

# 1 January 8

“Sometimes MVT stands for ‘most-valuable theorem’.”

## 1.1 Logistics

- Homework: Due on Fridays at the start of class (Canvas), posted on Thursday or Friday. List name of people you worked with at the top
  - Homework 0 is due this Friday, not due marks, but he is using to test automated test system to hand back
  - Some help: watch Monty Python Holy Grail
- Office hours: Zahl Wed. 1-2; TA TBD; TA2 TBD
- Weighting:
  - HW 30%
  - MT 30% Feb 14
  - Final 40%

## 320 Addendum

Typically, 321 picks up at integration after finishing with differentiation in 320. But we will pick up some missed material at the end of 320.

Recall

*Definition 1.*  $f: [a, b] \rightarrow \mathbb{R}$ ,  $c \in [a, b]$ , we say that  $f$  is *differentiable at  $c$*  if  $\lim_{x \rightarrow c} \frac{f(x) - f(c)}{x - c}$  exists (as a real number). We denote this by  $f'(c)$ .

This is nice, but could even do in high school. But we can go up many levels of abstraction with our limit.

- $c$  is a limit point in  $[a, b]$  (for every  $\varepsilon$ , a point not  $c$  exists inside the open ball)
- $g(x) = \frac{f(x) - f(c)}{x - c}$  is a function with domain  $[a, b] \setminus \{c\}$  (thankfully,  $c$  is still a limit point of this).
- If  $c \in (a, b)$ , the high school definition of the limit works. If  $c = a, b$ , then one-sided limit.

*Definition 2.* If  $f: [a, b] \rightarrow \mathbb{R}$  is differentiable at every point  $c \in [a, b]$ , then we say that  $f$  is *differentiable on  $[a, b]$*  and this gives us a new function  $f': [a, b] \rightarrow \mathbb{R}$ .

*Definition 3.* If  $f'$  is differentiable at  $c \in [a, b]$ , write  $f''(c) = (f')'(c)$ . Alternate notations:

$$\begin{array}{ccccccc} f(c), & f'(c), & f''(c), & f'''(c) \\ f^{(0)}(c), & f^{(1)}(c), & f^{(2)}(c), & f^{(3)}(c), & \dots & f^{(k)}(c) = \frac{d^k}{dx^k} f(x)|_{x=c} \end{array}$$

Some questions to consider:

- Why have codomain  $\mathbb{R}$ ? Why not  $\mathbb{C}$ ? Field  $F$ ? General set / metric space?
- Why make domain a closed interval?? More general subset of  $\mathbb{R}$ ?  $\mathbb{C}$ ? Set / metric space?

The derivative is one of the most important concepts, and so makes sense people have thought about making it more general. Sometimes it works, sometimes it doesn't.

We've used the field structure of  $\mathbb{R}$  in an important way (not necessarily the order structure)... more than just a metric space. Seems ambitious to have a topological definition of a derivative, because a derivative is a quantitative rate of change, and we don't get that in a topology. You can probably find people who have constructed a topological derivative, but will have needed to give up desired properties.

## 1.2 Taylor's Theorem

Recall the special case to MVT that is used to prove MVT:

**Theorem 1** (Roll's Theorem). *Let  $f: [a, b] \rightarrow \mathbb{R}$  be differentiable, with  $f(a) = f(b)$ . Then  $\exists c \in (a, b)$  such that  $f'(c) = 0$ .*

If you don't remember how to prove this, good exercise to go through that uses a lot of material from Math 320.

We will use this to prove Taylor's theorem, which seems really strong and handles most of our everyday functions, but easy to step on landmines (slightly rewording a true statement gives a really wrong one). We are going to give sufficient hypotheses, could technically weaken, but not as clear.

**Theorem 2** (Taylor's Theorem (5.15 in Rudin)). *Let  $f: [a, b] \rightarrow \mathbb{R}$ , let  $n \geq 0$  be an integer. Suppose that  $f$  is  $(n+1)$  times differentiable on  $[a, b]$ . Let  $x_0$  and  $x$  be points in  $[a, b]$  with  $x_0 \neq x$ . Then there exists a point  $c$  strictly between  $x_0$  and  $x$  such that*

$$f(x) = \sum_{k=0}^n \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k + \frac{f^{(n+1)}(c)}{(n+1)!} (x - x_0)^{n+1} \quad (1)$$

We hope the error term is small so we can control it. We won't prove this today because of the time. But we will say  $P_n(x) = \sum_{k=0}^n \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k$  is the "degree  $n$  Taylor expansion of  $f$  around  $x_0$ ". When you are choosing notation, there are competing goals: don't want it to be flowery so that it becomes more complicated  $P_n^{f, x_0}(x)$ , but will have to remember what we are hiding. And also because what is of interest might be when we fix  $f, x_0$ , and sending  $n$  to infinity.

Is this equation helpful? Will depend on how small the "error term" gets (could dominate!). But polynomial if  $x$  is close to  $x_0$  is going to zero geometrically, and factorial is going to zero faster than geometric, so actually for most functions, we get to say something nice.

Question: let  $f: \mathbb{R} \rightarrow \mathbb{R}$ , infinitely differentiable. Suppose  $f^{(k)}(0) = 0$  for all  $k$ . Is it true that  $f$  must be the zero function? Because then the function is just the error term, and so Taylor's theorem fails in the worst way.

## 2 January 10

### 2.1 Proof of Taylor's Theorem

How do we go about proving statement's like this? There are many different proofs of this result. One thing to do when confronted by these statements is to consider special cases.

Warm-up ( $n = 0$ ): (1) becomes  $f(x) = f(x_0) + f'(c)(x - x_0)$ . This is just MVT, which we have already proved. So in some sense, more general than MVT. We can write this in a clever way to make the proof use Rolle's instead.

*Proof.* Define  $A \in \mathbb{R}$  by

$$f(x) - P_n(x) = \frac{A}{(n+1)!} (x - x_0)^{n+1}$$

This  $A$  exists, since we can just bring the other factors on the right to the left and get  $A$  explicitly. Our goal is to show there exists  $c$  between  $x_0$  and  $x$  such that  $f^{(n+1)}(c) = A$ .

Define  $g(t) = f(t) - P_n(t) - \frac{A}{(n+1)!} (t - x_0)^{n+1}$ . Note to use Rolle's theorem, we have the freedom to shrink our endpoints so that they are equal to each other (you might have seen this trick to prove MVT from Rolle's). Common to construct a new function that meets the hypotheses of a theorem we want, and use that to inform us about our original function. We claim that we can shrink it to  $x$  and  $x_0$ . Observe  $g(x_0) = f(x_0) - P_n(x_0) - 0 = f(x_0) - f(x_0) = 0$  and  $g(x) = 0$  by the definition of  $A$ .

For  $j = 0, \dots, n$ , then  $g^{(j)}(x_0) = f^{(j)}(x_0) - P_n^{(j)}(x_0) - \underbrace{\frac{d^j}{dt^j} \frac{A}{(n+1)!} (t - x_0)^{n+1}}_0 \Big|_{t=x_0}$ . But also  $P_n^{(j)}(x_0) = f^{(j)}(x_0)$ ,

since if  $k < j$ , then the derivatives kill the term, and if  $j > k$ , then we have a  $(x_0 - x_0) = 0$  as a factor. Hence,  $g^{(j)}(x_0) = f^{(j)}(x_0) - f^{(j)}(x_0) = 0$ . We also have  $g^{(n+1)}(t) = f^{(n+1)}(t) - 0 - A$ . So we could reword our goal to be to find  $c$  such that  $g^{(n+1)}(c) = 0$ .

Now it is time to start using Rolle's. We have  $g(x_0) = 0$  and  $g(x) = 0$ . Then by Rolle's, there exists  $c_1$  between  $x$  and  $x_0$  such that  $g'(c_1) = 0$ . But we can apply Rolle's theorem again, since  $g'(x_0) = 0$  and  $g'(c_1) = 0$ . So there exists  $c_2$  between  $c_1$  and  $x_0$  such that  $g''(c_2) = 0$ . We can repeat this  $n$  times to get  $g^{(n)}(x_0) = g^{(n)}(c_n) = 0$ . Hence, there exists  $c_{n+1}$  between  $x_0$  and  $c_n$  such that  $g^{(n+1)}(c_{n+1}) = 0$ . Let  $c = c_{n+1}$ , and so we have achieved our goal.  $\square$

Now this seems a bit magic. You had to cook up a magic function  $g$  and it solves it for you. But Zähl would do this by looking at  $n = 1$ , maybe  $n = 2$  and  $n = 3$ , find something that makes the derivative vanish. How do we remember how to prove this? You really want to chunk things into a small number of ideas: you want to use Rolle's theorem (many times), want to cook up an auxiliary function  $g$  that satisfies Rolle's theorem hypothesis each time, and given 20 minutes, you could do it.

Given random people and chess grandmasters and random chessboards, did equally bad in recreating the board by memory. But given a chessboard that was halfway into a game, chess grandmasters did significantly better. The takeaway: the grandmasters weren't remembering the exact location of every piece, they were remembering by chunking and creating a narrative.

Examples ( $x_0 = 0$ ):

1.  $f$  a polynomial of degree  $D$ .  $P_n(t)$  is the first terms of  $f$ , up to degree  $n$ . So the Taylor expansion eventually becomes the polynomial.
2.  $f(t) = e^t$ .  $P_n(t) = \frac{1}{0!} + \frac{t}{1!} + \frac{t^2}{2!} + \cdots + \frac{t^n}{n!}$ .
3.  $f(t) = \sin t$ .  $P_n(t) = 0 + t + 0 - \frac{t^3}{3!} - 0 + \frac{t^5}{5!}$ .

We have talked about convergence in metric spaces.  $P_n$  is a sequence, so how do we talk about  $P_n$  converging to  $f$ ? When does it? What metric space are we in? We could let our metric space be infinitely differentiable real functions, with the metric  $d(f, g) = \sup_{x \in D} |f(x) - g(x)|$ . The first example converges, since it is eventually 0. What about  $e^t$ ? Well, if we take  $n$  before  $x$ , not true, exponential vs polynomial. But if fix a closed interval, then we do get an  $n$ . This is the difference between pointwise and uniform convergence. Our metric was normally used for bounded functions, so this is necessary. We will pick up this discussion next time.

### 3 January 12

We ended last class with some discussion of convergence of functions. This is something he wants to defer until later in the term, but we get some of this on the homework.

Recall he asked two lectures ago if  $f: \mathbb{R} \rightarrow \mathbb{R}$  (or  $[-1, 1] \rightarrow \mathbb{R}$ ), and  $f(0) = 0, f'(0) = 0, \dots, f^{(k)}(0) = 0$  for all  $k$ , must it be true that  $f(t) = 0$  for all  $t$ . Taylor's theorem makes us think this is true: otherwise,  $f(x)$  is wholly its error term always. But consider

$$f(x) = \begin{cases} 0, & \text{if } x \leq 0 \\ e^{-1/x} & \text{if } x > 0 \end{cases}$$

We can see that this is infinitely differentiable on all of  $\mathbb{R}$  (consider the domains separately). For  $x \leq 0$ ,  $f^{(k)}(x) = 0$  and for  $x > 0$ ,  $f^{(k)}(x) = Q(x)e^{-1/x}$  where  $Q(x)$  is a rational function. And exponential will shrink always faster than  $Q(x)$ , so goes to 0 at origin as well. Proving this is something he's given in homework, but it eats a whole week, and so probably not this year.

It is valuable in analysis (or all math) to have a large bank of interesting examples at your finger tips that exemplify the weird behaviour of how functions behave (or things behave). Some theorems say this is what happens with these four examples, and then everything else can be derived from these interesting examples.

### 3.1 The Riemann and Riemann-Stieltjes Integral

#### 3.1.1 The Riemann Integral

*Definition 4* (Rudin 6.1). A *partition* of  $[a, b]$  is a finite set

$$P = \{x_0, x_1, x_2, \dots, x_n\}$$

with  $a = x_0 < x_1 < x_2 < \cdots < x_n = b$ .

For  $i = 1, \dots, n$ , let  $\Delta x_i = x_i - x_{i-1}$ . For  $f: [a, b] \rightarrow \mathbb{R}$  bounded, define

$$M_i = \sup\{f(x) : x \in [x_{i-1}, x_i]\}$$

$$m_i = \inf\{f(x) : x \in [x_{i-1}, x_i]\}$$

And these will always exist in  $\mathbb{R}$ , since the interval is nonempty and  $f(x)$  is bounded. Define the upper and lower Riemann sums of the partition  $P$  and function  $f(x)$ :

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

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

Intuition: drawing boxes for each  $\Delta x_i$  whose height is  $M_i$  or  $m_i$ . As we take smaller widths of the boxes, we would like to see these two values converge.

We define the upper Riemann integral

$$\overline{\int_a^b} f dx = \inf_P U(P, f)$$

and the lower Riemann integral

$$\underline{\int_a^b} f dx = \sup_P L(P, f)$$

where the supremum and infimum is taken over all partitions of  $[a, b]$ .

*Definition 5.* We say  $f: [a, b] \rightarrow \mathbb{R}$  is *Riemann integrable* if

$$\overline{\int_a^b} f dx = \underline{\int_a^b} f dx$$

(both should always exist), in which case we denote this number by

$$\int_a^b f dx$$

and we say  $f \in \mathcal{R}[a, b]$  (the set of Riemann integrable functions on  $[a, b]$ ).

Example:  $[a, b] = [0, 1]$ ,  $f(x) = x$ . If  $P = \{x_0, \dots, x_n\}$  is a partition,  $M_i = x_i$  and  $m_i = x_{i-1}$ . For the sake of concreteness, consider the evenly-spaced partition  $\{0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n}{n}\}$ . Then

$$U(P, f) = \sum_{i=1}^n \frac{i}{n} \frac{1}{n} = \frac{1}{n^2} \sum_{i=1}^n i = \frac{1}{n^2} \frac{1}{2} n(n+1) = \frac{1}{2} + \frac{1}{2n}$$

In particular,  $\overline{\int_a^b} f dx \leq \inf\{\frac{1}{2} + \frac{1}{2n} : n \in \mathbb{N}\} = \frac{1}{2}$ . Also

$$L(P, f) = \sum_{i=1}^n \frac{i-1}{n} \frac{1}{n} = \frac{1}{n^2} \frac{1}{2} n(n-1) = \frac{1}{2} - \frac{1}{2n}$$

In particular,  $\underline{\int_a^b} f dx \geq \sup\{\frac{1}{2} - \frac{1}{2n} : n \in \mathbb{N}\} = \frac{1}{2}$ . Can we now conclude that  $f \in \mathcal{R}[0, 1]$  and  $\int_a^b f dx = \frac{1}{2}$ ? Well, we need that  $\underline{\int_a^b} f dx \leq \overline{\int_a^b} f dx$  always, which we don't have yet. [Is this not a result of the  $\liminf \leq \limsup$  inequality?] We will prove this actually for a more general class of integrals.

### 3.1.2 The Riemann-Stieltjes Integral

*Definition 6* (Rudin 6.2). Let  $\alpha: [a, b] \rightarrow \mathbb{R}$  be (weakly) monotone increasing, let  $P = \{x_0, \dots, x_n\}$  be a partition of  $[a, b]$ .

For  $i = 1, \dots, n$ , let  $\Delta\alpha_i = \alpha(x_i) - \alpha(x_{i-1})$  (if  $\alpha(x) = x$ , then  $\Delta\alpha_i = \Delta x_i$ ).

For  $f: [a, b] \rightarrow \mathbb{R}$  bounded define

$$U(P, f, \alpha) = \sum_{i=1}^n M_i \Delta\alpha_i$$

$$L(P, f, \alpha) = \sum_{i=1}^n m_i \Delta\alpha_i$$

(where  $M_i, m_i$  are the same as before). As before, we define

$$\overline{\int_a^b} f d\alpha = \inf_P U(P, f, \alpha)$$

$$\underline{\int_a^b} f d\alpha = \sup_P L(P, f, \alpha)$$

If

$$\overline{\int_a^b} f d\alpha = \underline{\int_a^b} f d\alpha$$

then we denote the common value by  $\int_a^b f d\alpha$  and we say  $f \in \mathcal{R}_\alpha[a, b]$ . When  $\alpha(x) = x$ , we get the Riemann integral from before. But consider  $\alpha(x) = \begin{cases} 0, & x \leq 0 \\ 1, & x > 0 \end{cases}$ . What does  $\int_0^1 f d\alpha$  look like ( $f$  continuous). This actually evaluates to  $f(0)$  (Dirac delta from physics).

## 4 January 15

*Definition 7* (Rudin 6.3). Let  $P$  and  $P^*$  be partitions of  $[a, b]$ . We say  $P^*$  is a *refinement* of  $P$  if  $P \subset P^*$ . If  $P_1$  and  $P_2$  are partitions of  $[a, b]$ , the *common refinement* is the partition  $P_1 \cup P_2$ .

**Theorem 3** (Rudin 6.4). *Let  $P^*$  be a refinement of  $P$ . Then  $L(P, f, \alpha) \leq L(P^*, f, \alpha) \leq U(P^*, f, \alpha) \leq U(P, f, \alpha)$ .*

*Proof.* The middle inequality, we have already seen.

Since partitions are finite, it suffices to prove the inequality when  $P^*$  has one additional point (i.e.  $x_i < x^* < x_{i+1}$  and  $P^* = P \cup \{x^*\}$ ).

Lets compare  $L(P, f, \alpha)$  vs  $L(P^*, f, \alpha)$ . Recall  $m_j = \sup\{f(x) : x \in [x_{j-1}, x_j]\}$ . See

$$L(P, f, \alpha) = \sum_{j=1}^n m_j \Delta\alpha_j$$

$$L(P^*, f, \alpha) = \sum_{j=1}^i m_j \Delta\alpha_j + (\inf\{f(x) : x \in [x_i, x^*]\})(\alpha(x^*) - \alpha(x_i))$$

$$+ (\inf\{f(x) : x \in [x^*, x_{i+1}]\})(\alpha(x_{i+1}) - \alpha(x^*))$$

$$+ \sum_{j=i+2}^n m_j \Delta\alpha_j$$

Then

$$\begin{aligned}
 L(P^*, f, \alpha) - L(P, f, \alpha) &= \left( \inf_{x \in [x_i, x^*]} f(x) \right) (\alpha(x^*) - \alpha(x_i)) + \left( \inf_{x \in [x^*, x_{i+1}]} f(x) \right) (\alpha(x_{i+1}) - \alpha(x^*)) - m_{i+1} \Delta \alpha_{i+1} \\
 &\geq \left( \inf_{x \in [x_i, x_{i+1}]} f(x) \right) (\alpha(x^*) - \alpha(x_i)) + \left( \inf_{x \in [x_i, x_{i+1}]} f(x) \right) (\alpha(x_{i+1}) - \alpha(x^*)) - m_{i+1} \Delta \alpha_{i+1} \\
 &= m_{i+1} (\alpha(x^*) - \alpha(x_i) + \alpha(x_{i+1}) - \alpha(x^*)) - m_{i+1} \Delta \alpha_{i+1} \\
 &= m_{i+1} \Delta \alpha_{i+1} - m_{i+1} \Delta \alpha_{i+1} = 0
 \end{aligned}$$

where we are using the fact that adding points to a set can only decrease its infimum. Lastly, the final inequality is proven similarly.  $\square$

Recall we wanted to prove something before we moved to the Riemann-Stieltjes integral.

**Theorem 4** (Rudin 6.5). *Let  $f: [a, b] \rightarrow \mathbb{R}$  bounded,  $\alpha: [a, b] \rightarrow \mathbb{R}$  monotone increasing. Then*

$$\int_a^b f d\alpha \leq \overline{\int_a^b f d\alpha}$$

*Proof.* Let  $P_1$  and  $P_2$  be partitions of  $[a, b]$ , let  $P^* = P_1 \cup P_2$ . By Theorem 6.4,  $L(P_1, f, \alpha) \leq U(P_2, f, \alpha)$ , hence

$$\int_a^b f d\alpha = \sup_{P_1} L(P_1, f, \alpha) \leq U(P_2, f, \alpha)$$

See that this is true for every  $P_2$ .

$$\int_a^b f d\alpha \leq \inf_{P_2} U(P_2, f, \alpha) = \overline{\int_a^b f d\alpha}$$

$\square$

This was the missing piece to show  $\int x dx = \frac{1}{2}$ .

