = \int_{\mathbb{R}^k} f(y)dy.$$

We can think of $|\det T'(x)|$ as $|\frac{dy}{dx}|$, though this is only really true in our pictures, to get the usual notational system for change of variables.

There are some maps we've already done for this: for instance, of T permutes coordinates, then this is just the Fubini theorem. We can also change one variable at a time: In 10.7 of Rudin, choose m, and consider $\{x_1, ..., x_n\} \mapsto \{x_1, ..., x_{m-1}, y_m, x_{m+1}, ..., x_n\}$, where $y_m = y_m(x_1, ..., x_n)$ is at least $\mathcal{C}^1$. By simply writing this out in a matrix, we see that the determinant of this change is

$$\det(\{x_1, ..., x_n\} \mapsto \{x_1, ..., x_{m-1}, y_m, x_{m+1}, ..., x_n\}) = \frac{\partial}{\partial x_m} y_m.$$

We will need to therefore require that this is nonzero.

Idea: we can now switch one variable at a time, until we've switched all the x_i to y_i. By pre- and post-composing by translations, we may assume that our neighborhood is centered at 0. Then we will possibly need to permute coordinates such that the dependence of y_n is actually on x_n, and so on. Then an inductive-style argument: given our matrix of transitions from x_i to y_i, we want to find a bi-codimension (1, 1) minor, so we can apply our inductive step. We will need our minor to not have dependence on the outside column, and outside row, however. To see this is possible, just note that the entire thing having invertible derivative shows the same for minors, and then we can do something with matrices.

probably shoul
ally explain thi

29.1. Get Stoked.

In the one variable case, we showed

$$\int_a^b F'(t)dt = F(b) - F(a) = -\int_b^a F'(t)dt = -(F(b) - F(a)).$$

This tells us that, in the general case, we might be able to evaluate an integral just by looking at things on the boundary.

Take $E \subset \mathbb{R}^n$, and $F : E \to \mathbb{R}$ a $\mathcal{C}^1$ function. For a path γ contained in E, we will argue $\int_\gamma dF = F(\gamma(b)) - F(\gamma(a))$.

Definition 29.1. A path is the image of an interval $[a, b]$ under a continuous (usually we will assume $\mathcal{C}^1$) map.

To define an integral along a path, we will say

$$\int_\gamma dF := \int_a^b (F \circ \gamma)' dt.$$

By the chain rule, $(F \circ \gamma)' = F'(\gamma(t)) \circ \gamma'(t)$. Since we're shamelessly using coordinates right now, we can further express this as the integral over

$$\sum_{i=1}^{n} \frac{\partial}{\partial x_i} F(\gamma(t)).$$

We will denote the summands as $\gamma_i'(t)$. Thus

$$\begin{aligned} \int_\gamma dF &= \int_a^b (F\gamma)' dt \\ &= \int \sum \frac{\partial}{\partial x_i} F(\gamma(t)) \gamma_i'(t) dt. \end{aligned}$$

What we're starting to see here is a *differential 1-form*: formally, let $\omega = \sum_i f_i dx_i$. Then we're defining

$$\int_\gamma \omega := \sum_i \int f_i \gamma_i'(t) dt.$$

This is a really nice construction: you can (and you should!) check that if we precompose our map by something so that it's coming from a different interval, say, then the integral remains the same. We'll define a bit more notation now: let

$$dF := \sum D_i f dx_i,$$

then our above claimed integral makes sense. That is, we can really think of $\int_\gamma dF$ as the integral of F over the path γ.

