MATH 321 Notes

Sushrut Tadwalkar

April 9, 2024

Quotes of the day: 01/08/2024 by Dr. Joshua Zahl

"Sometimes MVT stands for most valuable theorem."

"LaTeXis the language math is written in."

1 320 Review

Definition: Differentiable at point

Recall for some $f : [a, b] \to \mathbb{R}$, $c \in [a, b]$, we say that f is ***differentiable at*** c if $\lim_{x \to c} \dfrac{f(x) - f(c)}{x - c}$ exists (as a real number); we denote this by $f'(c)$.

This is a very elementary definition of what it means for something to be differentiable, but we look a bit deeper into what it means for the limit of a function. In particular, consider the case of the limit mentioned in the definition; what does it mean for this limit to exist?

- It satisfies the $\varepsilon - \delta$ definition of a limit.
- c is a limit point in $[a, b]$; in a metric space this means that any ball about the point c has a non empty intersection with the set $[a, b]$.
- $g(x) = \dfrac{f(x) - f(c)}{x - c}$ is a function with domain $[a, b]\backslash\{c\}$.

We might ask ourselves why go through all these layers of abstraction, when the high school definition of a limit works. Well, we have to make sure that the high school definition is consistent with what we have laid out so far: for any $c \in (a, b)$, the high school definition is just fine, but back then we had to separately check the end-points $c = a$ and $c = b$ with one sided limits, which we don't have to do when we satisfy one of the things laid out above. Hence, it is worth to delve into the abstraction.

Definition: Differentiable on a set

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

Furthermore, we can keep iterating this definition: if f' is differentiable at $c \in [a, b]$, we write $f''(c) = (f')'(c)$.

Notation 1. Some alternate notations for derivatives are:

- $f(c), f'(c), f''(c), \ldots$
- $f^{(0)}(c), f^{(1)}(c), f^{(2)}(c), \ldots, f^{(k)}(c)$.

Food for thought 1. Why have co-domain $\mathbb{R}$? Why not $\mathbb{C}$, or some arbitrary field F? Why not a general set/metric space?

Similarly, why make the domain a closed interval? Why not a more general subset of $\mathbb{R}$, or even $\mathbb{C}$? Why not a general set/metric space?

We cannot really have a notion of a derivative in a topological space, because in a TS we have no notion of a distance, only open and closed sets, so it does not really make sense to be talking about the rate of change of something as we get closer to a point. This is not a complete answer, but it's hard to give a better answer at this point in time. If we google a topological derivative, there might be some constructions that come close, but nothing that is a true generalization of a derivative using arbitrary topological spaces.

Theorem: Rolle's theorem

Let $f : [a, b] \to \mathbb{R}$ be differentiable with $f(a) = f(b)$. Then, there exists $c \in (a, b)$ such that $f'(c) = 0$.

We go on to showcase one of the more important theorems in differentiation:

Theorem: Taylor's theorem

Let $f : [a, b] \to \mathbb{R}$, and $n \geq 0$ 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

$$f(x) = \underbrace{\left[\sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k\right]}_{P_n(x)} + \frac{f^{(n+1)}(c)}{(n + 1)!}(x - x_0)^{n+1}. \tag{$\dagger$}$$

Call $P_n(x)$ the "degree n Taylor expansion of f around x_0".

Note (Choice of notation)**.** While choosing notation, we have many things competing for the "attention" of the notation; for example in case of $P_n(x)$, technically it is dependent on n, f, x_0, so it should be $P_n^{f,x_0}(x)$, but this is clunky. As we do more math, we get better with choosing what information notation should encode, and what can be omitted. In this particular case, we would generally know the f and x_0 and the more important part that needs to be encoded is the degree.

Food for thought 2. Let $f : \mathbb{R} \to \mathbb{R}$ which is infinitely differentiable. Suppose $f^{(k)}(0) = 0$ for all k; is it true that f must be the zero function?

Quotes of the day: 01/10/2024

No quotes today :(

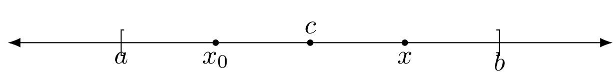

Figure 1: Visualization of points in Taylor's theorem.