Now we are going to look at some facts about the Riemann-Stieltjes integral, and they will slowly get less intuitive.

**Theorem 5** (Rudin 6.6). *Let  $f: [a, b] \rightarrow \mathbb{R}$  bounded and  $\alpha: [a, b] \rightarrow \mathbb{R}$  monotonically increasing. Then*

$$f \in \mathcal{R}_\alpha[a, b] \iff \forall \varepsilon, \exists P \text{ such that } U(P, f, \alpha) - L(P, f, \alpha) < \varepsilon$$

*Proof.* Forward direction: by hypothesis

$$\sup_P L(P, f, \alpha) = \int_a^b f d\alpha = \inf_P U(P, f, \alpha)$$

Let  $\varepsilon > 0$ . Then  $\exists$  a partition  $P_1$  such that

$$L(P_1, f, \alpha) > \int_a^b f d\alpha - \varepsilon/2$$

$\exists P_2$  such that  $U(P_2, f, \alpha) < \int_a^b f d\alpha + \varepsilon/2$ . Let  $P = P_1 \cup P_2$ . By theorem 6.4

$$L(P_1, f, \alpha) \leq L(P, f, \alpha) \leq U(P, f, \alpha) \leq U(P_2, f, \alpha)$$

Hence  $U(P, f, \alpha) - L(P, f, \alpha) < \varepsilon$ .

Backwards direction follows from definition.  $\square$

## 5 January 17

We are going to continue understanding the Riemann-Stieltjes integral. We showed when it existed (upper and lower Riemann-Stieltjes integral agrees), but we don't know a lot about functions that are Riemann-Stieltjes integrable. We show this now for a large class of functions.

**Theorem 6** (Rudin 6.8). *Let  $\alpha: [a, b] \rightarrow \mathbb{R}$  be monotonically increasing. Let  $f: [a, b] \rightarrow \mathbb{R}$  be continuous. Then  $f \in \mathcal{R}_\alpha[a, b]$ , i.e.  $C([a, b]) \subset \mathcal{R}_\alpha[a, b]$ .*

*Proof.* Let  $f$  be continuous, and  $[a, b]$  compact. Hence,  $f$  is uniformly continuous. Hence, for all  $\varepsilon_1 > 0$ ,  $\exists \delta > 0$  such that  $|f(x) - f(y)| < \varepsilon_1$  for all  $x, y$  with  $|x - y| < \delta$ . Thus, if  $P$  is a partition with  $\Delta x_i < \delta$  for all  $i$ , then  $M_i - m_i < \varepsilon_1$  for all  $i$  (hmm might be important here for the strict inequality that  $f$  attain max/min because compact set... closed subset of compact). Hence  $U(P, f, \alpha) - L(P, f, \alpha) = \sum_i M_i \Delta \alpha_i - \sum_i m_i \Delta \alpha_i \leq \sum_{i=1}^n \varepsilon_1 \Delta \alpha_i = \varepsilon_1 (\alpha(b) - \alpha(a))$ . Given  $\varepsilon > 0$ , select  $\varepsilon_1$  sufficiently small so that  $\varepsilon_1 (\alpha(b) - \alpha(a)) < \varepsilon$ . Choose  $P$  as above for the corresponding  $\varepsilon_1$ . We have shown: for  $\varepsilon > 0$ ,  $\exists$  a partition  $P$  such that  $U(P, f, \alpha) - L(P, f, \alpha) < \varepsilon$ . So by Theorem 6.6,  $f \in \mathcal{R}_\alpha[a, b]$ .  $\square$

Question: Can we describe/characterize  $\mathcal{R}_\alpha[a, b]$  or  $\mathcal{R}[a, b]$ ? We will pick this up later.

For now, we will expand the functions that are Riemann integrable.

**Theorem 7** (Rudin 6.9). *Let  $f: [a, b] \rightarrow \mathbb{R}$  be monotone (increasing or decreasing) and  $\alpha: [a, b] \rightarrow \mathbb{R}$  monotone (weakly) increasing and continuous. Then  $f \in \mathcal{R}_\alpha[a, b]$ .*

*Remark 1.* This theorem neither implies or is implied by the previous theorem.

*Proof.* Let  $n \in \mathbb{N}$ . By the intermediate value theorem,  $\exists$  a partition  $P$  such that  $\Delta \alpha_i = \frac{\alpha(b) - \alpha(a)}{n}$  for all  $i = 1, \dots, n$ .

$$\begin{aligned} U(P, f, \alpha) - L(P, f, \alpha) &= \sum_{i=1}^n (M_i - m_i) \Delta \alpha_i \\ &= \frac{\alpha(b) - \alpha(a)}{n} \sum_{i=1}^n (M_i - m_i) \end{aligned}$$

Suppose, WLOG, that  $f$  is monotone increasing, then  $M_i = f(x_i)$ ,  $m_i = f(x_{i-1})$ . So our value from before becomes

$$\begin{aligned} U(P, f, \alpha) - L(P, f, \alpha) &= \frac{\alpha(b) - \alpha(a)}{n} \sum_{i=1}^n (f(x_i) - f(x_{i-1})) \\ &= \frac{\alpha(b) - \alpha(a)}{n} (f(x_1) - f(x_0) + f(x_2) - f(x_1) + \dots + f(x_n) - f(x_{n-1})) \\ &= \frac{\alpha(b) - \alpha(a)}{n} (f(x_n) - f(x_0)) \\ &= \frac{\alpha(b) - \alpha(a)}{n} (f(b) - f(a)) \\ &= \frac{1}{n} \underbrace{(\alpha(b) - \alpha(a)) (f(b) - f(a))}_{\in \mathbb{R}} \end{aligned}$$

Given  $\varepsilon > 0$ , select  $n \in \mathbb{N}$  such that

$$|(\alpha(b) - \alpha(a)) (f(b) - f(a))| < \varepsilon$$

For such a partition  $P$ ,  $U(P, f, \alpha) - L(P, f, \alpha) < \varepsilon$ . By Theorem 6.6,  $f \in \mathcal{R}_\alpha[a, b]$ . If  $f$  is monotonically decreasing, an identical proof works.  $\square$

**Theorem 8** (Rudin 6.10). *Let  $f: [a, b] \rightarrow \mathbb{R}$  be bounded and continuous at all but finitely many points. Let  $\alpha: [a, b] \rightarrow \mathbb{R}$  be monotone increasing and must be continuous at every point where  $f$  is not continuous. Then  $f \in \mathcal{R}_\alpha[a, b]$ .*

Probably won't prove today, but some examples. ff

And then tangent into monotonically increasing functions with countably many discontinuities. Can show that the Devil's staircase  $\alpha: [0, 1] \rightarrow [0, 1]$  is monotonically increasing and discontinuous. Remember he mentioned that it is useful to have examples in your head. For Riemann-Stieltjes integral,  $\alpha(x) = x$  is a good one (Riemann integral), but also this Cantor-Lebesgue function.

## January 22

### 5.1 Properties of the Riemann-Stieltjes Integral

Assume that  $\alpha: [a, b] \rightarrow \mathbb{R}$  is monotonically increasing, and  $f, f_1, f_2: [a, b] \rightarrow \mathbb{R}$  are  $f, f_1, f_2 \in \mathcal{R}_\alpha[a, b]$ . Then we get some more functions that are integrable.

- (a). Linearity:  $f_1 + f_2 \in \mathcal{R}_\alpha[a, b]$  and  $\int_a^b (f_1 + f_2) d\alpha = \int_a^b f_1 d\alpha + \int_a^b f_2 d\alpha$ ;  $c \in \mathbb{R}$  then  $cf \in \mathcal{R}_\alpha[a, b]$  and  $\int_a^b cf d\alpha = c \int_a^b f d\alpha$ .
- (b). Non-negativity: If  $f(x) \geq 0$  for all  $x \in [a, b]$  then  $\int_a^b f d\alpha \geq 0$ . This then implies if  $f_1(x) \leq f_2(x)$  for all  $x \in [a, b]$  then  $\int_a^b f_1 d\alpha \leq \int_a^b f_2 d\alpha$ .
- (c). For  $c \in (a, b)$ , then  $f \in \mathcal{R}_\alpha[a, c]$  and  $f \in \mathcal{R}_\alpha[c, b]$  and  $\int_a^c f d\alpha + \int_c^b f d\alpha = \int_a^b f d\alpha$ .
- (d). Boundedness: If  $|f| \leq M$  then  $\left| \int_a^b f d\alpha \right| \leq M(\alpha(b) - \alpha(a))$ .
- (e). Let  $\alpha_1, \alpha_2: [a, b] \rightarrow \mathbb{R}$  and they're monotone increasing, and  $f: [a, b] \rightarrow \mathbb{R}$  where  $f \in \mathcal{R}_{\alpha_1}[a, b]$  and  $f \in \mathcal{R}_{\alpha_2}[a, b]$ . Then  $f \in \mathcal{R}_{\alpha_1 + \alpha_2}[a, b]$  and  $\int_a^b f d(\alpha_1 + \alpha_2) = \int_a^b f d\alpha_1 + \int_a^b f d\alpha_2$ . If  $c \in \mathbb{R}$ , then  $f \in \mathcal{R}_{c\alpha_1}[a, b]$  and  $\int_a^b f d(c\alpha_1) = c \int_a^b f d\alpha_1$ .

Proof: See Rudin; this is theorem 6.12.

Now we will have an informal discussion about what this all tells us. Why bother with a more complicated version of an integral, and why this specific setup? Recall from 320  $C([a, b])$ , the space of continuous functions  $f: [a, b] \rightarrow \mathbb{R}$ . We define  $\|f\|_{C([a, b])} = \sup_{x \in [a, b]} |f(x)|$  and so  $d(f, g) = \|f - g\|_{C([a, b])}$  (also called the  $C_0$  norm). This space is a vector space, infinite dimensional (field is reals for now, but can extend to complex). Thus,  $(C([a, b]), \|\cdot\|_{C([a, b])})$  is a normed vector space. In linear algebra, we are often interested in linear functions from vector spaces, or here, linear transforms between normed vector spaces. Property (a) from theorem 6.12 says: If  $\alpha: [a, b] \rightarrow \mathbb{R}$  is monotone increasing, then the function  $T(f) = \int_a^b f d\alpha$  is a linear function from the vector space  $C([a, b])$  to  $\mathbb{R}$  (for now, we will restrict our domain), i.e.  $T(f + g) = T(f) + T(g)$ ,  $T(cf) = cT(f)$ . Property (d) says that  $T$  is bounded, i.e.  $|Tf| \leq A\|f\|_{C([a, b])}$ , or more specifically,  $|Tf| \leq (\alpha(b) - \alpha(a))\|f\|_{C([a, b])}$  (since  $|f| \leq M$  when  $f \in C([a, b])$  means  $\|f\|_{C([a, b])} \leq M$ ). Notation: people sometimes write  $Tf$  instead of  $T(f)$  (in linear algebra, we write  $Mv$ ). Property (b) says that  $T$  is non-negative, i.e. if  $f \in C([a, b])$  with  $f(x) \geq 0$  for all  $x \in [a, b]$ , then  $Tf \geq 0$ .

Why have we changed our wording from theorem 6.12 to this? In functional analysis (Math 421) and more generally in physics, we want to study linear functions whose domain is  $C([a, b])$  (or more general) and whose codomain is often  $\mathbb{R}$  or  $\mathbb{C}$ . Functions of this type are called “(linear) operators” or “(linear) functionals”. Can't say too much right now about why this is an interesting area of study, but a starting point in modern quantum mechanics is, instead of thinking of particles as points, we think of it as a function. In classic mechanics, a particle might be described with 6 numbers  $(x, y, z, v_x, v_y, v_z)$ , and if you were to ask what the position of this particle, we just take the first three (and likewise with the velocity/momentum). Instead, we now consider transformations on the function.

**Theorem 9** (Riesz Representation Theorem V1). *Let  $T: C([a, b]) \rightarrow \mathbb{R}$  be linear, bounded, and non-negative. Then, there exists  $\alpha: [a, b] \rightarrow \mathbb{R}$  monotonically increasing such that  $Tf = \int_a^b f d\alpha$ .*

The whole point of this theorem is that you can go in the other direction.  $f$  and  $\alpha$  uniquely define a linear bounded non-negative operator  $T: [a, b] \rightarrow \mathbb{R}$ , and  $T$  uniquely determines  $f, \alpha$ .

This non-negative condition seems a little weird... what happens when our codomain is  $\mathbb{C}$ ?

**Theorem 10** (Riesz Representation Theorem V2). *Let  $T: C([a, b]) \rightarrow \mathbb{R}$  be linear and bounded. “Let  $T$  be a real-valued linear functional on  $C([a, b])$ .” Then there exists  $\alpha, \beta: [a, b] \rightarrow \mathbb{R}$  monotone increasing such that*

$$T(f) = \int_a^b f d\alpha - \int_a^b f d\beta = \int_a^b f d(\alpha - \beta)$$

(but careful, because  $\alpha - \beta$  is not necessarily monotone increasing, but can just define difference of monotone increasing functions; call these “functions of bounded variation”).



How would we go about proving this? Comes from Hahn-Banach Theorem from functional analysis. Banach spaces are complete normed vector spaces. If you have a linear function from a dense subspace to the reals, can extend this to the entire space to the reals. Won't prove this... requires axiom of choice (even though Riesz Representation doesn't require it). And then we're in good shape... get  $T$  is the integral with various test functions on dense subspace.

## 6 January 24

**Theorem 11** (Rudin 6.13). *Let  $f, g \in \mathcal{R}_\alpha[a, b]$ .*

(a). *Then  $fg \in \mathcal{R}_\alpha[a, b]$ .*

(b). *Then  $|f| \in \mathcal{R}_\alpha[a, b]$  and  $|\int_a^b f d\alpha| \leq \int_a^b |f| d\alpha$*

*Proof.* First proof of (a). By theorem 6.11,  $\alpha(x) = x^2$ ,  $(f+g)^2, (f-g)^2 \in \mathcal{R}_\alpha[a, b]$ . By theorem 6.12a,  $(f+g)^2 - (f-g)^2 = 4fg \in \mathcal{R}_\alpha[a, b]$ . By theorem 6.12a,  $(c = \frac{1}{4}) fg \in \mathcal{R}_\alpha[a, b]$ .

Now proof of (b). By theorem 6.11  $\alpha(x) = |x|$ ,  $|f| \in \mathcal{R}_\alpha[a, b]$ . Let  $c = \operatorname{sgn} \int_a^b f d\alpha$  so  $|\int_a^b f d\alpha| = c \int_a^b f d\alpha = \int_a^b c f d\alpha \leq \int_a^b |f| d\alpha$  (by theorem 6.12a).  $\square$

**Theorem 12** (Rudin 6.15). *Let  $f: [a, b] \rightarrow \mathbb{R}$  be bounded. Let  $s \in (a, b)$  and suppose  $f$  is continuous at  $s$ . Let  $\alpha(x) = \begin{cases} 0 & x \leq s \\ 1 & x > s \end{cases}$ . Then  $f \in \mathcal{R}_\alpha[a, b]$  and  $\int_a^b f d\alpha = f(s)$ .*

*Proof.* Let  $P = \{x_0, x_1, x_2, x_3\}$  where  $a = x_0, s = x_1, b = x_3$ . Then  $U(P, f, \alpha) = \sum_{i=1}^3 M_i \Delta\alpha_i = M_2 = \sup_{x \in [x_1, x_2]} f(x)$  and  $L(P, f, \alpha) = \sum_{i=1}^3 m_i \Delta\alpha_i = m_2 = \inf_{x \in [x_1, x_2]} f(x)$ . Since  $f$  is continuous at  $s$ ,  $\forall \varepsilon > 0$ ,  $\exists \delta$  such that if  $|x - y| < \delta$ , then  $|f(x) - f(y)| < \frac{\varepsilon}{2}$ , so let  $x_2 \in (x_1, x_1 + \delta)$ . Then,

$$\sup_{x \in [x_1, x_2]} f(x) \leq f(x_1) + \varepsilon/2 \implies M_2 \leq f(x_1) + \varepsilon/2$$

$$\inf_{x \in [x_1, x_2]} f(x) \geq f(x_1) - \varepsilon/2 \implies m_2 \geq f(x_1) - \varepsilon/2$$

Hence,  $M_2 - m_2 \leq \varepsilon$ .  $\square$

How would we change proof if  $f(x) = 1$  for  $x \geq s$ ? We would get the same result, but would need to change which switch  $x_1, x_2$  roles. If defined at neither, then probably  $s \in [x_1, x_2]$ .

This step function is actually quite important in electrical engineering. Even has a special name:

*Definition 8* (Heavyside step function).

$$I(x) = \begin{cases} 0, & x \leq 0 \\ 1, & x > 0 \end{cases}$$

**Theorem 13** (Rudin 6.16). *Let  $\{c_n\}_{n=1}^\infty$  be positive real numbers, with  $\sum_{n=1}^\infty c_n < \infty$ . Let  $[a, b]$  be an interval, and let  $\{s_n\}_{n=1}^\infty \subset (a, b)$  be distinct points. Let  $\alpha(x) = \sum_{n=1}^\infty c_n I(x - s_n)$  (a bunch of steps at  $s_n$  by  $c_n$ ). Let  $f: [a, b] \rightarrow \mathbb{R}$  be continuous, so  $f \in \mathcal{R}_\alpha[a, b]$  (always true,  $f$  is continuous and  $\alpha$  monotone increasing). Then*

$$\int_a^b f d\alpha = \sum_{n=1}^\infty c_n f(s_n)$$

(We know the integral exists, and the sum exists because  $\sum c_n < \infty$  and  $f$  is bounded.) "I love nitpicking, because math is meant to be precise."

*Proof.* Let  $R_N = \int_a^b f d\alpha - \sum_{n=1}^N c_n f(s_n)$ . Goal:  $\forall \varepsilon > 0, \exists N_0$  such that  $\forall N \geq N_0, |R_N| < \varepsilon$ . So fix  $N$ , let  $\alpha_1 = \sum_{n=1}^N c_n I(x - s_n)$ ,  $\alpha_2 = \sum_{n=N+1}^\infty c_n I(x - s_n)$ . By Theorem 6.12e,  $\int_a^b f d\alpha = \int_a^b f d\alpha_1 + \int_a^b f d\alpha_2$  (we meet the hypothesis that  $f \in \mathcal{R}_{\alpha_1}, \mathcal{R}_{\alpha_2}$ , since  $f$  is continuous).

$$\int_a^b f d\alpha_1 = \sum_{n=1}^N \int_a^b f(x) d[c_n I(x - s_n)] = \sum_{n=1}^N c_n f(s_n)$$

(by Theorem 6.15). So  $R_N = \int_a^b f d\alpha_2$ . Let  $K = \sup_{x \in [a,b]} |f|$ . By 6.12b,  $\int_a^b f d\alpha_2 \leq K \int_a^b 1 d\alpha_2 = K(\alpha_2(b) - \alpha_2(a)) = K \sum_{n=N+1}^\infty c_n \rightarrow 0$  as  $N \rightarrow \infty$ . Overview: we wrote our difference in terms of a main term and tail. And we showed the tail goes to zero (we can make our proof more formal with  $\varepsilon$ , etc., but out of time).  $\square$

## 7 January 26

**Theorem 14** (Rudin 6.17). *Let  $f: [a, b] \rightarrow \mathbb{R}$  bounded,  $\alpha: [a, b] \rightarrow \mathbb{R}$  differentiable and monotone increasing. Suppose  $\alpha' \in \mathcal{R}[a, b]$ . Then  $f \in \mathcal{R}_\alpha[a, b] \iff f\alpha' \in \mathcal{R}[a, b]$  and if so*

$$\int_a^b f d\alpha = \int_a^b f \alpha' dx$$

Step 1: it suffices to show that

$$\overline{\int_a^b f d\alpha} = \overline{\int_a^b f \alpha' dx} \quad \text{and} \quad \underline{\int_a^b f d\alpha} = \underline{\int_a^b f \alpha' dx}$$

We will prove the first equality, second one is an exercise.

Step 2: since  $\alpha' \in \mathcal{R}[a, b]$  for all  $\varepsilon > 0$ ,  $P$  (of  $[a, b]$ ) such that

$$U(P, \alpha') - L(P, \alpha') < \varepsilon$$

This inequality continues to hold for every refinement  $P'$  of  $P$ .