Just like we can add functions, we can compose paths, if their endpoints meet. It's clear that the integral of the composite is equal to the sum of the integrals, just by plugging in to our formula and noting that the middle points cancel. This is highly suggestive of the case where we have a closed loop. For simplicity, think of a rectangle B, and integrate some arbitrary differential 1-form $\omega = f_1 dx_1 + f_2 dx_2$, with f_1, f_2 as $\mathcal{C}^1$ functions on some neighborhood of B. Then we see

$$\int_{\gamma_1 + \gamma_3} \omega = \int_{x_1}^{x_2} f_1(x, y_1) dx - \int_{x_1}^{x_2} f_1(x, y_2).$$

But then we already computed something like this–it's just equal to

$$\iint_B \frac{\partial f_i}{\partial y} dx dy.$$

The same is true with $\gamma_2 + \gamma_4$, which tells us that

$$\int_{\partial B} \omega = \iint_B (D_1 f_2 - D_2 f_1) dx \wedge dy.$$

We know that this should be zero, though, as everything should cancel out. Then if $f_1 = D_1 F, f_2 = D_2 f$, and F is a $\mathcal{C}^2$ function, then we have $\omega = dF$, and $d(d\omega) = 0$, so we've found another proof that this integral is 0.

30. LECTURE 30 – 14 APRIL 2017

As discussed last class, we can proceed from integrating on boxes to any object that we can partition into boxes. Then any interior edges will cancel with each other, and we just get an integral around the boundary. This leads us to think that we have some kind of pairing $(B, \omega) \to \mathbb{R}$, where we have linear combinations of boxes B, and one forms ω, and $\int_{aB+a'B'} \omega := a \int_B \omega + a' \int_{B'} \omega$, and similarly for ω. Similarly, we will think about ∂B, the boundary of B, as a linear combination of $\mathcal{C}^1$ paths. Then we've already verified

$$\int_{\partial B} \omega = \int_B d\omega,$$

where ∂ takes us from linear combinations of boxes to linear combinations of paths, and d takes us from 1-forms to 2-forms.

Remark 30.1. The fundamental theorem of calculus in a single variable can be restated as

$$\int_\gamma dF = \int_{\partial\gamma} F,$$

where we define the integral at a point p to be evaluation at that point.

Note both d, ∂ have the property that their square is 0. The fact that $d^2 = 0$ is very interesting: it tells us we should ask whether the map forms an exact sequence, or if there's something else going on. Perhaps surprisingly, this depends on the area we're working over. For instance, if we're working over a convex space, we can define

$$F(b) = F(a) + \int_a^b dF,$$

where the integral is taken over an arbitrary path from a to b.

Lemma 30.2. *This is well defined.*

Proof. Given any two paths, the difference between them can be viewed as the integral over a loop that goes back to the point a. Then we have a loop is the boundary of some object S, and so we have

$$\int_{\partial S} dF = \int_S d^2 F = \int_S 0,$$

hence this is well defined. ☐

Then given any 1-form, this gives us a 2-form so long as it isn't in the image of a 0-form, hence exactness.

30.1. Important Special Case.

Identify $\mathbb{C} \cong \mathbb{R}^2$ in the usual way, and suppose we have a $\mathcal{C}^1$ map $w(x + iy) = u + iv$. By Problem set 8, problem 8, we know w is differentiable as a complex variable if and only if $w' : \mathbb{R}^2 \to \mathbb{R}^2$ is $\mathbb{C}$-linear. An equivalent condition is $\frac{\partial v}{\partial y} = \frac{\partial u}{\partial x}$, and $\frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}$. Now consider

$$\int_\gamma w(z)dz = \int_\gamma (x + iv)(dx + idy).$$

Then $\operatorname{Re}\omega = udx = vdy$, and $\operatorname{Im}\omega = udy + vdx$. Then this corresponds exactly to our conditions on derivatives above. Thus if we're working in a convex space (or the image of a convex space under any function), this is the differential of some function. In particular, this means that the integral on the boundary of any space on which our function is defined will be 0. This gives us an incredibly powerful tool to study complex analysis. It also shows a problem with spaces that aren't convex (or more precisely, the problem is when our space isn't simple connected).

Example 30.3. Consider a square centered on the origin, with the origin excised. Then we can define $z \mapsto \log z$. Then $d \log z = \frac{dz}{z}$, so we can try to define the logarithm the way we did above, as just the integral along a path. However, we will find

$$\int_\gamma \frac{dz}{z} = 2\pi i$$

when γ is a path around the origin, rather than being 0. The reason our argument doesn't work is that our space isn't the boundary of something–it's missing a point!