Proof. We start by noting that for $n = 0$, *eq.* ($\dagger$) says $f(x) = f(x_0) + f'(c)(x - x_0)$.

Define $A \in \mathbb{R}$ by

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

Our goal here is to show that there exists a c between x_0 and x such that $f^{(n+1)}(c) = A$.

Define $g(t) = f(t) - P_n(t) - \dfrac{A}{(n + 1)!}(t - x_0)^{n+1}$.

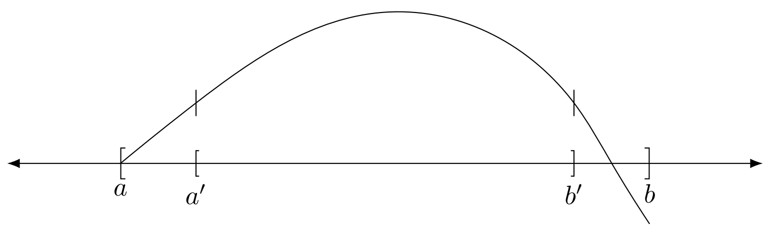

Figure 2: Visualization of how we shrink the interval to possibly apply Rolle's theorem.

Observe

$$\begin{aligned} g(x_0) =& f(x_0) - P_n(x_0) \\ =& f(x_0) - f(x_0) - 0 \\ =& 0, \end{aligned}$$

so $g(x) = 0$ by definition of A. Hence, for $j = 0, \ldots, n$,

$$\begin{aligned} g^{(j)}(x_0) =& f^{(j)}(x_0) - P_n^{(j)}(x_0) - \frac{d^j}{dt^j}\left\{\frac{A}{(n+1)!}(t-x_0)^{n+1}\right\}\Big|_{t=x_0} \\ =& f^{(j)}(x_0) - f^{(j)}(x_0) - 0 \\ =& 0, \end{aligned}$$

which tells us that $g^{(n+1)}(t) = f^{(n+1)}(t) - 0 - A$. Now, our goal is to find a c such that $g^{(n+1)}(c) = 0$.

Note that

$g(x_0) = 0, g(x) = 0$ by Rolle's thoerem, there exists c_1 between x_0 and x such that $g'(c_1) = 0$.
$g'(x_0) = 0, g'(x) = 0$ by Rolle's thoerem, there exists c_2 between x_0 and c such that $g''(c_2) = 0$.
$\vdots$
$g^{(n)}(x_0) = 0, g^{(n)}(x) = 0$ by Rolle's thoerem, there exists c_{n+1} between x_0 and c_n such that $g^{(n+1)}(c_{n+1}) = 0$.

Finally, set $c := c_{n+1}$ to conclude the proof. $\square$

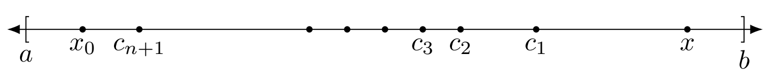

Figure 3: Visualization of the iterative process to find c_{n+1}.

Example 1. Why is Taylor's theorem so useful? We look at a few examples which illustrate this: set $x_0 = 0$,

1. f is a polynomial of degree D; $P_n(t)$ will be the first terms of f up to degree n.
2. If $f(t) = e^t$, we get
$$P_n(t) = \frac{1}{0!} + \frac{t}{1!} + \frac{t^2}{2!} + \frac{t^3}{3!} + \cdots + \frac{t^n}{n!}.$$
3. If $f(x) = \sin x$, we get
$$P_n(t) = 0 + t + 0 - \frac{t^3}{3!} + 0 + \frac{t^5}{5!} + \ldots$$

Quotes of the day: Dr. Joshua Zahl 01/12/2024

No quotes today :(

Recall food for thought 2: for a function $f : \mathbb{R} \to \mathbb{R}$ or $f : [-1, 1] \to \mathbb{R}$, given that $f(0) = 0$, and $f^{(k)}(0) = 0$ for all $k \in \mathbb{N}$, must it be true that $f(t) = 0$ for all t?