We have  $U(P, \alpha') - L(P, \alpha') = \sum_{i=1}^n (A_i - a_i) \Delta x_i$ , where  $A_i = \sup\{\alpha'(x) : x \in [x_{i-1}, x_i]\}$  and likewise for  $a_i$ . By the mean value theorem for each  $i = 1, \dots, n$   $\exists t_i \in [x_{i-1}, x_i]$  such that  $\Delta \alpha_i = \alpha'(t_i) \Delta x_i$ . Now this suggests  $\overline{\int f d\alpha} = \overline{\int f \alpha' dx}$ , but need to be careful.

For every  $s_i \in [x_{i-1}, x_i]$ , we have  $|\alpha'(s_i) - \alpha'(t_i)| \leq A_i - a_i$  (literally did this so many times on the homework lol) so  $\sum_{i=1}^n |\alpha'(s_i) - \alpha'(t_i)| \Delta x_i \leq \sum_{i=1}^n (A_i - a_i) \Delta x_i$  for every choice of  $s_i \in [x_{i-1}, x_i]$ ,  $i = 1, \dots, n$ . Let  $K = \sup_{x \in [a,b]} |f|$ , then

$$\sum_{i=1}^n |f(s_i) \alpha'(s_i) \Delta x_i - f(s_i) \alpha'(t_i) \Delta x_i| \leq K\varepsilon \quad (2)$$

Hence

$$\sum_{i=1}^n f(s_i) \Delta \alpha_i \leq \sum_{i=1}^n f(s_i) \alpha'(s_i) \Delta x_i + K\varepsilon \leq U(P, f\alpha') + K\varepsilon$$

Recall triangle inequality: if  $|a + b| \leq c$ , then  $a \leq b + c$  and  $b \leq a + c$ .

When we have an inequality, might wonder where the inequality is “sharp” or “tight”: i.e. equality or the closest to equality as possible. Taking the supremum of  $s_i \in [x_{i-1}, x_i]$ ,  $i = 1, \dots, n$ , we conclude (if  $M_i = \sup\{f(x) : x \in [x_{i-1}, x_i]\}$ )

$$U(P, f, \alpha) = \sum_{i=1}^n M_i \Delta \alpha_i \leq U(P, f\alpha') + K\varepsilon$$

Hence

$$\overline{\int_a^b f d\alpha} \leq U(P, f, \alpha) \leq U(P, f\alpha') + K\varepsilon$$

So for every  $\varepsilon$ , we found some partition  $P$  that makes the inequality hold. Recall that it also holds for every refinement  $P'$  of  $P$ .

How can we make this inequality be as close to equality as possible (how much strength can squeeze out of inequality). Taking the infimum over all refinements  $P'$  of  $P$ , we have

$$\overline{\int_a^b f d\alpha} \leq \int_{P'} U(P', f\alpha) + K\varepsilon$$

We are taking a more restricted set of partitions, will this give us the infimum we want? Yes: for any non-refinement partition, just union it with  $P$  to get a refinement, i.e.  $\forall \varepsilon > 0, \overline{\int_a^b f d\alpha} \leq \overline{\int_a^b f \alpha' dx} + KE \implies \overline{\int_a^b f d\alpha} \leq \overline{\int_a^b \alpha' dx}$ .

Once you've seen this, can do the other three inequalities. Will just say a word or two about the reverse inequality, i.e.

$$\overline{\int_a^b f d\alpha} \geq \overline{\int_a^b f \alpha' dx}$$

Taking (2), get

$$\sum_{i=1}^n f(s_i) \alpha'(s_i) \Delta x_i \leq \sum_{i=1}^n f(s_i) \Delta \alpha_i + K\varepsilon$$

Etc.

## 8 January 29

For the next theorem, note that a continuous bijective map must be strictly monotone (good exercise).

**Theorem 15** (Change of Variables (Rudin 6.19)). *Let  $\phi: [A, B] \rightarrow [a, b]$ , strictly increasing, onto (i.e.  $1-1$ ), continuous. Let  $\alpha: [a, b] \rightarrow \mathbb{R}$  monotone increasing. Let  $f \in \mathcal{R}_\alpha[a, b]$ . Define  $g = f \circ \phi: [A, B] \rightarrow \mathbb{R}$ ,  $\beta: \alpha \circ \phi: [A, B] \rightarrow \mathbb{R}$ . Then  $g \in \mathcal{R}_\beta[A, B]$  and*

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

*Remark 2.* Let  $\alpha(x) = x$ , and  $\phi$  is differentiable. Then  $d\beta = \phi'(x)dx$  (why??), i.e.

$$\int_a^b f d\alpha = \int_{\phi^{-1}(a)}^{\phi^{-1}(b)} f(x) \phi'(x) dx$$

And this lines up with our normal change of variables formula.

*Proof.* Partitions  $P = \{a = x_0, x_1, \dots, x_n = b\}$  of  $[a, b]$  and partitions  $Q$  of  $[A, B]$  are in  $1-1$  correspondance via  $x_i = \phi(y_i)$ . We have  $\alpha(x_i) = \alpha \circ \phi(y_i) = \beta(y_i)$  and  $\{f(x): x \in [x_{i-1}, x_i]\} = \{g(y): y \in [y_{i-1}, y_i]\}$ . Then  $U(P, f, \alpha) = U(Q, g, \beta)$  and  $L(P, f, \alpha) = L(Q, g, \beta)$ .

Let  $\varepsilon > 0$ . Since  $f \in \mathcal{R}_\alpha[a, b]$ , there exists a partition  $P$  of  $[a, b]$  such that  $U(P, f, \alpha) - L(P, f, \alpha) < \varepsilon$ , hence  $U(Q, g, \beta) - L(Q, g, \beta) < \varepsilon$ . Then  $g \in \mathcal{R}_\beta[A, B]$ . So first part of theorem done ✓

Finally,

$$\int_A^B g d\beta = \inf_Q U(Q, g, \beta) = \inf_P U(P, f, \alpha) = \int_a^b f d\alpha$$

□

Recall the hypotheses of this theorem:  $\phi$  is continuous,  $1-1$  and onto. Where did we use them? Where we got a  $1-1$  correspondance between  $P$  and  $Q$ , and makes the rest of the proof fall.

Now we start inching towards the fundamental theorem of calculus.

**Theorem 16** (Rudin 6.20). *Let  $f \in \mathcal{R}[a, b]$ . For  $x \in [a, b]$ , define  $F(x) = \int_a^x f(t) dt$ ,  $F(a) = 0$ . Then:  $F$  is continuous on  $[a, b]$ . And if  $c \in [a, b]$  and  $f$  is continuous at  $c$ , then  $F$  is differentiable at  $c$ , and  $F'(c) = f(c)$ .*

*Proof.* Proving continuity: Let  $K = \sup_{t \in [a,b]} |f(t)|$ . By theorem 6.12c (linear in the domain), for  $a \leq x \leq y \leq b$

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

Essentially, the biggest this can be is  $K$ , and so the largest the integral can be is  $\int_a^b Kdt$ , but have to do this carefully. Thus

$$|F(y) - F(x)| = \left| \int_x^y f(t)dt \right| \leq \int_x^y |f(t)|dt \leq \int_x^y Kdt = K(y - x)$$

(where we used theorem 6.13 for the first inequality, and 6.12b for the second). Hence, for every  $\varepsilon > 0$ , select  $\delta = \varepsilon/K$  (or  $\delta = \varepsilon$  if  $K = 0$ ). Then if  $|x - y| < \delta$ ,  $|F(x) - F(y)| < \varepsilon$ .

Proving differentiability at  $c$ : suppose  $c \neq b$ , i.e.  $c \in [a, b)$ . Let us compute  $\lim_{h \searrow 0} \frac{F(c+h) - F(c)}{h}$ . For  $h > 0$ , we have

$$\left| \frac{1}{h} [F(c+h) - F(c)] - f(c) \right| = \left| \frac{1}{h} \int_c^{c+h} f(t)dt - f(c) \right|$$

In analysis, we often want to exploit some cancellation between two quantities we think of as close together. We are sampling values of  $f$  quite close to  $c$ , and comparing them to  $f(c)$ . How do we exploit cancellation? One common technique when a quantity is not changing much over a region is take its average and bring it inside an integral to cancel. So, since  $f(c) = \frac{1}{h} \int_c^{c+h} f(c)dt$ , we get

$$= \left| \frac{1}{h} \int_c^{c+h} (f(t) - f(c))dt \right| \leq \frac{1}{h} \int_c^{c+h} |f(t) - f(c)|dt$$

(triangle inequality for integrals?) Since  $f$  is continuous at  $c$ ,  $\forall \varepsilon > 0$ ,  $\exists \delta > 0$  such that for all  $y \in [a, b]$  with  $|y - c| < \delta$ , we have  $|f(c) - f(y)| < \varepsilon$ . So for  $h < \delta$ , we have

$$\frac{1}{h} \int_c^{c+h} |f(t) - f(c)|dt < \frac{1}{h} \int_c^{c+h} \varepsilon dt = \varepsilon$$

i.e.  $\forall \varepsilon > 0$ ,  $\exists \delta > 0$  such that  $\forall 0 < h < \delta$ ,

$$\left| \frac{1}{h} (F(c+h) - F(c)) - f(c) \right| < \varepsilon$$

i.e.  $\lim_{h \searrow 0} \frac{F(c+h) - F(c)}{h} = f(c)$ .

If  $c \neq a$ , i.e. if  $c \in (a, b]$ , an identical argument shows  $\lim_{h \nearrow 0} \frac{F(c+h) - F(c)}{h} = f(c)$ .  $\square$

Question: If  $f$  is not continuous at  $c$ , is  $F$  never differentiable at  $c$ , maybe differentiable at  $c$ , or is always differentiable at  $c$  (our trichotomy).

*Remark 3.* If  $f(x) = g(x)$  except at one point, then  $\int_a^b f(x)dx = \int_a^b g(x)dx$ . I.e. the Riemann integral does not see the value of the function at one point (not true generally for Riemann-Stieltjes, we saw  $\alpha$  is Heaviside step function only cared about the function at a single point).

When we have a jump discontinuity,  $F$  will never be continuous at a point.

## 9 January 31!

Typo from last class: ff

Last class: If  $f: [a, b] \rightarrow \mathbb{R}$  continuous, and  $F(x) = \int_a^x f(t)dt$ , then  $F'(x) = f(x)$  for all  $x \in [a, b]$ .

**Theorem 17** (Fundamental Theorem of Calculus (Rudin 6.21)). *Let  $f \in \mathcal{R}[a, b]$ , let  $F: [a, b] \rightarrow \mathbb{R}$  be differentiable and suppose  $F'(x) = f(x)$  for  $x \in [a, b]$ . Then  $\int_a^b f(x)dx = F(b) - F(a)$ .*

*Proof.* By MVT (“Most Valuable Theorem”), for any partition  $P = \{x_0, \dots, x_n\}$  there are numbers  $t_i \in [x_{i-1}, x_i]$  for  $i = 1, \dots, n$  such that  $F'(t_i) = (F(x_i) - F(x_{i-1}))/\Delta x_i$ . So

$$\begin{aligned} F(b) - F(a) &= \sum_{i=1}^n (F(x_i) - F(x_{i-1})) \\ &= \sum_{i=1}^n F'(t_i) \Delta x_i \\ &= f f \end{aligned}$$

Then  $\left| \int_a^b f dx - (F(b) - F(a)) \right| \leq U(P, f) - L(P, f)$ .  $f \in \mathcal{R}[a, b]$  implies that for all  $\varepsilon > 0$ , there exists a partition  $P$  such that  $U(P, f) - L(P, f) < \varepsilon$  and hence  $\left| \int_a^b f dx - (F(b) - F(a)) \right| < \varepsilon$ .  $\square$

Now we can go and prove the things we know about integration. The first one, integration by parts, we kinda already have.

**Theorem 18** (Integration by Parts (Rudin 6.22)). *Let  $F, G: [a, b] \rightarrow \mathbb{R}$  be differentiable. Let  $f = F'$ ,  $g = G'$ , and suppose  $f, g \in \mathcal{R}[a, b]$ . Then*

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

*Proof.* Let  $H(x) = F(x)G(x)$ . Then  $H'(x) = f(x)G(x) + F(x)g(x) \in \mathcal{R}[a, b]$ . Apply Theorem 6.21 to  $H$ , then  $H(b) - H(a) = f f$   $\square$

In both of these results, we have this hypothesis that  $f, g$  are Riemann integrable. Question: If  $F: [a, b] \rightarrow \mathbb{R}$  is differentiable and  $F' = f$ , do we need to repeat  $f \in \mathcal{R}[a, b]$ , or does this hold automatically, i.e. is  $F' \in \mathcal{R}[a, b]$  for every  $F: [a, b] \rightarrow \mathbb{R}$  differentiable?

Does there exist  $F: [a, b] \rightarrow \mathbb{R}$  differentiable, so that  $F'$  is discontinuous at every  $x \in [a, b]$ ? “We’ve replaced a hard question with a harder question.” Won’t figure this out in this class, but the answer is NO. It is an interesting question: which sets can be the set of discontinuities of a derivative? He gives you  $S \subset [0, 1]$ , can you give a  $F'$  discontinuous at  $S$  (where  $F: [0, 1] \rightarrow \mathbb{R}$  is differentiable). These are called F- $\sigma$  sets.

Perhaps we want the derivative to blow up, then it isn’t Riemann integrable. Here is a function to keep in your back pocket:

$$F(x) = \begin{cases} 0, & x = 0 \\ x^2 \sin \frac{1}{x^2}, & x \neq 0 \end{cases}$$

This function is differentiable, but its derivative is unbounded. On  $x_n = \frac{1}{\sqrt{\pi n}}$ ,  $F'(x_n)$  blows up.

Another type of counter-example:  $F'$  is bounded, but  $F'$  is discontinuous at so many places that it is not Riemann integrable. Uncountable isn’t enough: they might still be Riemann integrable. The condition is that it is discontinuous at points with positive Lebesgue measure: try to cover all the discontinuities with open intervals, the smallest we can make the intervals will always add up to a positive value. But this is for Math 420.

For the end of class, we will get some definitions to explore when the interval of integration is not a closed interval.

**Definition 9.** If  $f: [a, \infty) \rightarrow \mathbb{R}$  satisfies  $f \in \mathcal{R}[a, b]$  for all  $b > a$ , then we define

$$\int_a^b f(x)dx = \lim_{b \rightarrow \infty} \int_a^b f(x)dx$$

If the limit exists (as a real number), we say that the  $\int_a^\infty f(x)dx$  converges. If  $\int_a^\infty |f|dx$  exists (as a real number), then we say  $\int_a^\infty f(x)dx$  converges *absolutely*.

(This is the same idea as conditional/absolute convergence of a sequence.) We can make an equivalent definition for  $\int_{-\infty}^b f(x)dx$ .

If  $f: \mathbb{R} \rightarrow \mathbb{R}$  and both  $\int_0^\infty f(x)dx$  and  $\int_{-\infty}^0$  converges (absolutely), we define  $\int_{-\infty}^\infty f(x)dx = \int_{-\infty}^0 f(x)dx + \int_0^\infty f(x)dx$ , and we say  $\int_{-\infty}^\infty f(x)dx$  converges (absolutely).

Can we construct a function that converges but not absolutely? Taking inspiration from series, we can take a step function of  $\frac{(-1)^n}{n}$ .

Next time, we'll finally look sequences of functions.

## 10 Feburary 2

### 10.1 Sequences and series of functions

The setup: Let  $E$  be a set ( $E = [a, b]$ ,  $E = \mathbb{R}$ ). Let  $\{f_n\}_{n=1}^\infty$  be a sequence of functions  $f_n: E \rightarrow S$  (usually  $S = \mathbb{R}$ , or  $\mathbb{C}$ , or some metric space  $(M, d)$ ) and let  $f: E \rightarrow S$ .

What does it mean for  $f_n$  to converge to  $f$ , i.e.  $f_n \rightarrow f$ ? If  $f_n \rightarrow f$ , and the functions  $f_n$  all have some property, must it be true that  $f$  also has that property?

#### 10.1.1 Pointwise convergence

ff missed a whole chunk late to class

*Definition 10* (Pointwise Convergence (Rudin 7.1)). Let  $\{f_n\}_{n=1}^\infty$  be a sequence of functions  $f_n: E \rightarrow M$ . If the sequence  $\{f_n(x)\}_{n=1}^\infty$  converges for every  $x \in E$ . Then we say  $\{f_n\}$  converges pointwise (on  $E$ ). If this happens, define  $f(x) = \lim_{n \rightarrow \infty} f_n(x)$ ,  $f(x): E \rightarrow M$ , and we say  $f_n \rightarrow f$  pointwise (on  $E$ ).

Note that there are two definitions one might use for convergence of functions.

- $\forall \varepsilon > 0, \exists N, \forall n \geq N, \forall x \in E, d(f_n(x) - f(x)) < \varepsilon$ .
- $\forall x \in E, \varepsilon > 0, \exists N, \forall n \geq N, d(f_n(x) - f(x)) < \varepsilon$ .

The second is the definition we did above. The first is called uniform convergence, and we will cover later.

Now, let us consider the second question we asked: for pointwise convergence, what properties hold?

- Continuity:  $E = \mathbb{R}$ . If each  $f_n$  is continuous and  $f_n \rightarrow f$  pointwise, must it be true that  $f$  is continuous? No:  $f_n = e^{-nx^2}$  and  $f = \begin{cases} 1, & x = 0 \\ 0, & x \neq 0 \end{cases}$ , but  $f_n \rightarrow f$ . Also,  $f_n = x^n$  and  $f = \begin{cases} 1, & x = 1 \\ 0, & \text{elsewhere} \end{cases}$  where  $E = [0, 1]$ , but  $f_n \rightarrow f$ . In fact, we can make our points of discontinuity arbitrarily where we want.
- Boundedness: If each  $f_n$  is bounded and  $f_n \rightarrow f$  pointwise, must  $f$  be bounded? The converse doesn't hold ( $f_n = \frac{x}{n} \rightarrow 0$ ). We have  $f_n(x) = \begin{cases} x, & x \in [-n, n] \\ n, & x \in (n, \infty) \\ -n, & x \in (-\infty, -n) \end{cases}$  which pointwise converges to  $f(x) = x$ .
- Quantitative boundedness: If each  $f_n$  is bounded by 1, i.e.  $|f_n(x)| \leq 1, \forall x \in E$ , and  $f_n \rightarrow f$  pointwise, must it be true that  $|f| \leq 1, \forall x \in E$ . The answer is yes: contradiction, assume  $|f| > 1$ , so  $|f(x)| = 1 + \varepsilon > 1$ , there will be some  $f_n(x) > 1$ .

If we change our previous condition to  $< 1$ , this is no longer true, can get  $|f| = 1$ .

- Integrability:  $f_n \in \mathcal{R}[0, 1]$ ,  $f_n \rightarrow f$  pointwise, must it be true that  $f \in \mathcal{R}[0, 1]$ . We don't know many functions that fail to be integrable, but one is  $f = \begin{cases} 1, & x \in \mathbb{Q} \\ 0, & x \notin \mathbb{Q} \end{cases}$ . Note that  $f_n(x) = \begin{cases} 1, & x \in \{q_1, \dots, q_n\} \\ 0, & \text{elsewhere} \end{cases}$ , where  $q_n$  is an enumeration of the rationals.
- Food for thought: assuming  $f_n$  and  $f$  are both integrable, must the values of their integrals converge?

## 11 February 5

Say  $f_n \rightarrow f$  pointwise. Recall we found that integrability of  $f_n$  does not imply integrability of their limit. But we asked: If  $\{f_n\}$  are Riemann integrable and  $f$  Riemann integrable, must it be true that  $\lim_{n \rightarrow \infty} \int_0^1 f_n dx \rightarrow \int_0^1 f dx$ . My counter-example (inspired by height of  $\frac{1}{n}$  and width of  $n$ ):

$$g_n(x) = \begin{cases} n, & 0 < x < \frac{1}{n} \\ 0, & \text{otherwise} \end{cases}$$

It is clear the sequence converges to the 0 function, but the area under the curve remains 1. If we wanted to make it so our limit doesn't even exist, then we could use  $n^2$ , or something like  $(-1)^n n$ . On the homework, we add in some extra conditions that make it true, that is monotonicity. However, a monotone sequence of functions doesn't necessarily mean the integrals converge:  $f$  may not be in  $\mathcal{R}[a, b]$ , as we saw from the counter-example last time. But if we assume  $f \in \mathcal{R}[a, b]$ , then it is enough.