In general, if we take the integral around the origin of something like $\frac{f(z)}{dz}$, we will get $2\pi i f(0)$.

31. Lecture 31 – 17 April 2017

Last time we spent awhile discussing the Cauchy Integral Formula. A nonempty open set in $\mathbb{C}$ is often called a region. Then we were focused on $f : \{|z - z_0| < R\} \to \mathbb{C}$. Then taking $r < R$, we applied Green's Theorem to show

$$\frac{1}{2\pi i} \oint_{|z-z_0|=r} f(z) \frac{dz}{z - z_0} = f(z_0).$$

A lot of facts from complex analysis follow from this formula. Indeed, using the power series expansion of $\frac{1}{z-z_0}$, we get

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n,$$

where

$$a_n = \frac{1}{2\pi i} \oint_{|z-z_0|=r} \frac{f(z)dz}{(z - z_0)^{n+1}}.$$

Theorem 31.1. *Every continuously complex differentiable function is analytic.*

Proof. We have a power series decomposition of any complex differentiable funciton given above. Then we have

$$|a_n| \leq \frac{1}{2\pi} \cdot 2\pi r \cdot A \frac{1}{r^{n+1}},$$

where $A = \sup_{|z-z_0|=r} |f(z)|$ is some constant, which by direct computation is bounded by inverse powers of r. Thus our power series makes sense on the region, and hence f is analytic. □

Suppose we have a function defined on some region, and that it extends to a slightly larger region. Then this extension is actually unique! (Do you see why?) Radii of convergence now actually necessarily find barriers to convergence, which they didn't necessarily in the real case, as well.

We can rewrite the Cauchy Integral formula as

$$f(z_0) = \int_0^{2\pi} f(z_0 + re^{i\theta}) d\theta.$$

(Exercise: check this by rewriting the Cauchy Integral Formula!) In particular, f is the average value on any circle around it. This leads us to what's called the Max/Min principles.

We can consider the function $z \mapsto |f(z)|$. Suppose $|f|$ attains its maximum at some point z_0. Then $f(z_0)$ is equal to the average value on some circle containing z_0, if z_0 is an interior point. But then if there's any variance from the norm of z_0, then there's a point with higher norm. (Rigorize this argument.) This leads us to

Lemma 31.2. (Max/Min Principle) *If f attains a local maximum, then it's a constant function.*

If $f(z_0) \neq 0$, then $\frac{1}{f}$ is analytic in a neighborhood of z_0. Then applying the maximum principle for $\frac{1}{f}$ gives us the minimum principle, with the exception that there can be a minimum if $f(z_0) = 0$.

Corollary 31.3. (Fundamental Theorem of Algebra) *Every polynomial has a zero.*

Note that polynomials eventually are dominated by their highest order term, so if a polynomial f has no zeroes, then $\frac{1}{f}$ is entire with local maxima, since it goes to 0 as $|z|$ gets large, which is a contradiction.

We say f is entire if it is a complex differentiable map defined on all of $\mathbb{C}$.

Theorem 31.4. (Liouville) *A bounded entire function is constant.*

Proof. Suppose $f : \mathbb{C} \to \mathbb{C}$ is differentiable, and there's some B such that for every z, $|f(z)| < B$. □

Proof. It suffices to check that $f'(z) = 0$. We have a formula for f' from the Cauchy integral formula. In particular, if we let $f'(z_0) = a_1$, then $|a_1| < \frac{B}{r}$ for every $r \in \mathbb{R}^+$, so taking r arbitrarily large we must get $a_1 = 0$. □

If more generally $|f(z)| < B|z|^\alpha$, then $f(z) = \sum a_n z^n$, and $|a_n| < B\frac{r^\alpha}{r^n}$, then $a_n = 0$ for every $n > \alpha$. This makes f a polynomial of degree at most α.