Solution. If we apply Taylor's theorem, we get

$$f(x) = \cancel{\sum_{k=0}^{n} \frac{f^{(k)}(0)}{k!}} x^n + \frac{f^{(n+1)}(c)}{(n+1)!} x^{n+1},$$

where the $P_n(x)$ term dies, but clearly the remainder term here could behave in unexpected ways (like blowing up), which would then be a function that fits our specification but is not identically zero. $\square$

Consider an example:

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

For $x > 0$, by applying chain rule, we get that $f^{(k)}(x) = Q(x)e^{-1/x}$, where $Q(x)$ is rational function in x. This function is in fact infinitely differentiable at the origin (good exercise); the intuition behind this is that the exponential function will always beat any rational function in decay at the origin, and the derivative at the origin will always be zero. However, just from Taylor's theorem, it would appear that the function is not zero at the origin, which in this case is not true. The point here is that while Taylor's theorem can aid in reconstructing a function by only using information about it at the origin, it can at times be misleading, and isn't as strong as it might seem.

2 The Riemann and Riemann-Stieltjes Integral

2.1 The Riemann integral

Definition: Partition

A ***partition*** of $[a, b]$ is a finite set $\mathcal{P} = \{x_0, x_1, \ldots, x_n\}$ with $a = x_0 < x_1 < \cdots < x_n = b$.

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

$$\begin{aligned} M_i &= \sup\{f(x) : x \in [x_{i-1}, x_i]\} \\ m_i &= \inf\{f(x) : x \in [x_{i-1}, x_i]\}; \end{aligned}$$

also, define

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

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

We define

$$\text{Upper Riemann integral} : \overline{\int_a^b} f\, dx = \inf_{\mathcal{P}} U(\mathcal{P}, f)$$

$$\text{Lower Riemann integral} : \underline{\int_a^b} f\, dx = \sup_{\mathcal{P}} L(\mathcal{P}, f);$$

the sup and inf are taken over all partitions of $[a, b]$.

Definition: Riemann integrable

We say $f : [a, b] \to \mathbb{R}$ is ***Riemann integrable*** if $\inf_{\mathcal{P}} U(\mathcal{P}, f) = \sup_{\mathcal{P}} L(\mathcal{P}, f)$, in which case we denote this number by $\int_a^b f\,dx$, and we say that $f \in \mathcal{R}[a, b]$: set of Riemann integrable functions on $[a, b]$.

A natural question that follows is what kinds of functions are Riemann integrable? We look at an example:

Example 2. Let $[a, b] = [0, 1]$, $f(x) = x$. If $\mathcal{P} = \{x_0, \ldots, x_n\}$ is a partition, $M_i = x_i$, $m_i = x_{i-1}$.

Consider $\mathcal{P} = \left\{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n}{n}\right\}$. In this case,

$$\begin{aligned}
U(\mathcal{P}, f) =& \sum_{i=1}^{n} \underbrace{\frac{i}{n}}_{M_i} \frac{1}{n} \\
=& \frac{1}{n^2} \sum_{i=1}^{n} i \\
=& \frac{1}{n^2} \frac{1}{2} n(n+1) \\
=& \frac{1}{2} + \frac{1}{2n}.
\end{aligned}$$

In particular,

$$\overline{\int_0^1} x\,dx \leq \inf\left\{\frac{1}{2} + \frac{1}{2n} : n \in \mathbb{N}\right\} = \frac{1}{2}.$$

Similarly,

$$\begin{aligned}
L(\mathcal{P}, f) =& \sum_{i=1}^{n} \frac{i-1}{n} \frac{1}{n} \\
=& \frac{1}{n^2} \frac{1}{2} n(n-1) \\
=& \frac{1}{2} - \frac{1}{2n}.
\end{aligned}$$

In particular,

$$\underline{\int_0^1} x\,dx \geq \frac{1}{2}.$$

At this point, we cannot really conclude that it is Riemann integrable, since we still need the inequality $U(\mathcal{P}, f) \geq L(\mathcal{P}, f)$, which we have not proved yet. However, rather than proving this, we will now define the Riemann-Stieltjes integral first, prove it for that, and we get it for the Riemann integral as a special case.