We have seen examples of  $f_n \rightarrow f$  pointwise, where  $f_n$  are continuous or integrable, and the limit is not. What is going on? Suppose  $f_n \rightarrow f$ ,  $f: \mathbb{R} \rightarrow \mathbb{R}$ . Let  $f_n$  continuous at  $c \in \mathbb{R}$ , is  $f$  continuous at  $c$ ? Well, we know  $\lim_{x \rightarrow c} f(x) = f(c)$ , so  $f$  continuous at  $c$  if

$$\lim_{x \rightarrow c} \lim_{n \rightarrow \infty} f_n(x) = \lim_{n \rightarrow \infty} \lim_{x \rightarrow c} f_n(x)$$

But we cannot interchange the limit, we know examples where this fails, and this leads to the failure of  $f$  being continuous.

We can do something similar with Riemann integrability, but we have to be a bit more careful with it. Recall,  $f$  is Riemann integrable (on  $[0, 1]$ ) if  $\lim_{m \rightarrow \infty} U(P_m, f) - L(P_m, f) = 0$  (from homework).  $f_n \rightarrow f$  pointwise, is it true that

$$\lim_{m \rightarrow \infty} U(P_m, \lim_{n \rightarrow \infty} f_n) = \lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} U(P_m, f_n)$$

(and likewise with  $L(P_m, f_n)$ ). Even though the limit is inside, this is taking the limits in different order. The examples we have been providing have been using the failure to interchange limits. We saw Lipschitz continuity last term: it also fails to hold in a pointwise limit, because of this interchange of limits. .

One final example showing you *cannot* (in general) interchange limits (an example to keep in mind).

$$a_{n,m} = \frac{n}{n+m}, \quad n, m \in \mathbb{N}$$

Easy to see that

$$\begin{aligned} \lim_{n \rightarrow \infty} a_{n,m} = 1 &\implies \lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} a_{n,m} = 1 \\ \lim_{m \rightarrow \infty} a_{n,m} = 0 &\implies \lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} a_{n,m} = 0 \end{aligned}$$

This is somehow the most simple example. Most of the examples we gave yesterday were of this flavour.

### 11.1 Uniform convergence

It is harder for functions to converge uniformly, but many of the properties we want carry over.

**Definition 11** (Uniform convergence (Rudin 7.7)). Let  $E$  be a set, (i.e.  $\mathbb{R}$ , or  $[a, b]$ ),  $(M, d)$  a metric space (i.e.  $\mathbb{R}, \mathbb{C}$ ). Let  $\{f_n\}$  be a sequence of functions  $f_n: E \rightarrow M$  and  $f: E \rightarrow M$ . We say  $f_n \rightarrow f$  *uniformly* if:  $\forall \varepsilon > 0, \exists N$  such that  $\forall n > N, x \in E, d(f_n(x), f(x)) < \varepsilon$ .

We talked in 320 about sequences that converge, and talked about sequences that are Cauchy. If we have a complete metric space, our Cauchy sequences and convergent sequences are the same. Something similar happens here:

**Theorem 19** (Cauchy Criteria for Uniform Convergence (Rudin 7.8)). Let  $E$  be a set,  $(M, d)$  a complete metric space. Let  $\{f_n\}$  be a sequence,  $f_n: E \rightarrow M$ . Then  $\{f_n\}$  converges uniformly (to some  $f: E \rightarrow M$ ) iff  $\forall \varepsilon > 0, \exists N$  such that  $\forall n, m \geq N, x \in E, d(f_n(x), f_m(x)) < \varepsilon$ .

Digression: if we consider the functions as a metric space, say  $(X, \rho)$  where  $X$  are the functions  $f: [0, 1] \rightarrow \mathbb{R}$  and  $\rho(f, g) = \sup|f(x) - g(x)|$ , this in fact is a complete metric space and so we do get that Cauchy sequences are the convergent sequences. But this is actually the metric corresponding to uniform convergence: to prove that this is a complete metric space, probably use the Cauchy Criteria. If we had changed  $X$  to  $f: [0, 1] \rightarrow \mathbb{Q}$ ,  $(X, \rho)$  is no longer a complete metric space, as Cauchy Criteria would predict.

*Proof.* Forwards: Suppose  $f_n \rightarrow f$  uniformly. Let  $\varepsilon > 0$ ,  $\exists N$  such that  $\forall n \geq N, x \in E, d(f_n(x), f(x)) < \varepsilon/2$ . So  $\forall n, m \geq N, x \in E$ ,

$$d(f_n(x), f_m(x)) \leq d(f_n(x), f(x)) + d(f_m(x), f(x)) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

Backwards: Suppose  $\{f_n\}$  satisfies the Cauchy Criteria. For each  $x \in E$ ,  $\{f_n(x)\}$  is a Cauchy sequence in  $M$  and  $(M, d)$  is complete, so  $\{f_n(x)\}$  converges, i.e.  $\lim_{n \rightarrow \infty} f_n(x)$  exists. Define  $f(x) = \lim_{n \rightarrow \infty} f_n(x)$ . We have shown pointwise convergence. Final goal: show that  $f_n \rightarrow f$  *uniformly*. Let  $\varepsilon > 0$ . Since  $\{f_n\}$  satisfies the Cauchy Criteria,  $\exists N$  such that  $n, m \geq N, x \in E$  we have

$$d(f_n(x), f_m(x)) < \frac{\varepsilon}{2}$$

Hence,  $d(f_n(x), f(x)) \leq d(f_n(x), f_m(x)) + d(f_m(x), f(x)) \leq \frac{\varepsilon}{2} + d(f_m(x), f(x))$ . Since  $f_m(x) \rightarrow f(x)$  (for this particular  $x$ ), we can select  $m \geq N$  such that  $d(f_m(x), f(x)) < \varepsilon/2$ , hence

$$d(f_n(x), f(x)) < \varepsilon$$

□

Subtle point: we first use the fact that  $M$  is complete to show that  $f_n \rightarrow f$  pointwise, and then we choose  $m$  for the Cauchy Criteria, but that is okay because it is for fixed  $x$ .

There is a cluster of true theorems around this: i.e. don't need complete metric space if  $f_n \rightarrow f$  pointwise, and also an equivalent Cauchy Criteria changing quantifiers for pointwise convergence.

## 12 February 7

**Theorem 20** (Rudin 7.11). *Let  $(M_1, d_1)$  and  $(M_2, d_2)$  be metric spaces, with  $(M_2, d_2)$  complete, i.e.  $\mathbb{R}$  or  $\mathbb{C}$ . Let  $E \subset M_1$ , let  $\{f_n\}$  be a sequence of functions  $f_n: E \rightarrow M_2$ , and suppose  $f_n \rightarrow f$  uniformly on  $E$ . Let  $E \in M_1$  be a limit point of  $E$ . Suppose  $\lim_{t \rightarrow x} f_n(t) = y_n$  exists for each  $n$ . Then  $\{y_n\}$  is a convergent sequence, i.e.  $y_n \rightarrow y \in M_2$ , and  $\lim_{t \rightarrow x} f(t) = y$ , i.e.*

$$\lim_{t \rightarrow x} \underbrace{\lim_{n \rightarrow \infty} f_n(t)}_{f(t)} = \lim_{n \rightarrow \infty} \underbrace{\lim_{t \rightarrow x} f_n(t)}_{y_n}$$

**Corollary 1.** *Let  $(M_1, d_1), (M_2, d_2), \{f_n\}, f, E$  be as before. If each  $f_n$  is continuous on  $E$ , and  $f_n \rightarrow f$  uniformly, then  $f$  is continuous on  $E$ . "The uniform limit of continuous functions is continuous."*

We first prove that the theorem implies the corollary.

*Proof.* Isolated points are always continuous, so we need only to consider limit points  $x \in E \cap E'$ . For every such  $x$ , Theorem 7.11 implies

$$f(x) = \lim_{n \rightarrow \infty} f_n(x) = \lim_{n \rightarrow \infty} \underbrace{\lim_{t \rightarrow x} f_n(t)}_{f_n \text{ cts}} = \lim_{t \rightarrow x} \lim_{n \rightarrow \infty} f_n(t) = \lim_{t \rightarrow x} f(t)$$

□

We now proceed with the proof of the theorem.



*Proof.* Step 1: Show that  $\{y_n\}$  converges (to some  $y \in M_2$ ). It suffices to  $\{y_n\}$  is Cauchy. Let  $\varepsilon > 0$ . Choose  $N$  such that  $\forall n, m \geq 2N, t \in E, d_2(f_n(t), f_m(t)) < \varepsilon/2$  by the Cauchy Criterion. Then  $d_2(y_n, y_m) \leq d_2(y_n, f_n(t)) + d_2(f_n(t), y_m) \leq d_2(y_n, f_n(t)) + d_2(f_n(t), f_m(t)) + d_2(f_m(t), y_m)$ . We can choose  $t$  such that the above is at most  $< \varepsilon/3 + \varepsilon/3 + \varepsilon/3$ . Now we call this the “ $\varepsilon$  over 3 trick” (won’t show full details in the future). Conclusion:  $\forall \varepsilon > 0, \exists N$  such that  $n, m \geq N, d_2(y_n, y_m) < \varepsilon$ , i.e.  $\{y_n\}$  is Cauchy, hence convergent ( $(M_2, d_2)$  complete).

Step 2: Prove that  $f(t) \rightarrow y$  as  $t \rightarrow x$ . We have  $\forall t \in E$  and  $n, d_2(f(t), y) \leq d_2(f(t), f_n(t)) + d_2(f_n(t), y_n) + d_2(y_n, y)$ . We could first fix  $N$  large, and then make  $t$  close to  $x$ , or we could do it in the reverse order. It will turn out that the former will be more fruitful. Let  $\varepsilon > 0$ . Since  $f_n \rightarrow f$  uniformly,  $\exists N_1$  such that  $\forall n \geq N_1, t \in E, d_2(f(t), f_n(t)) < \varepsilon/3$ . Since  $y_n \rightarrow y$ ,  $\exists N_2$  such that  $\forall n \geq N_2, d_2(y_n, y) < \varepsilon/3$ . Let  $n = \max(N_1, N_2)$ . Applying A1 (with this choice of  $n$ ), we have

$$d_2(f(t), y) < \varepsilon/3 + d_2(f_n(t), y_n) + \varepsilon/3$$

Since  $\lim_{t \rightarrow x} f_n(t) = y_n$ ,  $\exists \delta > 0$  such that  $\forall t \in E, d_1(t, x) < \delta$ , we have  $d_2(f_n(t), y_n) < \delta$ , we have  $d_2(f(t), y) < \varepsilon$ .  $\square$

Now, where did we use uniform convergence in this proof? We used it in Step 1 to prove that our  $y_n$  converge (Cauchy criterion). So what if we add the additional hypothesis that  $y_n$  converges, but we weaken it to pointwise convergence. Well, we use it by making  $N$  large for all  $t$ , so Step 2 goes wrong as well. This proof is not fixable with pointwise convergence.

## 12.1 Series of functions

Series are not as important as sequences of functions, but it is useful in everyday life.

*Definition 12.* Let  $E$  be a set, let  $\{f_n\}$  be a sequence of functions  $f_n: E \rightarrow \mathbb{R}$  or  $E \rightarrow \mathbb{C}$  and let  $g: E \rightarrow \mathbb{R}$  or  $E \rightarrow \mathbb{C}$ . We say that  $\sum_{n=1}^{\infty} f_n$  converges pointwise (resp. uniformly) to  $g$  on  $E$  if the sequence  $S_n = \sum_{i=1}^n f_i$  converges pointwise (resp. uniformly) to  $g$ .

Can’t do this in full arbitrary of  $f_n$ , needs to support addition. I guess you can do it in some group, but this is abstraction for no good reason.

Example: The series  $1 + \sum_{n=1}^{\infty} \frac{x^n}{n!}$ . Converges to  $g(x) = e^x$ :

- pointwise (on  $\mathbb{R}$ )
- uniformly on any bounded set  $E \subset \mathbb{R}$ , or specifically compact sets  $E \subset \mathbb{R}$ .

**Theorem 21** (Weierstrass M-test (Rudin 7.10)). *Let  $E$  be a set,  $E \rightarrow \mathbb{R}$  or  $E \rightarrow \mathbb{C}$ . If  $|f_n(x)| \leq M_n, \forall n \geq N_0, x \in E$ , and if  $\sum_{n=1}^{\infty} M_n < \infty$ , then  $\sum_{n=1}^{\infty} f_n$  converges uniformly.*

Proof is an exercise, but useful is Cauchy criterion. This theorem is almost trivial, but very practical.

## 13 February 12

Housekeeping: Midterm on Wednesday (bring ID, everything up to and including pointwise convergence). HW4 is graded, but bonus problem will be graded soon (by him).

Recall last class we talked about  $(X, d)$  a metric space, and  $\mathcal{C}(X)$  be the bounded continuous functions  $f: X \rightarrow \mathbb{C}$ . We talked about how  $\mathcal{C}(X)$  is a normed vector space that is complete.

What about  $\mathcal{C}(X, Y)$ , the bounded continuous functions  $f: X \rightarrow Y$  ( $X, Y$  are metric spaces) (bounded here means  $f(X)$  is contained in some  $r$ -ball in  $Y$ ). Our metric is  $d(f, g) = \sup_{x \in X} d_Y(f(x), g(x))$  (check this is a metric). Is  $\mathcal{C}(X, Y)$  complete? Yes iff  $Y$  is complete (same proof as before; consider constant function of Cauchy sequence that doesn’t converge).

### 13.1 Properties of Uniform Convergence

**Theorem 22** (Rudin 7.16). *Let  $\alpha: [a, b] \rightarrow \mathbb{R}$  monotone increasing. Let  $\{f_n\}$  a sequence  $f_n \in \mathcal{R}_\alpha[a, b]$ . Let  $f: [a, b] \rightarrow \mathbb{R}$  and suppose  $f_n \rightarrow f$  uniformly. Then  $f \in \mathcal{R}_\alpha[a, b]$  and  $\lim_{n \rightarrow \infty} \int_a^b f_n d\alpha = \int_a^b f d\alpha$ .*

*Proof.* We first show that  $f \in \mathcal{R}_\alpha[a, b]$ . We will show  $\int_a^b f d\alpha = \overline{\int_a^b f d\alpha}$ . Note that we can always assume that these upper and lower integrals exist: we just need to show that they are bound, so take  $\varepsilon = 1$ ,  $\exists N$  such that  $|f(x) - f_N(x)| \leq 1$  for all  $x$ , and  $|f_N(x)| < K$  since Riemann integrable, so  $|f(x)| \leq |f(x) - f_N(x)| + |f_N(x)| < K + 1$ .

Now, to show the upper and lower integrals are equal, we could show that their difference is less than  $\varepsilon$  for all  $\varepsilon > 0$ . Let  $\varepsilon > 0$ . Since  $f_n \rightarrow f$  uniformly, there exists  $N$  such that  $\forall n \geq N$ ,  $x \in [a, b]$ ,  $|f(x) - f_n(x)| \leq \varepsilon$ . Note  $f_n(x) - \varepsilon \leq f(x) \leq f_n(x) + \varepsilon$ , so

$$\int_a^b (f_n - \varepsilon) d\alpha \leq \underline{\int_a^b f d\alpha} \leq \overline{\int_a^b f d\alpha} \leq \int_a^b (f_n + \varepsilon) d\alpha$$

since each  $m_i$  is the greatest lower bound, it is greater than  $f_n - \varepsilon$  on each interval for any partition, so  $\int (f_n - \varepsilon) d\alpha \leq \underline{\int f d\alpha}$  but the upper and lower converge because it's Riemann integrable (this is only the first inequality, but the other follows the same). Rearranging, we get

$$\overline{\int_a^b f d\alpha} - \underline{\int_a^b f d\alpha} \leq \int_a^b (f_n + \varepsilon) d\alpha - \int_a^b (f_n - \varepsilon) d\alpha = \int_a^b 2\varepsilon d\alpha = 2\varepsilon(\alpha(b) - \alpha(a))$$

Since  $\varepsilon > 0$  is arbitrary,  $\underline{\int_a^b f d\alpha} = \overline{\int_a^b f d\alpha}$ .

Now we need to show that the integral agrees with the limit of the integrals of  $f_n$  (which we don't get for free, as we saw with the pointwise case). Rearranging our inequality chain, we get

$$\int_a^b f_n d\alpha - \int_a^b \varepsilon d\alpha \leq \int_a^b f d\alpha \leq \int_a^b f_n d\alpha + \int_a^b \varepsilon d\alpha$$

And so

$$\left| \int_a^b f d\alpha - \int_a^b f_n d\alpha \right| < \varepsilon(\alpha(b) - \alpha(a))$$

□

**Corollary 2.** If  $f_n \in \mathcal{R}_\alpha[a, b]$ , and  $\sum_{n=1}^\infty$  converges uniformly on  $[a, b]$  to  $f$ , then  $f \in \mathcal{R}_\alpha[a, b]$ , and

$$\int_a^b f d\alpha = \sum_{n=1}^\infty \int_a^b f_n d\alpha$$

*Proof.* Let  $g_n = \sum_{i=1}^n f_i$  and  $g_n \rightarrow f$  uniformly. Apply Theorem 7.16. (Exercise: write this down precisely yourself.) □

You will be moving an infinite sum inside an integral ( $\lim_{m \rightarrow \infty} \sum_{n=1}^m \int_a^b f_n d\alpha$  and move the sum inside), which we can't normally do, but the point of Theorem 7.16 is when we can do this.

Our next theorem will take some effort to prove. Note that we can have a series of differentiable functions that converge uniformly but is not differentiable. The functions  $f_0 = \sin(x)$ ,  $f_1 = \frac{1}{2} \sin(4x)$ ,  $f_n = \frac{1}{2^n} \sin(4^n x)$  are all differentiable, but  $\sum_{n=1}^\infty f_n$ , which uniformly converges (Weierstrass  $M$ -test) is not differentiable (actually not differentiable anywhere, hard to show, Weierstrass function??). The derivative of  $f_n$  is  $\frac{4^{n^2}}{2^n} \cos(4^n x)$ , and basically show that the cos have the conspire and cancel out in such a way that it just cannot happen.

If we assume a lot, however, we can say some things about differentiability.

**Theorem 23** (Rudin 7.17). Let  $\{f_n\}$  be a sequence of functions  $f_n: [a, b] \rightarrow \mathbb{R}$ . Suppose that

1. each  $f_n$  is differentiable on  $[a, b]$
2.  $\exists x_0 \in [a, b]$  such that  $\{f_n(x_0)\}$  converges
3.  $f'_n$  converge uniformly on  $[a, b]$

Then  $\exists f$  such that  $f_n \rightarrow f$  uniformly on  $[a, b]$ , and  $f'(x)$  exists for all  $x \in [a, b]$  and  $f'(x) = \lim_{n \rightarrow \infty} f'_n(x)$  (i.e.  $f'_n \rightarrow f'$  uniformly).

We need the last hypothesis, because our example above fails it. The second is also important, as  $f_n = n$  is differentiable and has derivatives, but it does not converge.

Won't finish proof right now, but we'll get started with an important estimate. Step 1:  $f_n \rightarrow f$  uniformly. Let  $\varepsilon > 0$ , let  $N$  large enough that  $|f_n(x_0) - f_m(x_0)| < \varepsilon$  for all  $n, m \geq N$  (convergent sequences are Cauchy) and  $|f'_n(x) - f'_m(x)| < \varepsilon$  for all  $n, m \geq N$ , for all  $x \in [a, b]$ . Here is the crucial idea of this proof: apply the Mean Value Theorem to the difference  $(f_n - f_m)$ . For  $x, t \in [a, b], x \neq t, \exists c$  between  $x$  and  $t$  such that

$$|(f_n(x) - f_m(x)) - (f_n(t) - f_m(t))| = |(f'_n - f'_m)(c)||x - t| \leq \varepsilon|x - t|$$

How do we use this inequality? We have two consequences:

$$(1) |((f_n(x) - f_m(x)) - (f_n(t) - f_m(t)))| < \varepsilon|b - a| \text{ (useful for uniform convergence)}$$

$$(2) \frac{|(f_n(x) - f_m(x)) - (f_n(t) - f_m(t))|}{|x - t|} < \varepsilon \text{ (useful for differentiability)}$$

(I don't know if this one is right, some power tripping Math 101 TA kicked us out.)

## 14 February 16

Recall what we had last time: in the process of proving theorem 7.17.