A somewhat less obvious trick of a similar nature is the following: suppose we have some entire function f, and it doesn't take the value w. Then we can consider $\frac{1}{f(z)-w}$ is also an entire function. If $|f(z) - w| > r$ for some r, then this would be a bounded entire function, so it follows that f gets arbitrarily close to w, and hence must have values dense in $\mathbb{C}$. A more difficult thing to prove is the little Picard theorem, that this can miss at most one point.

Proposition 31.5. *There is at most one way to extend any holomorophic function to a larger neighborhood.*

Proof. Let $f : E \to \mathbb{C}$ be analytic. If $z_n \to z$, then $f(z_n) = 0$ implies f is 0 in aa neighborhood of z. To see this, we know $f(z) = 0$, since continous functions commute with limits. Then we can define $f_1(w) = \frac{f(w)}{w-z}$, and this still vanishes at all the z_n. Then define $f_2(w) = \frac{f(w)}{(w-z)^2}$. Then this gives us a power series for some analytic function g, and we get $f(z_n) = g(z_n) \implies f = g$. That is, if two analytic functions agree on infinitely many points, then they must be equal. □

We can apply this to the Γ function: recall

$$\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt.$$

We've defined this so far for $z \in \mathbb{R}$. If we allow z to be a complex number, we know there's at most one extension to $\mathbb{C}$. We do need to define $t^{z-1} = \exp((z-1)\log t)$, where we know how to take the log of a positive real number. We can check that this still converges, and it's immediate that $z\Gamma(z) = \Gamma(z+1)$. Then using the fact that $\Gamma(z) = \frac{\Gamma(z+1)}{z}$, we get a way to extend Γ to all of $\mathbb{C}$, except for the nonpositive integers, since there were would be dividing by 0.

We will continue studying a complex differentiable function on some ball around a point z_0. Recall

$$f(z_0) = \frac{1}{2\pi i}\oint_{|z-z_0|=r} f(z)\frac{dz}{z-z_0}.$$

We had noted that this integral was therefore invariant as we decrease r, as a consequence of Green's theorem, so it makes sense that the integral converges to the value $f(z_0)$ as $r \to 0$.

Suppose f is analytic on $0 < |z - z_0| < r$. Of course, we know

$$|\oint_{|z-z_0|=r_1} f(z)dz| < 2\pi r_1 \max f(z).$$

If this approaches 0 as $r_1 \to 0$, then we can conclude $\oint_{r_2} f(z)dz = 0$.

Suppose $f(z)$ is bounded as $z \to z_0$ (we are not, however, assuming that the limit is well defined). Then we can consider $\frac{1}{2\pi i}\oint f(z)\frac{dz}{z-z_0}$. Then, as we did last class, we can from this create a Taylor series, which defines a function g. We would like to show this coincides with our original function. Instead of considering f, consider the function $(z-z_0)f$. Then we have $(z-z_0)f = \sum b_i(z-z_0)^n$, and f bounded as $z \to z_0$ implies $b_0 = 0$. It follows that $f = \frac{b}{z-z_0} +$ something analytic.

Now we can weaken this by assuming $f(z) << \frac{1}{|z-z_0|^k}$. Then $(z-z_0)^k f(z)$ is bounded, which implies $f(z) = \sum a_n(z-z_0)^n$, except now n is allowed to start at $-k$. Of course, this won't work for z_0, but will work everywhere else. We will say in this case that f has a pole of order k at z_0 (assuming $a_k \neq 0$). As a note on terminology, a simple pole is a pole of order 1, and a pole of order zero means that the function is analytic at the point. If there is no minimal k with $a_k \neq 0$, we say there's an essential singularity at z_0. (For example, consider $e^{\frac{1}{z-z_0}}$.) We call these function with poles meromorphic functions, so long as the set of singularities is discrete.

Remark 32.1. Both the holomorphic and meromorphic functions form a ring–and in fact, both have the structure of an algebra. However, it's immediate that we can divide meromorphic functions (other than the zero function), just by their construction. From this, it follows that meromorphic functions form a field.

