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. $f\colon [a,b] \to \mathbb{R}$, $c \in [a,b]$, we say that f is *differentiable at* c if $\lim_{x\to c} \frac{f(x)-f(c)}{x-c}$ 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 ε, 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 $f\colon [a,b] \to \mathbb{R}$ 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 $f'\colon [a,b] \to \mathbb{R}$.

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

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

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 ambitiuous 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'es Theorem)**.** *Let $f\colon [a,b] \to \mathbb{R}$ 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 $f\colon [a,b] \to \mathbb{R}$, 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*

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

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 $f\colon \mathbb{R} \to \mathbb{R}$, 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 explictly. 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 $j = 0, \ldots, n$, then $g^{(j)}(x_0) = f^{(j)}(x_0) - P_n^{(j)}(x_0) - \underbrace{\frac{d^j}{dt^j} \frac{A}{(n+1)!}(t - x_0)^{n+1}\Big|_{t=x_0}}_{0}$. 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+1} 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 Zahl 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 axuilliary 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 $f : \mathbb{R} \to \mathbb{R}$ (or $[-1, 1] \to \mathbb{R}$), and $f(0) = 0, f'(0) = 0, \ldots, f^{(k)}(0) = 0$ 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

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

We can see that this is infinitely differentiable on all of $\mathbb{R}$ (consider the domains seperately). For $x \le 0$, $f^{(k)}(x) = 0$ 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

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

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

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

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

And these will always exist in $\mathbb{R}$, since the interval is nonempty and $f(x)$ is bounded. Define the upper and lower Reimann 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 Δx_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 infinum is taken voer all partitions of $[a,b]$.

Definition 5. We say $f\colon [a,b] \to \mathbb{R}$ 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 $P = \{x_0, \ldots, x_n\}$ is a partition, $M_i = x_i$ and $m_i = x_{i-1}$. For the sake of concreteness, consider the evenly-spaced partition $\{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n}{n}\}$. 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, $\overline{\int_a^b} f dx \le \inf\{\frac{1}{2} + \frac{1}{2n} \colon n \in \mathbb{N}\} = \frac{1}{2}$. 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, $\underline{\int_a^b} f dx \ge \sup\{\frac{1}{2} - \frac{1}{2n} \colon n \in \mathbb{N}\} = \frac{1}{2}$. Can we now conclude that $f \in \mathcal{R}[0,1]$ and $\int_a^b f dx = \frac{1}{2}$? Well, we need that $\underline{\int_a^b} f dx \le \overline{\int_a^b} f dx$ always, which we don't have yet. [Is this not a result of the $\liminf \le \limsup$ inequality?] We will prove this actually for a more general class of itnegrals.

3.1.2 The Riemann-Stieltjes Integral

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

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

For $f\colon [a,b] \to \mathbb{R}$ 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 evalues 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 finnite, it sufficies 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 $m_j = \in \{f(x)\colon x \in [x_{j-1}, x_j]\}$. See

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

$$\begin{aligned} L(P^*, f, \alpha) = \sum_{j=1}^{i} m_j \Delta\alpha_j &+ (\inf\{f(x)\colon x \in [x_i, x^*]\})(\alpha(x^*) - \alpha(x_i)) \\ &+ (\inf\{f(x)\colon x \in [x^*, x_{i+1}]\})(\alpha(x_{i+1}) - \alpha(x^*)) \\ &+ \sum_{j=i+2}^{n} m_j \Delta\alpha_j \end{aligned}$$

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 infinum. Lastly, the final inequality is proven similarly. □

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

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

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

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

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

See that this is true for every P_2.

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

□

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 $f\colon [a,b] \to \mathbb{R}$ bounded and $\alpha\colon [a,b] \to \mathbb{R}$ 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. □

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 $\alpha\colon [a,b] \to \mathbb{R}$ be monotonically increasing. Let $f\colon [a,b] \to \mathbb{R}$ 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 ε_1 sufficiently small so that $\varepsilon_1(\alpha(b) - \alpha(a)) < \varepsilon$. Choose P as above for the corresponding ε_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, $f \in \mathcal{R}_\alpha[a,b]$. ☐

Question: Can we describe/characterze $\mathcal{R}_\alpha[a,b]$ 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 $f\colon [a,b] \to \mathbb{R}$ be monotone (increasing or decreasing) and $\alpha\colon [a,b] \to \mathbb{R}$ 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 $i = 1, \ldots, n$.

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

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

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

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

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

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. ☐

Theorem 8 (Rudin 6.10)**.** *Let $f\colon [a,b] \to \mathbb{R}$ be bounded and continuous at all but finitely many points. Let $\alpha\colon [a,b] \to \mathbb{R}$ 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 discontinuitys. Can show that the Devil's staircase $\alpha\colon [0,1] \to [0,1]$ 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 $\alpha\colon [a,b] \to \mathbb{R}$ is monotonically increasing, and $f, f_1, f_2\colon [a,b] \to \mathbb{R}$ 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 $cf \in cf \in \mathcal{R}_\alpha[a,b]$ 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 $\int f_1 d\alpha \leq \int f_2 d\alpha$

(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 $\alpha_1, \alpha_2\colon [a,b] \to \mathbb{R}$ and they're monotone increasing, and $f\colon [a,b] \to \mathbb{R}$ 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 $f\colon [a,b] \to \mathbb{R}$. 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 $\alpha\colon [a,b] \to \mathbb{R}$ 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 $T\colon C([a,b]) \to \mathbb{R}$ be linear, bounded, and non-negative. Then, there exists $\alpha\colon [a,b] \to \mathbb{R}$ 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 α uniquely define a linear bounded non-negative operator $T\colon [a,b] \to \mathbb{R}$, and T uniquely determines f, α.

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

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

$$T(f) = \int_a^b f d\alpha - \int_a^b f d\beta = \text{“} \int_a^b f d(\alpha - \beta) \text{”}$$

(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 Hanh-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 $|\int_a^b f d\alpha| \leq \int_a^b |f| \alpha$

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 $|\int_a^b f d\alpha = c\int_a^b f d\alpha = \int_a^b cf d\alpha \leq \int_a^b |f| d\alpha$ (by theorem 6.12a). □

Theorem 12 (Rudin 6.15)**.** *Let $f\colon [a,b] \to \mathbb{R}$ 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, $M_2 - m_2 \leq \varepsilon$. □

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 $f\colon [a,b] \to \mathbb{R}$ be continuous, so $f \in \mathcal{R}_\alpha[a,b]$ (always true, f is continuous and α monotone increasing). Then*

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

(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 $\forall N \geq N_0$m $|R_N| < \varepsilon$. 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, $\int_a^b f d\alpha_2 \leq K \int_a^B 1 d\alpha_2 = K(\alpha_2(b) - \alpha_2(a)) = K \sum_{n=N+1}^\infty c_n \to 0$ as $N \to \infty$. 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 ε, etc., but out of time). □

7 January 26

ff