*Proof.* What did we do to prove it last time? Fixed  $\varepsilon > 0$ , found  $N$  such that  $\forall n, m \geq N$ ,

$$|f_n(x_0) - f_m(x_0)| < \varepsilon \quad \text{and} \quad |f'_n(x) - f'_m(x)| < \varepsilon, \forall x$$

We used MVT  $x, t \in [a, b], x \neq t$  and got two consequences

$$(1) |((f_n(x) - f_m(x)) - (f_n(t) - f_m(t)))| \leq \varepsilon|b - a| \text{ (useful for uniform convergence)}$$

$$(2) \frac{|(f_n(x) - f_m(x)) - (f_n(t) - f_m(t))|}{|x - t|} \leq \varepsilon \text{ (useful for differentiability)}$$

We are splitting our step into several proofs. We first show that there is some  $f$  that  $f_n$  converges uniformly to, and then prove the statement for the derivative.

We will try to rearrange (1) to get the Cauchy criterion. By (1) for  $t = x_0$ , for  $n, m \geq N, x \in [a, b]$ , we get

$$|f_n(x) - f_m(x)| \leq |(f_n(x) - f_m(x)) - (f_n(x_0) - f_m(x_0))| + |f_n(x_0) - f_m(x_0)| \leq \varepsilon(b - a) + \varepsilon$$

Since  $\varepsilon$  can be arbitrarily small ( $\varepsilon'((b - a) + 1) = \varepsilon$ ),  $\{f_n\}$  satisfies the Cauchy Criterion for uniform convergence.

Now we prove that  $f' = \lim_{n \rightarrow \infty} f'_n$ . Fix  $x \in [a, b]$ . We need to show  $f$  is differentiable at  $x$  and  $f'(x) = \lim_{n \rightarrow \infty} f'_n(x)$ . Since the derivative involves limits, this will involve a careful interchange of limits. We write the derivative out explicitly:

$$\phi_n(t) = \frac{f_n(t) - f_n(x)}{t - x}, \quad \phi(t) = \frac{f(t) - f(x)}{t - x}$$

so  $f'_n(x) = \lim_{t \rightarrow x} \phi_n(t)$ , and similarly for  $f'(x)$ , if it exists. Inequality (2) says for  $n, m \geq N$ ,

$$|\phi_n(t) - \phi_m(t)| \leq \varepsilon$$

i.e. the functions  $\{\phi_n\}$  satisfies Cauchy criterion for uniform convergence, which implies  $\phi_n$  converges uniformly on the domain  $[a, b] \setminus \{x\}$  (but at least  $x$  is a limit point; important later). This says something about  $\phi_n$ , but we don't know if it converges to  $\phi(t)$  at any fixed  $t$ . Showing pointwise is enough. Fix  $t \in [a, b] \setminus \{x\}$ .

$$\phi_n(t) - \phi(t) = \left| \frac{(f_n(t) - f_n(x)) - (f(t) - f(x))}{t - x} \right| \leq \left| \frac{f_n(t) - f(t)}{t - x} \right| + \left| \frac{f_n(x) - f(x)}{t - x} \right|$$

which goes to 0 as  $n \rightarrow \infty$ . Conclusion,  $\phi_n \rightarrow \phi$  pointwise, hence uniformly on  $[a, b] \setminus \{x\}$ .

Theorem 7.11: properties being inherited by uniform limit, interchanging limits. Since  $x$  is a limit point of  $[a, b] \setminus \{x\}$ , and  $\phi_n \rightarrow \phi$  uniformly on  $[a, b] \setminus \{x\}$ , we apply theorem 7.11 to conclude

$$f'(x) = \lim_{t \rightarrow x} \phi(t) = \lim_{t \rightarrow x} \lim_{n \rightarrow \infty} \phi_n(t) = \lim_{n \rightarrow \infty} \lim_{t \rightarrow x} \phi_n(t) = \lim_{n \rightarrow \infty} f'_n(x)$$

if the limit exists, and it does. And this is a uniform limit???

□

## 14.1 The Weierstrass function

**Theorem 24** (Rudin 7.18). *There exists a continuous function  $f: \mathbb{R} \rightarrow \mathbb{R}$  such that  $f'(x)$  does not exist for any  $x \in \mathbb{R}$ .*

One way to do this is to construct such a function. Note that if we close our eyes and pick a function at random from  $C(\mathbb{R})$ , it will almost surely be a function that is not differentiable anywhere. But if we try to write any function down, it is usually differentiable. “It’s like looking for hay in a hay stack, but we never get the hay.” This happens often in math: can prove that almost always functions have a certain property, but can’t even write it out.

*Proof.* Consider the periodization of  $|x|$  on  $[-1, 1]$  to the whole real line, call this  $\phi(x)$ . This function is continuous, in fact, it is Lipschitz, with Lipschitz constant 1, or 1-Lipschitz:  $|\phi(x) - \phi(y)| \leq |x - y|$ . Note that being Lipschitz is not preserved under pointwise limits, but being  $k$ -Lipschitz, for a fixed  $k$ , is.

Let  $f(x) = \sum_{n=0}^{\infty} \left(\frac{3}{4}\right)^n \phi(4^n x)$ . This series converges absolutely by the Weierstrass  $M$ -test. Each of these terms are continuous, and since absolutely convergent series of continuous functions converges to a continuous function,  $f$  is continuous.

For  $n$  large, what does  $\left(\frac{3}{4}\right)^n \phi(4^n x)$  look like? Very small, but very spiky. But note that the  $4^n$  is getting large faster than  $\left(\frac{3}{4}\right)^n$  is getting small, i.e. if we multiply them, we get  $3^n$ , which still blows up. We get that  $\left(\frac{3}{4}\right)^n \phi(4^n x)$  is  $3^n$ -Lipschitz.

Fix  $x \in \mathbb{R}$ . When to show:  $f(x)$  is not differentiable at  $x$ . Suffices to find  $\delta_m \searrow 0$  such that

$$\left| \frac{f(x + \delta_m) - f(x)}{\delta_m} \right| \nearrow \infty$$

as  $m \rightarrow \infty$ . Trick is finding  $\delta_m$  such that bigger  $m$  cancel out (equal spots on the period), and for the smaller values of  $m$ , the Lipschitz constant is too small to make a difference: even if all the other peaks were working against it,  $3^j - \sum_{n=0}^{j-1} 3^j = \frac{1}{6}3^j$  or something like that.

Full proof in Rudin, or work through it yourself. □

Will talk about equicontinuity after the break. Administrative note: two homeworks, one walks you through the proof of DCT but not to be handed in, the other is the actual homework.

## 15 February 26

### 15.1 Equicontinuity

**Definition 13** (Equicontinuous).  $\mathcal{F}$  is a family of functions  $f: E \rightarrow \mathbb{C}$  where  $E$  is a set in a metric space  $X$ .  $\mathcal{F}$  is *equicontinuous* on  $E$  if for every  $\varepsilon > 0$ , there exists  $\delta > 0$  such that whenever  $x, y \in E$  where  $d(x, y) < \delta$  and  $f \in \mathcal{F}$ ,

$$\forall f \in \mathcal{F}, \quad |f(x) - f(y)| < \varepsilon$$

**Remark 4.** If  $\mathcal{F}$  is equicontinuous, each  $f \in \mathcal{F}$  is uniformly continuous)

**Remark 5.** Converse is false:  $X = [0, 1]$ ,  $\mathcal{F} = \{f_n\}_{n=1}^{\infty}$ , something something  $f_n = x^n$ ,  $f_n = nx$  ???

**Remark 6.** If  $\mathcal{F}$  is finite and each  $f \in \mathcal{F}$  is uniformly continuous, then  $\mathcal{F}$  is equicontinuous.

**Theorem 25** (Rudin 7.24). *Let  $K$  be a compact metric space,  $f_n: K \rightarrow \mathbb{C}$  continuous ( $f_n \in C(K)$ ) and  $\{f_n\}$  converges uniformly on  $K$ , then  $\{f_n\} = \mathcal{F}$  is equicontinuous on  $K$ .*

*Proof.* We will use the  $\varepsilon/3$  argument (just about splitting things up). Let  $\varepsilon > 0$ . Since  $\{f_n\}$  converges uniformly, there exists  $N$  such that  $\forall n, m \geq N$ ,  $x \in K$ ,  $|f_n(x) - f_m(x)| < \varepsilon/3$ . Since  $K$  is compact, each function is uniformly continuous. So  $\{f_1, \dots, f_N\}$  is equicontinuous, i.e.  $\exists \delta > 0$  such that  $\forall x, y \in K$  (with  $d(x, y) < \delta$ ) and all  $n \leq N$ ,  $|f_n(x) - f_n(y)| < \varepsilon/3$ . Next, if  $n > N$  and  $x, y \in K$  with  $d(x, y) < \delta$ ,  $|f_n(x) - f_n(y)|$  ff couldn’t see the bottom of the board, but the proof finishes as you would expect. □

**Theorem 26** (Rudin 7.16). *Let  $K$  be a compact metric space,  $\{f_n\}$  an equicontinuous family of functions  $f_n: K \rightarrow \mathbb{C}$ . If  $\{f_n\}_{n=1}^{\infty}$  converges pointwise on  $K$ , then  $\{f_n\}_{n=1}^{\infty}$  converges uniformly.*

Important in math, particularly in analysis, important to remember the proofs of the theorem. Will often come across a problem that hasn't a slightly different set up of theorem, and so modify proof slightly to get what you want. How do you remember hundreds of proof. You could remember every penstroke, but those form words, so easier to remember words; now we could remember every word, but those come together to form sentences, so easier to remember sentences; now could remember every sentence, but those come together to make arguments, so it is simpler to just remember the argument. In this case, we remember  $\varepsilon/3$  argument and try to expand out of that.

*Proof.*  $\varepsilon/3$  argument. Let  $\varepsilon > 0$ . Select  $\delta > 0$  such that  $\forall x, y \in K$  with  $d(x, y) < \delta$ , and for all  $f_n$ ,  $|f_n(x) - f_n(y)| < \varepsilon/3$ . Since  $K$  is compact, the open cover  $\{N_\delta(x)\}_{x \in K}$  has a finite subcover,  $N_\delta(x_1), \dots, N_\delta(x_l)$ . Thus, given  $x \in K$ , there exists  $x_j$  such that  $d(x, x_j) < \delta$  so for  $n, m \in N$ ,

$$|f_n(x) - f_m(x)| \leq \underbrace{|f_n(x) - f_n(x_j)|}_{< \varepsilon/3} + |f_n(x_j) - f_m(x_j)| + \underbrace{|f_m(x_j) - f_m(x)|}_{< \varepsilon/3}$$

So there is one more thing we have to bound by  $\varepsilon/3$ . Since  $\{f_n\}$  converges pointwise, for each  $j = 1, \dots, l$ ,  $\exists N_j$  such that  $\forall n, m \geq N_j$ ,  $|f_n(x_j) - f_m(x_j)| < \varepsilon/3$ . Let  $N = \max\{N_1, \dots, N_l\}$ . Then  $\forall n, m \geq N$ ,  $\forall j = 1, \dots, l$

$$|f_n(x_j) - f_m(x_j)| < \varepsilon/3$$

and the proof follows.  $\square$

We use compactness in a pretty essential way. In the previous theorem, we just used it for continuity thus uniform continuity. But here, we really needed it in the argument, so compactness is the natural environment for equicontinuity.

## 16 February 28

*Definition 14* (Rudin 7.19). Let  $X$  be a metric space,  $E \subset X$ , let  $\mathcal{F}$  be a family of functions  $f: E \rightarrow \mathbb{C}$ . We say  $\mathcal{F}$  is *pointwise bounded* if there exists  $\phi: E \rightarrow \mathbb{R}$  such that  $|f(x)| \leq \phi(x)$  for every  $x \in E$  and  $f \in \mathcal{F}$ . We say  $\mathcal{F}$  is *uniformly bounded* if there exists  $M \in \mathbb{R}$  such that  $|f(x)| \leq M$  for all  $x \in E$  and  $f \in \mathcal{F}$ .

How would we generalize this, i.e. for  $f: E \rightarrow Y$  where  $(Y, \rho)$  is any metric space. If not sure about pointwise, but perhaps for uniform, we say that  $f(E)$  is contained with a bounded set (above, our bounded set is centered at 0, but could be anywhere). But we will just stick with looking at functions to  $\mathbb{C}$ .

**Theorem 27** (Rudin 7.23). *Let  $X$  be a metric space,  $E \subset X$ ,  $E$  countable. Let  $f_n: E \rightarrow \mathbb{C}$ , and suppose  $\{f_n\}$  is pointwise bounded on  $E$ . Then there exists a subsequence that converges pointwise on  $E$ .*

Just as last class, proof is just one sentence: Cantor diagonalization argument. The idea is that we will get a subsequence that works for one element of  $E$ , then a subsubsequence that works for two elements of  $E$ , etc.

*Proof.* Write  $E = \{x_1, x_2, \dots\}$ . We know  $\{f_n(x_1)\}_{n=1}^\infty$  is a bounded sequence of complex numbers, and hence has a convergent subsequence  $\{f_{1,k}\}_{k=1}^\infty$  (note that this is a sequence of functions, NOT evaluated at  $x_1$ ). We will construct successive subsequences.

$$\begin{array}{ccccccc} f_{1,1} & f_{1,2} & f_{1,3} & f_{1,4} & \cdots \\ f_{2,1} & f_{2,2} & f_{2,3} & f_{2,4} & \cdots \\ f_{3,1} & f_{3,2} & f_{3,3} & f_{3,4} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{array}$$

where  $\{f_{i,k}\}_{k=1}^\infty$  is a subsequence of  $\{f_{i-1,k}\}_{k=1}^\infty$ , hence  $\{f_{i,k}\}_{k=1}^\infty$  converges. We take the diagonal sequence  $f_{i,i}$ . This converges at  $x_j$  for every  $j$ . Because  $\{f_{i,i}\}$  is a subsequence of  $\{f_{j,k}\}_{k=1}^\infty$ , except possibly for the first  $j-1$  elements.  $\square$

**Theorem 28** (Arzela-Ascoli (Rudin 7.25)). *Let  $K$  be a compact metric space, Let  $\{f_n\} \subset C(K)$  (i.e.  $f_n$  are continuous,  $f_n: K \rightarrow \mathbb{C}$ ) by equicontinuous and pointwise bounded. Then*

(a)  $\{f_n\}$  is uniformly bounded

(b)  $\{f_n\}$  has a uniformly convergent subsequence

Might recall an equivalent definition of compactness: every convergent sequence has a convergent subsequence. We can't really say this is compact (not quite a sequence in a metric space), but gives us some similar idea.

*Proof.* (a) We need to find  $M \in \mathbb{R}$  such that  $\forall x \in K, n \in \mathbb{N}, |f_n(x)| \leq M$ . Since  $\{f_n\}$  is equicontinuous, there exists  $\delta > 0$  ( $\varepsilon = 1$ ) such that  $\forall n \in \mathbb{N}, \forall x, y \in K$  with  $d(x, y) < \delta$ , we have  $|f_n(x) - f_n(y)| < 1$ . Since  $K$  is compact, the cover  $\{N_\delta(x)\}_{x \in K}$  has a finite subcover  $N_\delta(x_1), \dots, N_\delta(x_l)$ . For each  $i = 1, \dots, l$ ,  $\{f_n(x_i)\}_{n=1}^\infty$  is bounded by  $M_i$ . Let  $M = 1 + \max\{M_1, \dots, M_l\}$ . For any  $x \in K, n \in \mathbb{N}$ ,

$$|f_n(x)| \leq |f_n(x_i)| + |f_n(x) - f_n(x_i)| < M_i + 1 \leq M$$

where  $x_i$  is a point with  $x \in N_\delta(x_i)$ . Hence, we have uniform boundedness.

(b) Step 1: Let  $E$  be a countable, dense subset of  $K$ . (How do we know exists? Consider the cover of balls at each point with radius of the form  $1/n$ , take finite subcover for each  $1/n$ , and iterating over each  $n$  gives us a countable set of points. Easy to convince that it is dense: closure is equal to  $K$ .) By Theorem 7.23, there exists a subsequence of  $\{f_n\}$  that converges pointwise on  $E$ . Call this sequence  $\{g_i\}$ . We will show that  $\{g_i\}$  satisfies the Cauchy Criterion for convergence.

Step 2: Let  $\varepsilon > 0$ . By equicontinuity of  $\{g_i\}$ ,  $\exists \delta > 0$  such that  $\forall x, y \in K$  with  $d(x, y) < \delta$ ,  $\forall i \in \mathbb{N}$ ,  $|g_i(x) - g_i(y)| < \varepsilon/3$ . Since  $E \subset K$  is dense,  $\{N_\delta(y)\}_{y \in E}$  is a cover of  $K$ , thus there exists a finite subcover  $\{N_\delta(y_1), \dots, N_\delta(y_l)\}$ ,  $y_i \in E$ . For all  $x \in K$ ,  $\exists s$  such that  $d(x, y_s) < \delta$ .

Step 3: Now we just complete the proof ( $\varepsilon/3$  argument). For  $x \in K, i, j \in \mathbb{N}$ ,

$$|g_i(x) - g_j(y)| \leq \underbrace{|g_i(x) - g_i(y_s)|}_{< \varepsilon/3} + \underbrace{|g_i(y_s) - g_j(y_s)|}_{\text{converges}} + \underbrace{|g_j(y_s) - g_j(x)|}_{< \varepsilon/3}$$

If we choose  $N \in \mathbb{N}$  sufficiently large,  $\forall i, j \geq N$ , we have  $|g_i(y_s) - g_j(y_s)| < \varepsilon/3$  for all  $s \in 1, \dots, l$  (because there are only finite number of choices of  $s$ ).  $\square$

Used compactness in an essential way in part (a) and (b). Can be an exercise: find when  $K$  isn't compact, this fails.

## 17 March 1

(Notes via Sushrut.)

*Definition 15* (Convolution). Let  $f, g: \mathbb{R} \rightarrow \mathbb{R}$  or  $\mathbb{R} \rightarrow \mathbb{C}$  be Riemann integrable over all of  $\mathbb{R}$ , i.e. the following integrals exist:

$$\lim_{z \rightarrow -\infty} \int_{-z}^0 f(x) dx \quad \text{and} \quad \lim_{z \rightarrow \infty} \int_0^z f(x) dx$$

For  $x \in \mathbb{R}$  we define

$$f * g(x) = \int_{-\infty}^{\infty} f(t)g(x-t)dt$$

If the integral exists,  $f * g$  is a function whose domain is a subset of  $\mathbb{R}$ .

(For most examples that we care about, the domain is  $\mathbb{R}$ .) Exercise: If  $f * g(x)$  exists, then  $g * f(x)$  exists and  $f * g(x) = g * f(x)$ .

*Definition 16* (Approximate identity). We say that a sequence of functions  $\{f_n\}$ ,  $f_n: \mathbb{R} \rightarrow \mathbb{C}$  (or  $\mathbb{R} \rightarrow \mathbb{R}$ ) is called an *approximate identity* if they satisfy the following:

1.  $\int_{-\infty}^{\infty} f_n(t)dt = 1$  for all  $n$
2. There exists  $M \geq 0$  such that

$$\int_{-\infty}^{\infty} |f_n(t)|dt \leq M$$

for all  $n \in \mathbb{N}$ . (Note: this part is superfluous if  $f_n(t) \geq 0$  for all  $t \in \mathbb{R}$ , for all  $n \in \mathbb{N}$ .)

3. For all  $\delta > 0$ ,

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{-\delta} |f_n(t)| dt = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} \int_{\delta}^{\infty} |f_n(t)| dt = 0$$

Example: Consider the function  $f_n(t) = n f(nt)$ , such that

$$\int_{-\infty}^{\infty} f_n(t) dt = 1$$

(I feel like something is missing). The limit of this sequence obeys

$$\int_{-\infty}^{\infty} f(t) dt = \int_{-1}^1 f(t) dt$$

This sequence is slowly approaching the Dirac- $\delta$  function. We now build on the example further.

Example: Consider the function

$$g(x) = \begin{cases} 1, & x \in [-1, 1] \\ 0 & \text{otherwise} \end{cases}$$

We will now try to sketch the graphs of  $g * f_1(x)$  and  $g * f_5(x)$  and try to make a guesstimate of how the graph changes as we keep convolving with  $f_n$  for larger  $n$ .