This allows us to slightly formalize our previous discussion of the Gamma function. In particular, we know that there are poles at the nonpositive integers, all of which are order 1.

In the real case, we know lots of diffeomorphic functions $[0,1] \to [0,1]$. However we have far fewer examples of biholomorphic (holomorphic bijective maps, whose inverses are holomorphic) maps from the unit ball in $\mathbb{C}$ back to $\mathbb{C}$.

Theorem 32.2. (Schwarz Lemma) *Let $D = B(0,1)$. The only biholomorphic functions from $D \to D$ that send 0 to 0 are multiplication by a complex number of norm 1. In fact, we will prove the slightly stronger claim that, given a holomorphic map $D \xrightarrow{f} D$ which takes 0 to 0, we have*

$|f(z)| \leq |z|$, and equality for any $z \neq 0$ implies there is $c \in \mathbb{C}$ with $f(z) = dz$. Moreover, $|f'(0)| < 1$.

Proof. Given f, consider the function $\frac{f(z)}{z}$. We claim this is still a map $D \to D$. Fix $r < 1$, and consider $\frac{f(z)}{z}$ on the circle $|z| = r$. Then $|\frac{f(z)}{z}| < \frac{1}{r}$, and by the maximum principle, this also holds on the interior of the circle. This is true for every $r < 1$, so letting $r \to 1$, we get immediately $|f(z)| \leq |z|$.

In the equality case, suppose we have a point z_0 where the norms are equal. Then by the maximum modulus principle, we either have the function continues increasing, or is locally constant. Thus $|f(z)| = |z| \implies \frac{f(z)}{z}$ is constant.

Then we get the result on $f'(0)$ just by knowing it's the limit of $\frac{f(z)}{z}$ as $z \to 0$, and using the above. $\square$

33. LECTURE 33 – 21 APRIL 2017

"Does anyone know what the Riemann Hypothesis is?"

The Riemann ζ function is

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

The Riemann Hypothesis conjectures that all (nontrivial) zeroes of the zeta function have real part $\frac{1}{2}$. We would really like to verify this is holomorphic on some region so that we can do complex analysis with it. However, we mentioned last class that the uniform limit of holomorphic functions is holomorphic. So long as we keep the real part of s greater than 1, we have that this sum uniformly converges–in particular, we have uniform convergence of a right half plane.

Proposition 33.1. *Given $E \subset \mathbb{C}$ and $f_1, f_2, ...$ a sequence of maps $E \to \mathbb{C}$ converging uniformly to a function f, then if each f_n is analytic, then f is analytic, and for every $z_0 \in E$, and each $k = 0, 1, 2, ...$, the $(z - z_0)^k$ coefficient of f is $\lim_{n\to\infty}$ of the $(z - z_0)^k$ coefficient of f_n.*

Proof. Given $z_0 \in E$, there's some $r > 0$ such that $\overline{B_r}(Z_0) \subset E$. Then

$$f_n(z_0) = \frac{1}{2\pi i} \oint_r \frac{f_n(z)dz}{z - z_0},$$

and in general

$$a_k = \frac{1}{2\pi i} \oint_r \frac{f_n(z)dz}{(z - z_0)^{k+1}}.$$

Then this converges as $n \to \infty$ to something, which tells us that f satisfies the Cauchy Integral Formula. In particular, if $|z - z_0| < r$, then there's a power series

$$f(z) = \sum a_k (z - z_0)^k,$$

which shows that f is analytic. □

Using the formula $\sin z = \frac{e^{iz} - e^{-iz}}{2i}$, we already noted that all the zeroes of sin are on the real line. In fact, on the homework you showed the product formula

$$\sin z = z \prod \left(1 - \left(\frac{z}{\pi n}\right)^2\right).$$

Now we know that each of the partial products is analytic, and convergence is uniform, so the product is analytic. Then by analytic continuation we get this for all of $\mathbb{C}$.