2.2 The Riemann-Stieltjes integral

Let $\alpha : [a, b] \to \mathbb{R}$ be (weakly) monotone increasing; let $\mathcal{P} = \{x_0, x_1, \ldots, x_n\}$ 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 $f : [a, b] \to \mathbb{R}$ bounded, define

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

Define

$$\text{Upper Riemann-Stieltjes integral}: \overline{\int_a^b} f\,dx = \inf_{\mathcal{P}} U(\mathcal{P}, f, \alpha)$$
$$\text{Lower Riemann-Stieltjes integral}: \underline{\int_a^b} f\,dx = \sup_{\mathcal{P}} L(\mathcal{P}, f, \alpha).$$

Definition: Riemann-Stieltjes integrable

We say $f : [a, b] \to \mathbb{R}$ is ***Riemann-Stieltjes integrable*** if $\inf_{\mathcal{P}} U(\mathcal{P}, f, \alpha) = \sup_{\mathcal{P}} L(\mathcal{P}, f, \alpha)$, in which case we denote this number by $\displaystyle\int_a^b f\,d\alpha$, and we say that $f \in \mathcal{R}_\alpha[a, b]$: set of Riemann-Stieltjes integrable functions on $[a, b]$.

Does $\alpha(x)$ always have to be continuous? We look at an example:

Example 3. Consider

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

what does the integral $\displaystyle\int_{-1}^1 f\,d\alpha$ look like? It it literally just $f(0)$ (it is like the Dirac-δ “function”), and this showcases the power of the Riemann-Stieltjes integral, because $\alpha(x)$ does not have to be continuous.

Quotes of the day: Dr. Joshua Zahl 01/15/2024

No quotes today :(

Definition: Refinement and common refinement (Rudin 6.3)

Let $\mathcal{P}$ and $\mathcal{P}^*$ be partitions of $[a, b]$. We say $\mathcal{P}^*$ is a ***refinement*** of $\mathcal{P}$ is $\mathcal{P} \subseteq \mathcal{P}^*$.

If $\mathcal{P}_1$ and $\mathcal{P}_2$ are partitions of $[a, b]$, their ***common refinement*** is the partition $\mathcal{P}_1 \cup \mathcal{P}_2$.

Theorem: Baby Rudin 6.4

Let $\mathcal{P}^*$ is a refinement of $\mathcal{P}$. Then, $L(\mathcal{P}, f, \alpha) \leq L(\mathcal{P}^*, f, \alpha) \leq U(\mathcal{P}^*, f, \alpha) \leq U(\mathcal{P}, f, \alpha)$.

Proof. Middle inequality we have seen before (follows from the definition of inf and sup.) Proving the leftmost inequality is equivalent to proving the rightmost inequality, so we will just prove the leftmost one.

The only interesting case is when $\mathcal{P} \subsetneq \mathcal{P}^*$, since if they're the same set, we just get equality. So it suffices to prove the inequality when $\mathcal{P}^*$ has one additional point (the minimum for two sets to not be the same one; this can be extended to any number of points by induction.) Let the additional point be x^*, and let it be between two points x_i and x_{i+1} of $\mathcal{P}$.

We proceed by comparing the two lower sums $L(\mathcal{P}, f, \alpha)$ and $L(\mathcal{P}^*, f, \alpha)$:

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

Hence,

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

□

Theorem: Baby Rudin 6.5

Let $f : [a,b] \to \mathbb{R}$ be bounded, and $\alpha : [a,b] \to \mathbb{R}$ be monotone increasing. Then,

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

Proof. Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be partitions of $[a,b]$; hence, let $\mathcal{P}^* = \mathcal{P}_1 \cup \mathcal{P}_2$. By theorem 6.4, $L(\mathcal{P}_1, f, \alpha) \leq U(\mathcal{P}_2, f, \alpha)$. Hence,

$$\underline{\int_a^b} f\, d\alpha = \sup_{\mathcal{P}_1} L(\mathcal{P}_1, f, \alpha) \leq U(\mathcal{P}_2, f, \alpha).$$

Since this is true for *every* $\mathcal{P}_2$,

$$\underline{\int_a^b} f\, d\alpha \leq \inf_{\mathcal{P}_2} U(\mathcal{P}_2, f, \alpha) = \overline{\int_a^b} f\, d\alpha.$$

□

Note. This was the missing piece that we required to show that $\displaystyle\int_0^1 x\, dx = \frac{1}{2}$.