Note that if we look at  $g * f_1(100) = \int_{-\infty}^{\infty} g(t) f_1(100 - t) dt$ , we get this evaluates to zero, since  $g(t)$  is zero for all  $t$  outside  $I := [-1, 1]$  and  $f_1(100 - t) = 0$  for all  $t \in [-1, 1]$  since  $f_1$  has a width of 1 where it is non-zero, and it is centered at 100. Thus,  $g * f_1(100) = 0$ . So if we observe carefully, since  $g$  is non-zero only in  $I$ , we get that the only points where  $g * f_1(x) \neq 0$  is on the interval  $[-2, 2]$ . The graph will be wider than that of  $f_1$ , and the descent to zero begins immediately after you move off the origin on either side. Similarly, we can repeat this for higher  $n$ , and what we see is that the curve gets flatter on top because it equals 1 for all  $x$  up to  $1 - \frac{1}{n}$ , and then has a very fast descent to zero on the interval  $[1 - \frac{1}{n}, 1 + \frac{1}{n}]$ .

## 18 March 4

Recall the definition of an approximate identity:

*Definition 17.* Let  $\{f_n\}$  be a sequence of functions such that  $\int_{-\infty}^{\infty} f_n \leq M$  for all  $n$ , and  $\int_{-\delta}^{\delta} |f_n| \rightarrow 0$  and  $\int_{\delta}^{\infty} |f_n| \rightarrow 0$  (all of the mass of  $f_n$  is shrinking to the origin.)

Recall that  $f_n$  approaching dirac delta. If  $g$  is the square wave, then  $f_n * g$  approaches being a square wave as well???

Ex 2:  $f_n = \begin{cases} c_n(1 - x^2)^n, & -1 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$ . Choose  $c_n$  such that  $\int_{-\infty}^{\infty} f_n(t) dt = \int_{-1}^1 f_n(t) dt = 1$ . This a lot like

our previous example, but actually not infinitely differentiable. Claim:  $\{f_n\}$  is an approximate identity.

1.  $\int f_n = 1$ , true by definition
2.  $\int |f_n| = \int f_n = 1$  since  $f_n$  nonnegative
3.  $\int_{-\delta}^{\delta} f_n$  and  $\int_{\delta}^{\infty} |f_n| \rightarrow 0$ . This is just a calculation. Done in Rudin, gets a more accurate estimate, but it doesn't matter, we're just going to zero. Note  $c_n$  is probably growing with  $n$  since  $(1 - x^2) < 1$  so  $(1 - x^2)^n$  is shrinking with  $n$ . It would be difficult to compute  $c_n$ , but it is enough to approximate it. Note that  $f_n(x) \geq \frac{1}{2}$  on  $\left[\frac{-1}{2\sqrt{n}}, \frac{1}{2\sqrt{n}}\right]$ . So

$$\int_{-\infty}^{\infty} (1 - x^2)^n dx = \int_{-1}^1 (1 - x^2)^n dx \geq \int_{\frac{-1}{2\sqrt{n}}}^{\frac{1}{2\sqrt{n}}} (1 - x^2)^4 dx \geq \int_{\frac{-1}{2\sqrt{n}}}^{\frac{1}{2\sqrt{n}}} (1 - x^2)^4 dx \geq \frac{1}{2\sqrt{n}}$$

Then  $c_n \leq 2\sqrt{n}$  ( $100n^{100}$  would have also been fine). For  $\delta > 0$ , then on  $(-\infty, -\delta)$  and  $(\delta, \infty)$ :

$$c_n(1 - x^2)^n \leq 2\sqrt{n}(1 - \delta^2)^n \rightarrow 0$$

(the rate it goes to 0 does depend on  $\delta$ , but we fix it).<sup>4</sup> Thus  $\int_{-\infty}^{\delta} |f_n(x)| dx = \int_{-1}^{-\delta} f_n(x) dx \leq \int_{-1}^{-\delta} 2\sqrt{n}(1 - \delta^2)^n dx \leq 2\sqrt{n}(1 - \delta^2)^n \rightarrow 0$  as  $n \rightarrow \infty$ . Same for  $\int_{\delta}^{\infty} |f_n(x)| dx$ .

Good technique for approximating integrals. Restrict to a region that we understand well, and where the integral is big.

Why is this an important example? It is a polynomial (at least in some region), and we'll use it to show how to approximate functions with polynomials.

**Theorem 29 (C).** *Let  $\{f_n\}$  be an approximate identity. Let  $g: \mathbb{R} \rightarrow \mathbb{C}$  be bounded and uniformly continuous. Then  $f_n * g \rightarrow g$  uniformly on  $\mathbb{R}$ .*

Can probably prove this ourself: very similar flavour to the homework problem we did.

*Proof.* Select  $M_1$  such that  $\int_{-\infty}^{\infty} |f_n| \leq M_1$  for all  $n$  and  $M_2$  such that  $|g(x)| \leq M_2$  for all  $x \in \mathbb{R}$  (which we can do, since  $\{f_n\}$  is an approximate identity and  $g$  is bounded). Let  $\varepsilon > 0$ .  $\exists \delta > 0$  such that  $\forall x, y \in \mathbb{R}$  with  $|x - y| < \delta$ ,  $|g(x) - g(y)| < \varepsilon$ . For  $x \in \mathbb{R}$ , we have

$$|f_n * g(x) - g(x)| = \left| \int_{-\infty}^{\infty} f_n(t)g(x-t)dt - \int_{-\infty}^{\infty} f_n(t)g(x)dt \right|$$

(Integrals exist, since second one works because  $g(x)$  is fixed, and the first one converges since  $\int f_n$  is absolutely convergent, and so we can multiply it by a bounded function to and it is still absolutely convergent, which isn't true in general:  $\int_{-\infty}^{\infty} |f_n(t)g(x-t)|dt \leq \int_{-\infty}^{\infty} M_2|f_n(t)|dt \leq M_1M_2$ .) This is equivalent to

$$\left| \int_{-\infty}^{\infty} f_n(t)(g(x-t) - g(x))dt \right|$$

(Theorem 6.12 for  $\int_a^b h + \int_a^b c = \int_a^b (h+c)$  is for integrals over an interval, but also true for convergent integrals:  $\int_{-\infty}^{\infty} h + \int_{-\infty}^{\infty} i = \int_{-\infty}^{\infty} h+i$ ... probably already did on homework, so exercise.) Also

$$\leq \int_{-\infty}^{\infty} |f_n(t)(g(x-t) - g(x))|dt$$

(Theorem 6.13... in indefinite integral case, if not absolutely integrable its infinite, but inequality is still true anyway.) Finally, 6.12 gives

$$= \int_{-\infty}^{-\delta} |f_n(t)(g(x-t) - g(x))|dt + \int_{-\delta}^{\delta} |f_n(t)(g(x-t) - g(x))|dt + \int_{\delta}^{\infty} |f_n(t)(g(x-t) - g(x))|dt$$

(I think converges by definition of indefinite integral converging???) We then have

$$\leq 2M_2 \int_{-\infty}^{-\delta} |f_n(t)|dt + \varepsilon \int_{-\delta}^{\delta} |f_n(t)| + 2M_2 \int_{\delta}^{\infty} |f_n(t)|dt \leq 2M_2 \left( \int_{-\infty}^{-\delta} |f_n(t)|dt + \int_{\delta}^{\infty} |f_n(t)|dt \right) + \varepsilon \int_{-\infty}^{\infty} |f_n(t)|dt$$

(expanding domain of integral because nonnegative). Select  $N$  large such that  $\forall n \geq N$ ,  $\int_{-\infty}^{-\delta} |f_n(t)|dt < \varepsilon$  and  $\int_{\delta}^{\infty} |f_n(t)|dt < \varepsilon$  (since it's an approximate identity).

So for  $n \geq N$ ,  $x \in \mathbb{R}$  we have

$$|f_n * g(x) - g(x)| < \varepsilon(4M_2 + M_1)$$

and  $M_1, M_2$  fixed, so this proves the convergence we desired.  $\square$

He doesn't know the full history of approximate identities, but pretty sure that it wasn't this axiomatic way. There were a bunch of specific examples, families of functions, that they would use to prove things, and then realized that all of these were similar. One such example of proof we are going to do next time: Weierstrass approximation theorem. Given a continuous function compactly supported, then there is a sequence of polynomials that converge to it uniformly. We're gonna have a sequence of polynomials, and convolve it with the approximate identity from example 2.



## 19 March 6

**Definition 18** (Compact support).  $(X, d)$  a metric space (usually  $X = \mathbb{R}$ ),  $f: X \rightarrow \mathbb{C}$  or  $X \rightarrow \mathbb{R}$ . We say  $f$  has *compact support* if there is a compact set  $K$  such that  $f(x) = 0$  for all  $x \in X \setminus K$  (i.e. for all  $x \notin K$ ).

If a metric space is compact, then all of the functions defined on it have compact support, since just take the whole space. But that is usually not the case, since we normally work with  $\mathbb{R}$ . Heine-Borel tells us: If  $X = \mathbb{R}$ , then  $f: \mathbb{R} \rightarrow \mathbb{C}$  has compact support iff  $\exists R$  such that  $f(x) = 0$  for all  $x \notin [-R, R]$ .

Why do we like functions with compact support? See that if  $f, g$  have compact support, specifically in  $[-R, R]$ , then  $f * g(x) = \int_{-\infty}^{\infty} f(t)g(x-t)dt$  has compact support in  $[-2R, 2R]$ .

**Lemma 1** (D). Let  $f: \mathbb{R} \rightarrow \mathbb{C}$  or  $\mathbb{R} \rightarrow \mathbb{R}$  be compactly supported and integrable. Let  $Q: \mathbb{R} \rightarrow \mathbb{C}$  be a polynomial. Then  $f * Q$  is a polynomial. If  $f: \mathbb{R} \rightarrow \mathbb{R}$  and  $Q$  has real coefficients, then  $f * Q$  also has real coefficients.

*Proof.* What does it mean for  $Q$  to be a polynomial? Write  $Q(x) = \sum_{k=0}^n a_k x^k$  ( $x^0 = 1$ ). Formally (we haven't discussed convergence) we have  $f * Q(x) = \int_{-\infty}^{\infty} f(t)Q(x-t)dt$ , but we do in fact have equality, because  $f$  is compactly supported, so if we want to get a polynomial of  $x$  with coefficients in  $t$ , we can try to separate our  $x$  and  $t$ :

$$\begin{aligned} f * Q(x) &= \int_{-R}^R f(t)Q(x-t)dt \\ &= \int_{-R}^R f(t) \sum_{k=0}^n a_k (x-t)^k dt \\ &= \int_{-R}^R f(t) \sum_{k=0}^n a_k \sum_{l=0}^k \binom{k}{l} (-t)^{k-l} x^l dt \\ &= \int_{-R}^R \sum_{k=0}^n \sum_{l=0}^k \left( f(t) a_k \binom{k}{l} (-t)^{k-l} x^l \right) dt \\ &= \sum_{k=0}^n \sum_{l=0}^k \left( \int_{-R}^R f(t) a_k \binom{k}{l} (-t)^{k-l} dt \right) x^l \end{aligned}$$

(Theorem 6.12 in the last line.) And the coefficients are in  $\mathbb{C}$ , or are real iff  $\mathbb{R} \rightarrow \mathbb{R}$  and all  $a_k \in \mathbb{R}$ . □

### 19.1 Weierstrass Approximation Theorem

We have now built up the tools to prove the following theorem:

**Theorem 30** (Weierstrass approximation theorem (Rudin 7.26)). Let  $f: [a, b] \rightarrow \mathbb{C}$  or  $f: [a, b] \rightarrow \mathbb{R}$  be continuous. Then there exists a sequence of polynomials  $\{P_n\}$  so that  $P_n \rightarrow f$  uniformly on  $[a, b]$ . If  $f: [a, b] \rightarrow \mathbb{R}$ , then  $\{P_n\}$  can be chosen to have real coefficients.

The proof for this just uses the following three facts: (1) convolving with an approximate identity gives us a piecewise polynomial, (2) when you convolve a bounded function with a uniformly convergent function, you uniformly converge to the function that you started with, (3) Convolving a compactly supported function with a polynomial is also a polynomial. What stops us from just citing these facts, what is the difficulty? Take a function continuous on  $[a, b]$ , and extend it to the reals by making it 0 everywhere, not quite continuous. Also, when we convolve with our approximate identity, not quite a polynomial, but we will just have to go about this carefully. We will address these in the proof.

*Proof.* Step 1: We want to reduce Theorem 7.26 to the special case where  $[a, b] = [0, 1]$  and  $f(0) = f(1) = 0$ . So suppose the theorem is true for such functions, and let  $g: [a, b] \rightarrow \mathbb{C}$  be continuous. We'll make  $g$  have endpoints at 0, 1, and then subtract the linear function that passes through the end points. Let  $f_1(x) = g(x(b-a) + a)$ . Then  $f_2(x) = f_1(x) - (f_1(0)(1-x) + f_1(1)x)$ . Hence, if  $Q_n \rightarrow f_2$  uniformly, then let  $P_n(x) = Q_n\left(\frac{x-a}{b-a}\right) + f_1(0)\left(1 - \frac{x-a}{b-a}\right) + f_1(1)\left(\frac{x-a}{b-a}\right)$  (do we not need to cite that composition of uniformly convergent is uniformly

convergent??). Particularly, this is a polynomial (composing a polynomial with a polynomial is also a polynomial). Conclusion: it suffices to prove Theorem 7.26 for  $f: [0, 1] \rightarrow \mathbb{C}$  with  $f(0) = f(1) = 0$ . We will extend  $f: \mathbb{R} \rightarrow \mathbb{C}$  by setting  $f(x) = 0$  for  $x \notin [0, 1]$ . This  $f$  is uniformly continuous and bounded, so by Theorem C, if  $\{\tilde{Q}\}_{n \in \mathbb{N}}$  is an approximate identity, then  $\tilde{Q}_n * f \rightarrow f$  uniformly. We will use  $\tilde{Q}_n$  from example 2

$$\tilde{Q}_n(x) = \begin{cases} C_n(1 - x^2)^n, & -1 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Last Step: Prove  $\tilde{Q}_n * f = Q_n * f$ . Will pick this up next lecture. □

## 20 March 8

(Notes via Sushrut.)

**Theorem 31 ((B)).** For a polynomial function  $g_n$ , where  $\{q_n\}$  is an approximate identity, and  $f: [a, b] \rightarrow \mathbb{C}$  (or  $\mathbb{R}$ ) be a continuous function. Then  $q_n * f(x)$  is a polynomial function for each  $n$ .

*Proof.* Step 3: We claim that  $\tilde{q}_n * f(x) = q_n * f(x)$  for all  $x \in [0, 1]$ .

Let  $x \in [0, 1]$ , hence

$$\begin{aligned} \tilde{q}_n * f(x) &= f * \tilde{q}_n(x) = \int_{-\infty}^{\infty} f(t) \tilde{q}_n(x - t) dt \\ &= \int_0^1 f(t) \tilde{q}_n(x - t) dt, && \text{here } x - t \in [-1, 1] \\ &= \int_0^1 f(t) q_n(x - t) dt \\ &= \int_{-\infty}^{\infty} f(t) q_n(x - t) dt = f * q_n(x) = q_n * f(x) \end{aligned}$$

□

### 20.1 Stone's Generalization of the Weierstrass approximation theorem

*Definition 19 (Algebra).* Let  $\mathcal{A}$  be a set of functions  $f: E \rightarrow \mathbb{C}$  (or  $E \rightarrow \mathbb{R}$ ). We say  $\mathcal{A}$  is a (complex) *algebra* if for all  $f, g \in \mathcal{A}$ , for all  $c \in \mathbb{C}$ :

1.  $f + g \in \mathcal{A}$
2.  $f \cdot g \in \mathcal{A}$
3.  $cf \in \mathcal{A}$

Examples:

- Polynomial functions  $f: \mathbb{R} \rightarrow \mathbb{C}$
- $\mathcal{C}(\mathbb{R})$  the bounded continuous function  $f: \mathbb{R} \rightarrow \mathbb{C}$
- Trigonometric polynomial functions, which are polynomials of the form

$$p(x) = \sum_{k=0}^n (a_k \sin(kx) + b_k \cos(kx))$$

- Symmetric polynomial functions
- Piecewise polynomial functions

- Functions of the form

$$f(x) = \sum_{k=0}^n c_k e^{2\pi i k x}$$

- Functions of the form

$$f(x) = \sum_{k=-n}^n c_k e^{2\pi i k x}$$

- holomorphic functions over  $\mathbb{C}$  (or over simply connected subsets of  $\mathbb{C}$ )

**Definition 20** (Uniformly closed). We say  $\mathcal{A}$  is *uniformly closed* if for all uniformly convergent sequences  $\{f_n\} \subseteq \mathcal{A}$ , we have  $\lim f_n \in \mathcal{A}$ .

**Definition 21** (Uniform closure). Let  $\mathcal{A}$  be an algebra, and

$$\mathcal{B} := \{f: E \rightarrow \mathbb{C} : \text{there exist } \{f_n\} \subseteq \mathcal{A} \text{ such that } f_n \rightarrow f \text{ uniformly}\}$$

(the set of limit points of uniformly convergent sequences in  $\mathcal{A}$ ).  $\mathcal{B}$  is called the uniform closure of  $\mathcal{A}$ , which we will denote by  $\text{Cl}_u(\mathcal{A})$ .

It is natural that when an algebra is uniformly closed, it equals its uniform closure.

**Remark 7** (Consistency between definitions of uniform closure and closure). If  $\mathcal{A}$  is an algebra of bounded functions, then it has the metric  $\|\cdot\|_\infty$  (supremum norm), so  $(\mathcal{A}, d)$  is a metric space; it is subset of the metric space  $(\mathcal{X}, d)$ , where  $\mathcal{X}$  is the set of all bounded functions  $f: E \rightarrow \mathbb{C}$ .

The uniform closure of  $\mathcal{A}$  is the closure of  $\mathcal{A}$  in the metric space  $\mathcal{X}$ .

## 21 March 11

**Theorem 32** (Rudin 7.29). Let  $\mathcal{A}$  be an algebra of bounded functions. Then the uniform closure  $\overline{\mathcal{A}}$  is a uniformly closed algebra.

*Proof.* Recall  $(X, d)$  a metric space,  $\overline{\mathcal{A}}$  is the closure of  $\mathcal{A}$  in  $(X, d)$ , hence  $\overline{\mathcal{A}}$  is closed ( $X$  are bounded functions  $E \rightarrow \mathbb{R}$  or  $E \rightarrow \mathbb{C}$ ). We are not done: it is uniformly closed because in uniform metric. But it is not clear that it is an algebra, but it is straight forward: Let  $f, g \in \overline{\mathcal{A}}$ ,  $c \in \mathbb{R}$  or  $c \in \mathbb{C}$ . Let  $\{f_n\} \subset \mathcal{A}$ ,  $f_n \rightarrow f$  uniformly and  $\{g_n\} \subset \mathcal{A}$ ,  $g_n \rightarrow g$  uniformly. Show that  $f_n + g_n \rightarrow f + g$ ,  $f_n \cdot g_n \rightarrow f \cdot g$ , and  $cf_n \rightarrow cf$ , all uniformly (exercise).  $\square$

Idea for Stone-Weierstrass (extend to complex later): Let  $K$  be a compact set,  $\mathcal{A}$  a (real) algebra of continuous functions,  $f: K \rightarrow \mathbb{R}$ . Then:  $\overline{\mathcal{A}}$  is the algebra of *all* continuous functions  $f: K \rightarrow \mathbb{R}$ . (This is what Weierstrass approximation theorem says). Unless there an obvious obstruction, which we'll define below.

**Definition 22** (Rudin 7.30A). Let  $E$  be a set,  $\mathcal{A}$  a set of functions  $f: E \rightarrow \mathbb{R}$  or  $E \rightarrow \mathbb{C}$ . We say that  $\mathcal{A}$  *separates points* if for all  $x, y \in E$ , there exists  $f \in \mathcal{A}$  such that  $f(x) \neq f(y)$ .

Ex:  $E = [-1, 1]$  or  $E = \mathbb{R}$ .  $\mathcal{A}$  is polynomials works, but  $\mathcal{A}$  is even polynomials doesn't work, since  $f(x) = f(-x)$ . Failure to separate points is an obstruction to Stone-Weierstrass.