Similarly, the Γ function comes to us defined as a limit–if we want to make something that looks like factorials, we can write

$$\Gamma(z) = \lim_{n\to\infty} \frac{n^z n!}{z(z_1)...(z + n)}.$$

However, there's a problem: this can't be uniformly convergent. There are a lot of poles. However, we can remove neighborhoods of each pole, and eventually get convergence for each point except the poles. As you might guess from the formula, while the Γ functions has poles, it doesn't have any zeroes. Then we might study $\frac{1}{\Gamma(z)}$ which will be analytic. On the other hand, we could simply accept that Γ is meromorphic, rather than analytic.

Remark 33.2. One way to describe this is that we really are looking at functions to the Riemann sphere (isomorphic as a topological space to the one point compactification of $\mathbb{C}$), rather than to $\mathbb{C}$. This makes everything a lot nicer and makes us happy.

We know

$$\begin{aligned}|\Gamma(z)| &\leq \int_0^\infty t^{\Re z} e^{-t} dt \\ &= \Gamma(\Re(z)),\end{aligned}$$

when $\Re(z) > 0$. In fact, $|\Gamma(x+iy)|$ is a decreasing function of $|y|$. [This doesn't seem exactly right–it should be repeating on multiples of 2π, so it can't be just a decreasing function. It's certainly true, however, that it never becomes greater than $|\Gamma(x)|$.] Recall we also had

$$\Gamma(x)\Gamma(1-x) = \frac{\pi}{\sin \pi x}$$

for $x \in \mathbb{C} - \mathbb{Z}$. From this formula it's particularly clear that Γ has no zeroes, since sin has no poles, nor π any zeroes.

In the real case, we had Sterling's approximation,

$$\Gamma(x) \sim \sqrt{2\pi} x^{x-1/2}/e^x.$$

This more or less works in the complex numbers. If we're on some simply connected subset of $\mathbb{C}$ not containing the origin, we can define the logarithm. Assuming our region contains 1 as well for simplicity, on this region we can set

$$\log \Gamma(z) = \int_1^z \frac{\Gamma'(s)}{\Gamma(x)} ds.$$

33.1. **Residue Theorem.** Because all the folds of $\log f$ differ by $2\pi i$, the derivative $(\log f)'$ is well defined on $\mathbb{C}$. Then $f \mapsto \frac{f'}{f}$ is a homomorphism from the set of meromorphic functions under multiplication to the set of meromorphic functions under addition. However, note that if we integrate around any boundary, this only takes values in multiples of $2\pi i$.

Today we'll be doing several applications of complex analysis to familiar problems. For example, we know

$$\int_0^\infty \frac{dx}{x^2+1} = \frac{\pi}{2}$$

because it's the derivative of arctan. However, we can also integrate things like

$$\int_0^\infty \cos(cx)\frac{dx}{x^2+1} = \frac{\pi}{2}e^{-|c|},$$

which don't have closed form antiderivatives. Such results come naturally from contour integration.

Theorem 34.1. (Legendre-Bürmann) *Let $w = g(z), g(0) = 0, g'(0) \neq 0$. We showed if g is differentiable then $z = g^{-1}(w)$ is also differentiable. When g is holomorphic, we can actually expect the same of g^{-1}, so $z = \sum c_n w^n$.*

Example 34.2. Let $w = ze^{-z} = z - z^2 + z^3/2 - ....$ Then we can solve this for $z = w + w^2 + 3w^3/2 + ... + \frac{n^{n+1}}{n!}w^n + ...$ After some manipulation with generating functions, this gives us some interesting combinatorial results about counting trees.

Definition 34.3. Let f be complex differentiable on a punctured disk $U - z_0$. Then the *residue* of $f(z)dz$ is

$$\operatorname{Res}(f)_{z_0} = \frac{1}{2\pi i}\oint_\gamma f(z)dz,$$

where γ is a path in U that surrounds z_0.

