

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 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 $f: [a, b] \rightarrow \mathbb{R}$ 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 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 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 $m_j = \inf\{f(x) : x \in [x_{j-1}, x_j]\}$. 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 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, $f \in \mathcal{R}_\alpha[a, b]$. \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 α uniquely define a linear bounded non-negative operator $T: [a, b] \rightarrow \mathbb{R}$, 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 $T: C([a, b]) \rightarrow \mathbb{R}$ 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 α 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 ε , 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 ε , 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 ϕ 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: ϕ 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

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

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 α 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- σ 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, ρ) 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, ρ) 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 “ ε 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 ε 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

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

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

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

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

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 $\{f_n(x)\}$ 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) $|(f_n(x) - f_m(x)) - (f_n(t) - f_m(t))| < \varepsilon|b - a|$ (useful for uniform convergence)
- (2) $\frac{|f_n(x) - f_m(x) - f_n(t) + f_m(t)|}{|x - t|} < \varepsilon$ (useful for differentiability)

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