**Definition 23** (Rudin 7.30B). Let  $E$  and  $\mathcal{A}$  as before. We say that  $\mathcal{A}$  *vanishes at no point of  $E$*  if for each  $x \in E$ , there is a function  $f \in \mathcal{A}$  such that  $f(x) \neq 0$ .

Ex.  $E = [-1, 1]$  or  $E = \mathbb{R}$ .  $\mathcal{A}$  is polynomials works. Odd polynomials do not work, since  $f(-x) = -f(x) \implies f(0) = 0$ . (What is the smallest algebra that contains all odd polynomials? It is the ideal generated by  $x$ . What is the uniform closure of this? All continuous functions vanishing at 0, per SW actually). Vanishing at a point of  $E$  is an obstruction to Stone-Weierstrass.

**Theorem 33** (Real Stone-Weierstrass (Rudin 7.32)). Let  $K$  be a compact metric space,  $\mathcal{A}$  an algebra of continuous functions  $f: K \rightarrow \mathbb{R}$ . Suppose that  $\mathcal{A}$  separates points and vanishes at no point of  $K$ . Then the uniform closure  $\overline{\mathcal{A}}$  is the set/algebra of all continuous functions  $f: K \rightarrow \mathbb{R}$ .

This generalizes the real version of the Weierstrass approximation theorem. We had some closed interval, that's  $K$ . Can show that the set of polynomials form an algebra, and can check it separates and no vanishes.

Doing math research: write down something hope to be true. Then guess what obstructions are, what they would look like. Both genuine obstructions, or things that look like they could be obstructions. There are whole sequences of papers trying to find the obstructions of a theorem we would like to prove. And then this actually gives some ideas to the proof that we are going for.

Why  $\mathbb{R}$ , and not just any metric space? Well, crucially, we need the ring structure of  $\mathbb{R}$ , so that we can add and multiply functions in  $\mathbb{R}$  for our algebras. There are ways to generalize to more arbitrary spaces, but not a good introduction. We will go to  $\mathbb{C}$ , this will require considering another obstruction, but we will prove the real version first.

The proof has 3 main steps.

**Lemma 2.**  $K$  and  $\mathcal{A}$  as above. If  $f \in \mathcal{A}$ , then  $|f| \in \overline{\mathcal{A}}$ .

*Proof.* Let  $f \in \mathcal{A}$ , let  $M = \sup_{x \in K} |f(x)|$ ,  $M < \infty$ . Let  $\varepsilon > 0$ . Goal: find  $g \in \mathcal{A}$  with  $\sup_{x \in K} ||f(x)| - g(x)| < \varepsilon$ , then  $|f| \in \overline{\mathcal{A}}$ . By the Weierstrass approximation theorem (Theorem 7.26), or just an explicit computation, there exists a polynomial  $Q$  with real coefficients such that  $|Q(y) - |y|| < \varepsilon/2$  for all  $y \in [-M, M]$ . Let  $P(y) = Q(y) - Q(0) \implies P(0) = 0$ .

$$\begin{aligned} |P(y) - |y|| &= |Q(y) - Q(0) - |y| - |0|| \\ &\leq |Q(y) - |y|| + |Q(0) - |0|| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \end{aligned}$$

for all  $y \in [-M, M]$ . Write  $P(x) = \sum_{k=1}^n c_k x^k$  ( $c_0 = 0$ ).  $\mathcal{A}$  is an algebra, so

$$P(f) := \underbrace{\sum_{k=1}^n c_k f^k}_g \in \mathcal{A}$$

for  $x \in K$ ,  $|g(x) - |f(x)|| = |P(f)(x) - |f(x)|| = |P(f(x)) - |f(x)|| < \varepsilon$ . □

## 22 March 13

Last class:  $f \in \overline{\mathcal{A}} \implies |f| \in \overline{\mathcal{A}}$ . Stone-Weierstrass continued:  $K$  compact,  $\mathcal{A}$  an algebra of continuous functions  $f: K \rightarrow \mathbb{R}$  with separate points and vanishes at no points. Goal:  $\overline{\mathcal{A}}$  is the set of all continuous functions  $f: K \rightarrow \mathbb{R}$ .

**Lemma 3.**  $K, \mathcal{A}$  as before. Let  $f_1, \dots, f_n \in \overline{\mathcal{A}}$ , then  $\max\{f_1, \dots, f_n\} \in \overline{\mathcal{A}}$  and  $\min\{f_1, \dots, f_n\} \in \overline{\mathcal{A}}$ .

*Proof.* Induct on  $n$ .  $n = 1$  trivial. Now let  $f_1, \dots, f_n \in \overline{\mathcal{A}}$ , then  $\max\{f_1, \dots, f_n\} = \max\{\max\{f_1, \dots, f_{n-1}\}, f_n\}$ . Let  $g = \max\{f_1, \dots, f_{n-1}\}$ . It suffices to show then that if  $f, g \in \overline{\mathcal{A}}$ , then  $\max\{f, g\} \in \overline{\mathcal{A}}$ .

Note that  $\max\{x, y\} = \frac{x+y}{2} + \frac{|x-y|}{2}$ . Hence  $\max\{f, g\} = \frac{1}{2}(f+g) + \frac{1}{2}|f-g| \in \overline{\mathcal{A}}$ . Similarly,  $\min\{f, g\} = \frac{1}{2}(f+g) - \frac{1}{2}|f-g| \in \overline{\mathcal{A}}$ . □

We have not used the fact that  $\mathcal{A}$  separates points and vanishes nowhere. We use it in the proof of the following.

**Lemma 4** (Rudin Theorem 7.31).  $K, \mathcal{A}$  as before. Let  $x, y \in K$ ,  $x \neq y$ ,  $c, d \in \mathbb{R}$ , then  $\exists f \in \mathcal{A}$  such that  $f(x) = c$  and  $f(y) = d$ . (If  $x = y$ , true if  $c = d$ .)

*Proof.* Since  $\mathcal{A}$  separates points and vanishes at no point,  $\exists g, h, k \in \mathcal{A}$  such that  $g(x) \neq g(y)$ ,  $h(x) \neq 0$ ,  $k(y) \neq 0$ . Let  $u(z) = g(z)k(z) - \underbrace{g(x)k(z)}_{\in \mathbb{R}} = (g(z) - g(x))k(z) \in \mathcal{A}$  and  $v(z) = g(z)h(z) - g(y)h(y) = (g(z) - g(y))h(z) \in \mathcal{A}$ .

Then  $u(x) = 0$ ,  $u(y) \neq 0$ ,  $v(x) \neq 0$ ,  $v(y) = 0$ . Let

$$f(z) = \underbrace{\frac{c}{v(x)}}_{\in \mathbb{R}} v(z) + \frac{d}{u(y)} u(z) \in \mathcal{A}$$

Then  $f(x) = c$ ,  $f(y) = d$ . (If  $x = y$  and  $c = d$ , exercise.) □

So we have used the fact  $\mathcal{A}$  is an algebra (constantly), we have used that it separates points and vanishes nowhere, and used the fact that the functions were continuous to use the Weierstrass approximation theorem. We have yet to use the fact that  $K$  is compact.

**Lemma 5.**  $K, \mathcal{A}$  as before. Let  $f: K \rightarrow \mathbb{R}$  continuous,  $x \in K$ ,  $\varepsilon > 0$ . Then there exists  $g \in \overline{\mathcal{A}}$  such that

- $g(x) = f(x)$
- $g(t) - f(t) \geq -\varepsilon \forall t \in K$

*Proof.* For each  $y \in K$ , use Lemma 4 to find  $g_y \in \overline{\mathcal{A}}$  such that  $g_y(x) = f(x)$ ,  $g_y(y) = f(y)$ . Note  $g_y \in \mathcal{A} \implies g_y$  is continuous, and so  $g_y - f$  is continuous. Also,  $(g_y - f)(x) = 0$  and  $(g_y - f)(y) = 0$ . Hence, by continuity, there exists an open set  $U_y \in K$  such that  $g_y(t) - f(t) > -\varepsilon \forall t \in U_y$ .

$U_y$  contains  $y$ . If we let  $y$  range over all of  $K$ , we get a cover for  $K$ .  $K$  is compact, so we can get a finite subcover  $U_{y_1}, \dots, U_{y_n}$ . Let  $g = \max\{g_{y_1}, \dots, g_{y_n}\} \in \overline{\mathcal{A}}$  by Lemma 3. Let's verify

$$g(x) = \max\{g_{y_1}(x), \dots, g_{y_n}(x)\} = \max\{f(x), \dots, f(x)\} = f(x)$$

And for  $t \in K$ ,  $t$  is contained in some  $U_{y_j}$ ,  $1 \leq j \leq n$ , hence  $g(t) \geq g_{y_j}(t)$ , so  $g(t) - f(t) \geq g_{y_j}(t) - f(t) > -\varepsilon$ .  $\square$

Compactness was critical. Lemma 3 needed a finite number of functions to work, and we could only get this with compactness. We now mirror the proof to get a bound from above.

**Lemma 6.**  $K, \mathcal{A}$  as before. Let  $f: K \rightarrow \mathbb{R}$  continuous. Then  $\exists h \in \overline{\mathcal{A}}$  such that  $|h(t) - f(t)| \leq \varepsilon \forall t \in K$ .

*Proof.* For  $x \in K$ , let  $g_x \in \overline{\mathcal{A}}$  be as in Lemma 5 (same  $\varepsilon$ ).  $g_x - f$  is continuous and  $(g_x - f)(x) = 0$ . Hence,  $\exists$  an open set  $U_x$  containing  $x$  such that  $g_x(t) - f(t) < \varepsilon \forall t \in U_x$ .

$K$  is compact, select a finite subcover  $U_{x_1}, \dots, U_{x_m}$  and let  $h = \min\{g_{x_1}, \dots, g_{x_m}\} \in \overline{\mathcal{A}}$  by Lemma 3. By the same argument as Lemma 5,  $h(t) - f(t) < \varepsilon \forall t \in K$  from the fact that  $h(t) - f(t) \geq -\varepsilon \forall t \in K$  since this is true for each  $g_{x_1}, \dots, g_{x_m}$ .  $\square$

Next class, we will wrap up all the details, and also look at the complex version.

## 23 March 15

Recall we are proving Stone-Weierstrass:  $K$  is compact,  $\mathcal{A} \subset \mathcal{C}_{\mathbb{R}}(K)$  ( $\mathbb{R}$ -valued continuous, bounded functions on  $K$ ) where  $\mathcal{A}$  is an algebra that separates points and vanishes at no points. Want  $\overline{\mathcal{A}} = \mathcal{C}_{\mathbb{R}}(K)$ .

Recall the Lemma we just finished at the end of last class.

**Lemma 7.** For all  $f \in \mathcal{C}_{\mathbb{R}}(K)$ ,  $\forall \varepsilon > 0$ ,  $\exists g \in \overline{\mathcal{A}}$  such that  $\|f - g\| < \varepsilon$ .

(we want  $h \in \mathcal{A}$ ). The proof of Stone-Weierstrass:

*Proof.* Let  $f \in \mathcal{C}_{\mathbb{R}}(K)$ , for each  $n$  select  $g_n \in \overline{\mathcal{A}}$  such that  $\|f - g_n\| < \frac{1}{2n}$ . Select  $f_n \in \mathcal{A}$  such that  $\|f_n - g_n\| < \frac{1}{2n}$ . Hence,  $\|f - f_n\| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$ , i.e.  $f_n \rightarrow f$ .  $\square$

Conjecture: The above theorem is true with  $\mathcal{C}_{\mathbb{R}}(K)$  replaced by  $\mathcal{C}(K)$  (complex-valued included). This is false. Counter-example:  $K = S^1$  is the unit circle,  $\{z \in \mathbb{C}: |z| = 1\} = \{e^{it}: t \in [0, 2\pi]\}$ . Then some function  $f: S^1 \rightarrow \mathbb{C}$  can be represented as  $f(z)$  or  $f(e^{it})$  (require  $f(e^{i0}) = f(e^{i2\pi})$ ). Our algebra  $\mathcal{A}$  are the polynomials with  $\mathbb{C}$  coefficients,  $f(z) = \sum_{k=0}^n c_k z^k$  ( $c_k \in \mathbb{C}$ ) or  $f(e^{it}) = \sum_{k=0}^n c_k e^{ikt}$ . We have shown before that this is an algebra. It can be confirmed that  $\mathcal{A}$  that separates points and vanishes at no point. We want to show  $\overline{\mathcal{A}} \neq \mathcal{C}_{\mathbb{R}}(K)$ . We do this by showing there is some function in  $\mathcal{C}_{\mathbb{R}}(K)$  but is not in the closure of  $\mathcal{A}$ .

Let  $g(z) = \bar{z} \in \mathcal{C}(K)$ . We now dip into some complex analysis for motivation. Recall that if we take the integral along some contour (closed simple curve) and the function is holomorphic on the interior and contour, then the contour integral is 0. All of the polynomials are holomorphic, and so the integral is 0. The integral of  $\bar{z}$  will not be 0. But recall that if a sequence of functions converge uniformly, the integral also converges, and so must be 0

in this case. Which shows we can't have polynomials uniformly converge to  $\bar{z}$ . More formally, let  $P \in \mathcal{A}$ . Let's compute (secretly a contour integral)

$$\int_0^{2\pi} P(e^{it})ie^{it}dt = i \int_0^{2\pi} \sum_{k=0}^n c_k e^{i(k+1)t} dt = i \sum_{k=0}^n c_k \int_0^{2\pi} e^{i(k+1)t} dt = i \sum_{k=0}^n c_k \int_0^{2\pi} (\cos((k+1)t) + i \sin((k+1)t)) dt$$

and this evaluates to 0 on inspection of sin, cos. Next,

$$\int_0^{2\pi} g(e^{it})e^{it}ie^{it}dt = i \int_0^{2\pi} e^{-it}e^{it}dt = i \int_0^{2\pi} 1dt = 2\pi i$$

Hence, if there existed  $\{P_n\} \subset \mathcal{A}$  such that  $P_n \rightarrow g$  uniformly (on  $K$ ). Then  $P_n z \rightarrow gz$  uniformly (on  $K$ ), and hence, by Theorem 7.16,

$$\int_0^{2\pi} P_n(e^{it})e^{it}dt = 0 \rightarrow \int_0^{2\pi} g(e^{it})e^{it}dt = 2\pi$$

Contradiction. Hence,  $g(z) = \bar{z} \notin \mathcal{A}$ .

Intuitively, why does this not work? Polynomials in  $\mathbb{C}$  are all holomorphic, they all satisfy a certain PDEs. This is invariant, even with the operations of an algebra, we are always going to have another holomorphic function. But not all continuous functions are holomorphic.

What is the addition? The moral of the counter-example, is that, even though it is not the only counter-example, it presents the only obstruction to it. Recall that our original Stone-Weierstrass, we wanted it to be really simple. We saw that the most naive version, no restrictions on the algebra, did not work. If all the functions vanish at a point, this is an invariant you can't escape through the operations of an algebra. If all the functions are equal at a point, this is an invariant you can't escape through the operations of an algebra. And in this case, an algebra being holomorphic is an invariant you can't escape. One way to break out is conjugation: we've already seen it is not holomorphic.

**Theorem 34** (Complex Stone-Weierstrass (Rudin 7.33)). *Let  $K$  be a compact metric space,  $A \subset \mathcal{C}(K)$ ,  $\mathcal{A}$  is an algebra,  $\mathcal{A}$  separates points,  $\mathcal{A}$  vanishes at no point,  $\mathcal{A}$  is self-adjoint: if  $f \in \mathcal{A}$ , then  $\bar{f} \in \mathcal{A}$  ( $\bar{f}(z) = \overline{f(z)}$ ). Then,  $\overline{\mathcal{A}} = \mathcal{C}(K)$  (uniform complex closure of  $\mathcal{A}$ ).*

The proof of this is not bad, actually reduces to the real case.

It was clear that the real version of Stone-Weierstrass was a generalization the real Weierstrass theorem. But does complex Stone-Weierstrass generalize complex Weierstrass? (I.e. it follows from the other). It does: our original was complex-valued functions, but the domain was still  $[a, b] \subset \mathbb{R}$ . And if the domain is the real line, automatically self-adjoint because the conjugate is itself.

Next lecture will be the proof.

## 24 March 18

Last time, we were talking about the complex version of Stone-Weierstrass. Recall that we came across another obstruction: if all the functions in our algebra were holomorphic, there were non-holomorphic continuous functions that was not in the closure.

We are trying to prove that if  $K$  is compact,  $\mathcal{A} \subset \mathcal{C}(K)$ ,  $\mathcal{A}$  separates points, vanishes at no point, self-adjoint; then  $\overline{\mathcal{A}} = \mathcal{C}(K)$ . We are going to heavily rely on the real Stone-Weierstrass.

*Proof.* Let  $\mathcal{A}_{\mathbb{R}}$  = (real) algebra of real-valued functions in  $\mathcal{A}$ , i.e.  $\{f \in \mathcal{A} : \text{Im}(f) \equiv 0\}$ .

Step 1: Let  $f = u + iv \in \mathcal{A}$ , then  $u = \text{Re}(f) = \frac{1}{2}(f + \bar{f}) \in \mathcal{A}$ , hence  $\text{Re}(f) \in \mathcal{A}_{\mathbb{R}}$ .  $v = \text{Im}(f) = \frac{1}{2i}(f - \bar{f}) \in \mathcal{A}$ , hence  $\text{Im}(f) \in \mathcal{A}_{\mathbb{R}}$ . We are crucially using the fact here that the algebra is self-adjoint.

Step 2:  $\mathcal{A}_{\mathbb{R}}$  separates points: let  $x, y \in K$ , with  $x \neq y$ , by Lemma 4, there exists  $f \in \mathcal{A}$  such that  $f(x) = 0$ ,  $f(y) = 1$ . (Note that our proof for this does not rely on whether the functions are real or not. There is a reason Rudin put this as a separate theorem 7.31 before either proof; Zahl messed up making it seem like it relied on it being real functions by putting it inside of the real Stone-Weierstrass proof.) So  $\text{Re}(f)(x) = 0$ ,  $\text{Re}(f)(y) = 1$ . These are both in  $\mathcal{A}_{\mathbb{R}}$ , hence  $\mathcal{A}_{\mathbb{R}}$  separates points.

Step 3:  $\mathcal{A}_{\mathbb{R}}$  vanishes at no point: Let  $x \in K$ . Since  $\mathcal{A}$  vanishes at no point, there exists  $f \in \mathcal{A}$  such that  $f(x) \neq 0$ . At least one of  $\operatorname{Re}(f)(x)$  or  $\operatorname{Im}(f)(x)$  is not 0. (Slightly different way that Rudin uses: because  $f(x) = re^{i\theta} \implies e^{-i\theta}f(x) = r \in \mathcal{A}$ . This is a useful argument to be aware of.) (Essentially,  $f\bar{f} = |f|^2$  also works too. “Maybe this will make it into the next version of Rudin.”)

Step 4: By Stone-Weierstrass (Theorem 7.32),  $\overline{\mathcal{A}_{\mathbb{R}}} = \mathcal{C}_{\mathbb{R}}(K)$  = the space of continuous, bounded, real-valued functions  $f: K \rightarrow \mathbb{R}$ .

Step 5: Let  $f = u + iv \in \mathcal{C}(K)$ ,  $u, v \in \mathcal{C}_{\mathbb{R}}(K)$  (function from  $\mathbb{R}^n \rightarrow \mathbb{R}^n$  continuous only if functions continuous component-wise, proof in Rudin but also easy to do).  $\exists g, h \in \mathcal{A}_{\mathbb{R}}$  such that  $\|u - g\| < \varepsilon/2$ ,  $\|v - h\| < \varepsilon/2$ . We have then  $g, ih \in \mathcal{A}$ , so  $g + ih \in \mathcal{A}$  and  $\|f - (g + ih)\| \leq \|u - g\| + \|iv - ih\| < \varepsilon$ . Uniform limit of continuous functions is also continuous gives set inclusion the other way.  $\square$

## 24.1 Fourier Analysis

Fourier analysis is glorified linear algebra, but linear algebra in an infinite-dimensional vector space. But talking to some of you, he realized that not everyone was well-familiarized with linear algebra.