This is well defined by Green's theorem. Note if f can be extended to z_0 (i.e. is holomorphic on all of U) then again by Green's theorem the residue is 0. Then if our region contains multiple punctures/poles, we can divide it up into regions as used in the definition, and see that residues sum

$$\frac{1}{2\pi i}\oint_{\partial U} f(z)dz = \sum \operatorname{Res}(f)_z.$$

In summary, we have the following properties:

- If f is holomorphic at z_0 or has a removeable singularity, then the residue is 0
- If f is meromorphic, then

  $$f(z) = \sum_{n \geq -d} a_n(z-z_0)^n,$$

  and everything in our integral cancels except the $n = -1$ term. Then we see the residue we defined is just a_{-1}.

  This turns out to be true whether there is a pole of order d or an essential singularity, for more or less the same reason.

34.1. **Example 1.** Let's go back to our first example now,

$$\int_0^\infty \frac{dx}{x^2+1}.$$

By symmetry, we can rewrite this as

$$\frac{1}{2}\int_{-\infty}^\infty \frac{dx}{x^2+1},$$

which is a bit closer to something we can apply complex analysis to. We don't have a closed contour, so we'll formally add one as a hemisphere in the upper half plane. Then integrating over this contour, we must get just the sum of residues, but there's only one residue at $z = i$. It's clear by factoring that the residue is $\frac{1}{2i}$ by factoring $\frac{1}{x^2+1} = \frac{1}{x+i}\frac{1}{x-i}$ by evaluating $x+i$ at i. Then the integral over the boundary is equal to $2\pi i \cdot \frac{1}{2i} = \pi$. Then it suffices to check that the extra contour we added goes to 0 in order to evaluate the integral along the real line, which isn't hard to do. (Sample argument: our hemisphere has radius R, so the norm of $\frac{1}{x^2+1}$ gets very small, in a way faster than the integral should increase from the radius.) Then we just need to divide by two from the outside of our integral, and we do in fact get that the integral is $\frac{\pi}{2}$.

34.2. **Example 2.** We'll do one more example:

$$\int_0^\infty \cos cx \frac{dx}{x^2+1}.$$

We know $\cos cx = \frac{e^{icx}+e^{-icx}}{2}$. By symmetry, we can assume c is positive, and

$$\int_0^\infty \cos cx \frac{dx}{x^2+1} = \frac{1}{2}\int_{-\infty}^\infty \cos(cx)\frac{dx}{x^2+1} = \frac{1}{2}\operatorname{Re}\int_{-\infty}^\infty e^{icx}\frac{dx}{x^2+1}.$$

Now $|e^{icx}| = e^{-c\,\mathrm{Imx}} \leq 1$. Thus if we consider a hemisphere contour, the hemisphere part will go away leaving only the bottom. Now what's the residue? Again there's only one pole at $x = i$, and we now have a factorization

$$\frac{1}{x-i}\frac{1}{x+i}e^{icx}.$$

Evaluating for a_{-1} at $x = i$, we see that we get $\frac{1}{2i}e^{-c}$. This tells us that our integral is equal to $\frac{\pi}{2}e^{-|c|}$ for the same reason as before.

34.3. **Change of Variable Stuff.** Let g be a map from some neighborhood to some other neigborhood, $g'(w_0) \neq 0$, and $f : U \to \mathbb{C}$ has a pole at $z_0 = g(w_0)$. Then one thing we can consider is $f(g(w))dz(w) = f(g(w))g'(w)dw$. (This makes sense by using the inverse funciton theorem on some neighborhood???) We see that the residue is therefore invariant under change of variables, just by checking.

35. Lecture 35 – whatevs who cares about dates because I'm never on them

I'm andy and not wyatt and I'm so lonely life is miserable :(

As we've noted/used frequently, one very nice thing about holomorphic functions is that they're automatically analytic–we can at least locally use power series. More precisely, given f there is a power series expansion in $|z - z_0| \leq R$ if f is differentiable on some open enigbhorhod of $|z - z_0| \leq R$. A pole is the same thing as the normal power series, except we allow sums starting at negative numbers.

> **Example 35.1.** A pole, incidentally, is what I usually dance on for $$$.

