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:

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

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.   □

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 $f \colon \mathbb{R} \to \mathbb{R}$ (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

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

We can see that this is infinitely differentiable on all of $\mathbb{R}$ (consider the domains seperately). 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

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

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, $\overline{\int_a^b} f \, dx \leq \inf\{\frac{1}{2} + \frac{1}{2n} \colon n \in \mathbb{N}\} = \frac{1}{2}$. 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, $\underline{\int_a^b} f \, dx \geq \sup\{\frac{1}{2} - \frac{1}{2n} \colon n \in \mathbb{N}\} = \frac{1}{2}$. Can we now conclude that $f \in \mathcal{R}[0, 1]$ and $\int_a^b f \, dx = \frac{1}{2}$? Well, we need that $\underline{\int_a^b} f \, dx \leq \overline{\int_a^b} f \, dx$ 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 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).