### 24.1.1 Linear Algebra Review

Let  $F = \mathbb{R}$  or  $\mathbb{C}$  (more generally  $F$  a field). A vector space over  $F$  (a.k.a. a  $F$ -vector space) is a set  $V$  along with two operations, called vector addition, denoted by  $u + v$ , and scalar multiplication  $av$  ( $u, v \in V$ ,  $a \in F$ ); these are maps  $V \times V \rightarrow V$ ,  $F \times V \rightarrow V$ . These must satisfy the vector space axioms ( $u, v, w \in V$ ,  $a, b \in F$ ):

1. Addition is associative  $u + (v + w) = (u + v) + w$ .
2. Addition is commutative  $u + v = v + u$ .
3. There is an additive identity  $0 \in V$  such that  $u + 0 = 0 + u = u$ .
4. Every element has an additive inverse, i.e.  $\forall v \in V$ ,  $\exists u \in V$  such that  $v + u = u + v = 0$ . Call  $u$  “ $-v$ ”.
5. Scalar multiplication and field multiplication are compatible  $a(bv) = (ab)v$  (on the left, we have two scalar multiplication, on the right, we have field multiplication then scalar multiplication).
6. Scalar multiplication respects the field identity  $1v = v$  where  $1$  is the multiplicative identity of the field.
7. Distributivity:  $a(u + v) = au + av$  and  $(a + b)v = av + bv$ .

Say you have a floating point number in C++, it is storing it to some level of accuracy (“the only programming language I can speak to with any level of accuracy”). Here, depending on the size of  $u, v, w$ , associativity could break. Also, octonions don’t have associativity. We know that not all operations are commutative, e.g. matrix multiplication.

Some examples of vector spaces:

- $\mathbb{R}^d, \mathbb{C}^d$  are the  $d$ -tuples of  $\mathbb{R}$  or  $\mathbb{C}$ , with operations  $(x_1, \dots, x_n) + (y_1, \dots, y_n) = (x_1 + y_1, \dots, x_n + y_n)$  and  $a(x_1, \dots, x_n) = (ax_1, \dots, ax_n)$ . Can verify satisfies axioms.
- Sequences  $(x_n)_{n \in \mathbb{N}}$  (or  $(x_n)_{n \in \mathbb{Z}}$ ) where  $x_i \in \mathbb{R}$  or  $\in \mathbb{C}$  (four examples here). Component-wise operations as before. This is an infinite-dimensional version of the one we had before.
- Sequences  $(x_n)_{n \in \mathbb{N}}$  (or  $n \in \mathbb{Z}$ ) where  $x_n = 0$  for all  $n \geq N$ . “Sequences that are eventually null.”
- Sequences  $(x_n)_{n \in \mathbb{N}, \mathbb{Z}}$  such that  $\sum_{n \in \mathbb{N}, \mathbb{Z}} |x_n| < \infty$ . Or to generalize, when  $1 \leq p < \infty$ ,  $(\sum |x_n|^p)^{1/p} < \infty$  (addition is fine, since  $|x_n + y_n|^p \leq 2^{p-1}(|x_n|^p + |y_n|^p)$ ).
- $V = \mathcal{C}_{\mathbb{R}}(\mathbb{R}), \mathcal{C}(\mathbb{C})$  or  $\mathcal{C}_{\mathbb{R}}(K), \mathcal{C}(K)$ .
- $\mathcal{R}([a, b])$ .

Next class, we will add additional structure on to the vector spaces, and see how those structures interact.

## 25 March 20

### 25.1 More linear algebra

$F = \mathbb{R}$  or  $\mathbb{C}$ ,  $V$  is a vector space over  $F$ .

*Definition 24* (Norm). A *norm* is a function  $\|\cdot\|: V \rightarrow \mathbb{R}$  that satisfies the following properties:

1. Triangle inequality:  $\|x + y\| \leq \|x\| + \|y\|$ ,  $\forall x, y \in V$
2.  $\|ax\| = |a|\|x\|$ ,  $a \in F, x \in V$
3. Positive definiteness:  $\|x\| \geq 0$  and  $\|x\| = 0$  iff  $x = 0$  (this last one can be sneaky when showing something is a norm)

Examples:

- $V = \mathbb{R}^d$  or  $\mathbb{C}^d$  where  $\|x\|_2 = \left(\sum_{j=1}^n |x_j|^2\right)^{1/2}$  or more generally, when  $1 \leq p < \infty$ ,  $\|x\|_p = \left(\sum_{j=1}^n |x_j|^p\right)^{1/p}$  (the  $1/p$  power important not just for getting scalar right, but also triangle inequality)
- When  $1 \leq p < \infty$ ,  $V = \left\{a = (a_n)_{n \in \mathbb{N}}: \left(\sum_{n \in \mathbb{N}} |a_n|^p\right)^{1/p} < \infty\right\} =: \ell^p(\mathbb{N})$  (or swap the  $\mathbb{N}$ s with  $\mathbb{Z}$ ) where the norm is  $\|a\|_p = \left(\sum_{n \in \mathbb{N}} |a_n|^p\right)^{1/p}$ , or if  $f$  is the step function of  $a_n$ , we have  $\left(\int_{\mathbb{R}} |f|^p d\alpha\right)^{1/p}$
- $V = \mathcal{C}([0, 1])$ ,  $\|f\| = \int_0^1 |f(t)| dt$
- $V = \mathcal{C}([0, 1])$ ,  $1 \leq p < \infty$ ,  $\|f\|_{L^p} = \left(\int_0^1 |f(t)|^p dt\right)^{1/p}$ , and also have  $\|f\|_{L^\infty} = \sup_{x \in [0, 1]} |f(x)|$  (definition can be weird for  $L^\infty$ , but for continuous functions, this is consistent with all the ones out there)
- $V = \mathcal{R}([0, 1])$ ,  $1 \leq p < \infty$ , and the same  $\|\cdot\|_{L^p}$  as before. But there is a problem with positive definiteness: consider the function that is equal to 0 except for one point. Hence, we get what we call positive semi-definiteness (we don't get the  $\|x\| = 0$  iff  $x = 0$ ). We could retreat back to continuous function, this is what Rudin does, but it is actually common in the theory of Lebesgue integration to go another way: we mod out by an equivalence relation.
- Define  $\sim$  on  $\mathcal{R}([0, 1])$  with  $f \sim g$  if  $\int_0^1 |f(t) - g(t)| dt = 0$ . We can verify that this is an equivalence relation as an exercise. Now let  $V = \mathcal{R}([0, 1]) / \sim$  (if  $\sim$  is an equivalence relation on  $S$ ,  $x \in S$ ,  $[x] = \{y \in S: y \sim x\}$ , then  $V = \{[f]: f \in \mathcal{R}([0, 1])\}$ , ex.  $[0] = \left\{f \in \mathcal{R}([0, 1]): \int_0^1 |f(t)| dt = 0\right\}$ ). Now we can define  $\|[f]\|_p = \left(\int_0^1 |f(t)|^p dt\right)^{1/p}$ . Is this well-defined? We need to verify that if  $[f] = [g]$ , then  $\left(\int_0^1 |f(t)|^p dt\right)^{1/p} = \left(\int_0^1 |g(t)|^p dt\right)^{1/p}$ . This is an upcoming homework problem when  $p = 2$  (proof for other  $p$  is the same). (Note  $[f] = f + [0]$ ; also not every function has a continuous representative.)

*Definition 25* (Normed Vector Space). A pair  $(V, \|\cdot\|)$  is called a *normed vector space*. This induces a metric  $d(x, y) = \|x - y\|$ . If  $(V, d)$  is complete, then we call  $(V, \|\cdot\|)$  a *Banach space*.

There is an entire field of math that studies normed vector spaces, particularly Banach spaces, called functional analysis. There are a lot of theorems that only apply to Banach spaces.

E.g.  $(\mathbb{R}^d, \|\cdot\|_p)$  is a Banach space. However,  $\mathcal{C}([0, 1]), \|\cdot\|_{L^p}$ ,  $p < \infty$ , is *not* a Banach space (say we have a sequence that tends towards the step function, there is no continuous function that can be the limit). Whether  $\mathcal{R}([0, 1]) / \sim, \|\cdot\|_{L^p}$  is a Banach space is in the next homework.

#### 25.1.1 Inner product

*Definition 26* (Inner product space). If  $V$  is a vector space, and  $\langle \cdot, \cdot \rangle$  is an inner product. Then  $(V, \langle \cdot, \cdot \rangle)$  is called an *inner product space*. An inner product induces a norm  $\|x\| = (\langle x, x \rangle)^{1/2}$ . If this is a Banach space, then  $(V, \langle \cdot, \cdot \rangle)$  is called a *Hilbert space*.



Examples:

- $V = \ell^2(\mathbb{N})$  or  $\ell^2(\mathbb{Z})$ ,  $a = (a_n)_{n \in \mathbb{N}}$ ,  $b = (b_n)_{n \in \mathbb{N}}$ , then  $\langle a, b \rangle = \sum_{n \in \mathbb{N}} a_n \overline{b_n}$ , see  $\langle a, b \rangle = \overline{\langle a, b \rangle}$ .
- $V = \mathcal{C}([0, 1])$  and  $\langle f, g \rangle = \int_0^1 f(t) \overline{g(t)} dt$  (can also do  $[-\pi, \pi]$ ).

Next time, we will see how we can split up spaces into orthonormal functions.

## 26 March 22

The main vector space we're interested in is  $V = \mathcal{R}([0, 1])$ , but we want to put a norm, even better an inner product on this space. But we've already discussed how a norm can't be defined here, so we actually use  $V = \mathcal{R}([0, 1]) / \sim$  (or  $\mathcal{R}([a, b]) / \sim$ ). We will define the inner product:

$$\langle [f], [g] \rangle = \int_a^b f(x) \overline{g(x)} dx$$

The whole point, which you'll see on homework, is that we are taking the inner product of equivalence classes, but the definition uses a representative. We have to make sure that it is independent of the choice of representative. Instead of writing that, we write

$$\langle f, g \rangle = \int_a^b f(x) \overline{g(x)} dx$$

(but this is abuse of notation, otherwise only positive semi-definite).

In measure theory,  $L^p$  functions are actually equivalence classes of functions, but we almost never write it as such.

Rudin discusses some Fourier analysis with mostly the same notation Zahl would like to use, but not all of them.

*Definition 27* (Rudin 8.10). If  $\{\phi_n\}$  is a sequence of Riemann integrable functions on  $[a, b]$  that are orthonormal, i.e.

$$\langle \phi_n, \phi_m \rangle = \int_a^b \phi_n(x) \overline{\phi_m(x)} dx = \begin{cases} 1, & \text{if } n = m \\ 0, & \text{if } n \neq m \end{cases}$$

then we call  $\{\phi_n\}$  a “system of orthonormal functions” or “an orthonormal system” on  $[a, b]$ .

(For us, orthonormal systems of functions will always be infinite, but similar stuff hold for finite orthonormal systems.) This probably predates modern linear algebra (and that it would be relevant here). Fourier was looking to solutions to the heat equation, realized we could take linear combinations to get solutions (hint hint), but used orthonormal systems to construct solutions. Because of this difference in history, some notation might differ than what you would expect in linear algebra.

Examples of orthonormal systems:

- $\{e^{2\pi i n x}\}_{n \in \mathbb{Z}}$ . You can check

$$\langle e^{2\pi i n x}, e^{2\pi i m x} \rangle = \int_0^1 e^{2\pi i n x} \overline{e^{2\pi i m x}} dx = \int_0^1 e^{2\pi i (n-m)x} dx$$

Sometimes you might see as  $x$  goes from 0 to  $2\pi$ , might alter the exponent, or need to put a constant out front to make sure that it is orthonormal (regardless, going to be orthogonal).

- $\{\sin(n\pi x)\}_{n \in \mathbb{N}} \cup \{\cos(n\pi x)\}_{n \in \mathbb{N}}$

- Wavelet basis. We start with the mother wavelet:  $f = \begin{cases} 1 & x \in [0, 1/2] \\ -1 & x \in (1/2, 1] \end{cases}$ . Then we iterate:  $f_{1,0}$  is supported on the first half,  $f_{0,1}$  is supported on the second half (and we keep the total area under the curve 1)

$$f_{1,0} = \begin{cases} 2 & x \in [0, 1/4] \\ -2 & x \in (1/4, 1/2] \\ 0 & x \in (1/2, 1] \end{cases}, \quad f_{0,1} = \begin{cases} 0 & x \in [0, 1/2] \\ 2 & x \in (1/2, 3/4] \\ -2 & x \in (3/4, 1] \end{cases}$$

And next iteration: we get functions from  $f_{1,0}$  whose support is only  $[0, 1/4]$  (and  $1, -1$  again), and functions whose support is only  $(1/4, 1/2]$ , and repeat for  $f_{0,1}$ . This is actually how image resolution (used to???) work.

**Definition 28** (Fourier coefficient). If  $\{\phi_n\}$  is an orthonormal system on  $[a, b]$ , and if  $c_n = \langle f, \phi_n \rangle$ , then we call  $c_n$  the  $n$ -th Fourier coefficient of  $f$  relative to  $\{\phi_n\}$ . If  $[a, b] = [0, 1]$  and  $\phi_n = e^{2\pi i n x}$ , we write  $c_n = \hat{f}(n)$ .

What is the geometric interpretation of  $c_n \phi_n(x)$ , i.e.  $\langle f, \phi_n \rangle \phi_n$ . Let's understand  $\{\phi_n\}$  orthonormal,  $v \in V$ ,  $\langle v, \phi_n \rangle \phi_n$ . Ex.  $V = \mathbb{R}^2$ .  $\phi_1$  is a unit vector in  $\mathbb{R}^2$ .  $\langle v, \phi_n \rangle \phi_n$  is the projection of  $v$  on  $\text{span}\{\phi_1\}$ . This is because the inner product captures how orthogonal the  $v$  is, and it is a scalar multiple of  $\phi_n$ .

How about  $V = \mathbb{R}^3$ . For simplicity, let  $\phi_1 = (1, 0, 0)$  and  $\phi_2 = (0, 1, 0)$ . Then  $\text{span}\{\phi_1, \phi_2\} =$  the  $xy$  plane. We then have  $\langle v, \phi_1 \rangle \phi_1 + \langle v, \phi_2 \rangle \phi_2$  is the orthogonal projection of  $v$  onto the  $xy$  plane. Because of our choice of basis, if  $v = (v_1, v_2, v_3)$ , we have  $\langle v, \phi_1 \rangle = v_1$ ,  $\langle v, \phi_2 \rangle = v_2$ , and so its projection is  $(v_1, v_2, 0)$ . We might ask what is the point in the plane that is closest to  $v$ . It is precisely and uniquely the projected point in the plane. This makes sense, since distance comes from the norm, which comes from the inner product, so it all makes sense.

The next theorem will be this geometric observation in the context of orthonormal systems, which are infinite dimensional.

Tangent (in regards to a question whether there are other types of projections): well, there are other coefficients we can have in front of  $\phi_1, \phi_2$ . Could be some other "product". We can think of the fiber (inverse image) of a point in the plane as a straight line up, orthogonal to the plane. We could imagine that instead of straight line, it is a specific curve. This is relevant to some math problems, but no longer in the area of linear algebra.

**Definition 29** (Fourier series). If  $\{c_n\}_{n=1}^\infty$  are the Fourier coefficients of  $f$  relative to some orthonormal system  $\{\phi_n\}_{n=1}^\infty$ , then  $\sum_{n=1}^\infty c_n \phi_n$  is called the *Fourier series* of  $f$  (relative to  $\{\phi_n\}$ ) and we write  $f \sim \sum_{n=1}^\infty c_n \phi_n$  (not an equivalence, just overloading  $\sim$ ). Note that this is just some formal series, regardless of convergence.

**Theorem 35** (Rudin 8.11). Let  $\{\phi_n\}_{n=1}^\infty$  be an orthonormal system on  $[a, b]$ . Let  $f \in \mathcal{R}([a, b])$  with Fourier coefficients  $\{c_n\}_{n=1}^\infty$ . Let  $s_n(x) = \sum_{j=1}^n c_j \phi_j(x)$ , this is called the  $n$ -th partial Fourier sum of  $f$ .

Let  $d_1, \dots, d_n \in \mathbb{C}$  and  $t_n = \sum_{j=1}^n d_j \phi_j(x)$ . Then

$$\|f - s_n\| \leq \|f - t_n\|$$

with equality iff  $s_n = t_n$ , i.e.  $d_j = c_j$  for all  $j = 1, \dots, n$  (the norm is  $L^2$ ).

This might seem specific to functions, but for any inner product space, the projection of a vector on some subspace, that projection is the unique point that minimizes the distance to the vector on the subspace.

## 27 March 25

ff missed like the first half hour of class

Theorem 8.11 ff

Let's compute:

$$\|t_n\|^2 = \int_a^b \left( \sum_{j=1}^n d_j \phi_j \right) \overline{\left( \sum_{k=1}^n d_k \phi_k \right)} dx = \sum_{j=1}^n |d_j|^2$$

Now we can compute:

$$\begin{aligned}
 \|f - t_n\|^2 &= \int_a^b |f - t_n|^2 \\
 &= \int_a^b (f - t_n) \overline{(f - t_n)} \\
 &= \int |f|^2 - \sum_{j=1}^n c_j \overline{d_j} - \sum_{j=1}^n \overline{c_j} d_j + \sum_{j=1}^n d_j \overline{d_j} \\
 &= \int |f|^2 - \sum_{j=1}^n |c_j|^2 + \underbrace{\sum_{j=1}^n |d_j - c_j|^2}_{\geq 0} \\
 &\geq \int |f|^2 - \underbrace{\sum_{j=1}^n |c_j|^2}_{\|s_n\|^2}, \text{ with equality iff } d_j = c_j \forall j
 \end{aligned}$$

(fourth line had the computation:  $-c_j^2 + (d_j - c_j)^2 = -c_j^2 + d_j^2 - d_j \overline{c_j} - \overline{d_j} c_j + c_j^2$ ). So when  $d_j = c_j$  for all  $j$ , then

$$\|f - s_n\|^2 = \|f\|^2 - \|s_n\|^2 \quad (3)$$

(something like Pythagorean). So

$$\|f - t_n\|^2 = \int |f - t_n|^2 \geq \|f\|^2 - \|s_n\|^2 = \|f - s_n\|^2$$

with equality iff  $d_j = c_j, \forall j$ .

As a consequence of (3): ff Bessel's inequality.

From now on,  $[a, b] = [0, 1]$ ,  $\{\phi_n\}_{n \in \mathbb{Z}}, \phi_n(x) = e^{2\pi i n x} = e_n(x)$ .

**Definition 30** (L-periodic).  $L > 0$ . We say  $f: \mathbb{R} \rightarrow \mathbb{C}$  is *L-periodic* if  $f(x + L) = f(x)$  for all  $x \in \mathbb{R}$ .

Ex.  $e_n(x)$  is 1 periodic for all  $n \in \mathbb{Z}$ .

Let  $V = \{f: \mathbb{R} \rightarrow \mathbb{C}, f \text{ is 1-periodic and integrable on } [0, 1]\} / \sim$ . On  $V$ , we define  $\langle f, g \rangle = \int_0^1 f(x) \overline{g(x)} dx$ . Something ff about functions being in the same equivalence class as  $[0]$  when 1-periodic.

Interesting setting to look at Fourier transform, connections with group theory (not within scope of this course). So we originally saw these functions as eigenfunctions for our system from the heat equation, etc. But there was a different way to get to Fourier analysis.  $f: \mathbb{R} \rightarrow \mathbb{C}$  that is 1 periodic has a one to one correspondence with functions in  $f: \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{C}$  (these are the set of equivalence classes  $\mathbb{R}/\mathbb{Z} - 0$ , i.e.  $x \sim y$  if  $x - y \in \mathbb{Z}$ ). This is an abelian group, and  $e_n: G \rightarrow \mathbb{C}$ , or  $e_n: G \rightarrow \{\text{complex number of magnitude 1}\}$ . Group operation:  $e_n e_m = e_{n+m}(x) e^{i y}$ .

Many of the things we are doing can be written in this more abstract setting. We have a function that maps from groups to the complex numbers, and instead of a basis, we have characters, which map from  $G$  to complex number of magnitude 1. These characters form what is called the "dual group".

Examples: we had  $\mathbb{R}/\mathbb{Z} \leftrightarrow \mathbb{Z}$  ( $\mathbb{Z}$  is the dual group), but also  $\mathbb{Z}_p \leftrightarrow \hat{\mathbb{Z}}_p$  and  $\mathbb{R} \leftrightarrow \mathbb{R}$ . Finally,  $f: \mathbb{R} \rightarrow \mathbb{C}$ ,  $\hat{f} = \int e^{2\pi i x}$ .