An essential singularity occurs if we sum from $-\infty$: that is, our power series expansion is

$$\sum_{n=-\infty}^{\infty} a_n (z - z_0)^n.$$

Again we have a radius of convergence for this–if we imagine inverting the function and looking for things far away from ∞, we get a smaller circle inside our larger circle, outside which our sum is defined. Thus our power series converges on some annulus.

To see that this really works, translate so that $z_0 = 0$. Then we know $f(z)$ is minus the residue of $\frac{f(w)}{z-w}$ at $w = z$. We can't quite do this over just the outer circle, because the corresponding space isn't simply connected. Instead, applying Green's theorem, we see the integral must also incorporate the inner circle, evaluated with reversed orientation. Dividing the circle on some branch cut, we see this is the same as integrating over some rectangle. Then

$$|w| = R\frac{1}{w - z} = \frac{1}{w(1 - z/w)} = \frac{1}{w} + \frac{z}{w^2} + \ldots$$

and so

$$\frac{1}{2\pi i} \oint_R = \sum a_n z^n,$$

where a_n is given by our usual formula. We can also do the expansion where we invert z instead of w, which gives us an expansion in terms of $\frac{1}{z}$. We get an extra minus sign, which cancels with the orientation minus sign to get

$$-\frac{1}{2\pi i} \oint_r = \sum_{n=-\infty}^{-1} a - nz^n,$$

again with the usual formula for coefficients from the Cauchy Integral Formula. As an example, if we do this around a circle of radius 1, so $r < 1 < R$, we have

$$\begin{aligned} a_n &= \frac{1}{2\pi i} \int_0^{2\pi} f(e^{i\theta}) \frac{d(e^{i\theta})}{e^{(n+1)i\theta}} \\ &= \frac{1}{2\pi} \int_0^{2\pi} e^{-ni\theta} f(e^{i\theta}) d\theta, \end{aligned}$$

which really is just the Fourier series.

Suppose g is defined on some neighborhood of the unit disk, and is a map to $\mathbb{C}^*$. Then since disk is simply connected, we can define log on it in a well defined way, so

$$\log|g(0)| = \frac{1}{2\pi}\int_0^{2\pi} \log|g(e^{i\theta})|d\theta.$$

Now if $f(0) \neq 0$, f is defined on the same region as g, and f has zeroes at $z_1, ..., z_n$ with some multiplicity, so

$$f(z) = g(z)\prod(z - z_k)^{e_k}$$

for some g as above, we know how to compute $\log|g(0)|$. We'd like ot go from this to computing $\log|f(0)|$. In particular, we get

$$\log|f(0)| = \frac{1}{2\pi}\log|g(e^{i\theta})|d\theta + e_k\sum\log z_k = \frac{1}{2\pi}\int_0^{2\pi}\log f(e^{i\theta})d\theta.$$

This is known as *Jensen's Equation*.

Theorem 35.2. (Jensen's Inequality) *if $|f| \leq B$ on $|z| \leq 1$, then*

$$|f(0)| \leq B\prod|z_k|.$$

In particular, the more zeroes f has, and the closer they are to 0, the smaller $f(0)$ must be.

me andy gordon knows a thing or two about stuff being particularly small, so you can take this theorem on good authority from me.

That's a sketch because I'm running out of time and my time's up, eyes up, eyes up. I catch a glimpse of the other side–Lawrence leads a soldiers chorus on the other side my son is on the other side. Washington is watching from the other side–teach me how to say goodbye, eyes up, rise up, Eliza–my love take your time. I'll see you on the other side. (Raise a glass to freedom) [He aims a pistol at the sky] Wait!

I shoot him right between the ribs. I walk towards him, but I am ushered away. I row back across the Hudson–I get a drink. I hear wailing in the streets, somebody tells me 'you'd better hide.' They say Angelica and Eliza were both by his side when he died. Death doesn't discriminate between the sinners and the saints it takes and it takes and it takes and history obliterates every picture it paints it paints me in all my mistakes. Alexander may have been the first one to die but I'm the one who paid for it. I survived but I paid for it.