Theorem: Baby Rudin 6.6

Let $f : [a,b] \to \mathbb{R}$ be bounded, and $\alpha : [a,b] \to \mathbb{R}$ be monotone increasing. Then $f \in \mathcal{R}_\alpha[a,b] \iff$ for all $\varepsilon > 0$, there exists $\mathcal{P}$ such that $U(\mathcal{P}, f, \alpha) - L(\mathcal{P}, f, \alpha) < \varepsilon$.

Proof. By hypothesis,

$$\sup_{\mathcal{P}} L(\mathcal{P}, f, \alpha) = \int_a^b f\, d\alpha = \inf_{\mathcal{P}} U(\mathcal{P}, f, \alpha).$$

Let $\varepsilon > 0$, then there exists a partition $\mathcal{P}_1$ such that

$$L(\mathcal{P}_1, f, \alpha) > \int_a^b f\, d\alpha - \frac{\varepsilon}{2},$$

and there exists $\mathcal{P}_2$ such that

$$U(\mathcal{P}_2, f, \alpha) < \frac{\varepsilon}{2} + \int_a^b f\, d\alpha.$$

Let $\mathcal{P} = \mathcal{P}_1 \cup \mathcal{P}_2$. By theorem 6.4,

$$L(\mathcal{P}_1, f, \alpha) \leq L(\mathcal{P}, f, \alpha) \leq U(\mathcal{P}, f, \alpha) \leq U(\mathcal{P}_2, f, \alpha).$$

Hence,

$$U(\mathcal{P}, f, \alpha) - L(\mathcal{P}, f, \alpha) < \varepsilon.$$

The other direction follows from definition. □

Quotes of the day: Dr. Joshua Zahl 01/17/2023

No quotes today :(

Theorem: Baby Rudin 6.8

Let $\alpha : [a, b] \to \mathbb{R}$ be monotone increasing, and let $f : [a, b] \to \mathbb{R}$ be continuous. Then $f \in \mathcal{R}_\alpha[a, b]$, i.e., $C([a, b]) \in \mathcal{R}_\alpha[a, b]$.

Proof. Given that f is continuous, since $[a, b]$ is compact, f is uniformly continuous. Hence, for all $\varepsilon_1 > 0$, there exists some $\delta > 0$ such that $|f(x) - f(y)| < \varepsilon$ for all x, y with $|x - y| < \delta$.

Thus, if $\mathcal{P}$ is a partition with $\Delta x_i < \delta$ for all i, then $M_i - m_i < \varepsilon_1$ for all i. Hence,

$$U(\mathcal{P}, f, \alpha) - L(\mathcal{P}, f, \alpha) \le \varepsilon_1 \sum_{i=1}^{n} \Delta\alpha_i = \varepsilon\left(\alpha(b) - \alpha(a)\right).$$

Given $\varepsilon > 0$, select ε_1 sufficiently small, such that $\varepsilon_1\left(\alpha(b) - \alpha(a)\right) < \varepsilon$. Choose $\mathcal{P}$ as above for the corresponding ε_1. We have shown that for $\varepsilon > 0$, there exists a partition $\mathcal{P}$, such that $U(\mathcal{P}, f, \alpha) - L(\mathcal{P}, f, \alpha) < \varepsilon$. Therefore, by theorem 6.6, $f \in \mathcal{R}_\alpha[a, b]$. $\square$

Food for thought 3. Can we describe/characterize $\mathcal{R}_\alpha[a, b]$ or $\mathcal{R}[a, b]$?

Turns out there is a nice bi-conditional statement to characterize these sets, but we need to develop some more machinery before we can do so.

Theorem: Baby Rudin 6.9

Let $f : [a, b] \to \mathbb{R}$ be monotone (increasing or decreasing), $\alpha : [a, b] \to \mathbb{R}$ be monotone increasing, and continuous. Then $f \in \mathcal{R}_\alpha[a, b]$.

Proof. Let $n \in \mathbb{N}$; by the intermediate value theorem, there exists a partition $\mathcal{P}$ such that $\Delta\alpha_i = \dfrac{\alpha(b) - \alpha(a)}{n}$ for all $i = 1, \ldots, n$. Note that

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

Suppose, without loss of generality, f is monotone increasing; we have $M_i = f(x_i)$ and $m_i = f(x_{i-1})$, so

$$\begin{aligned} U(\mathcal{P}, f, \alpha) - L(\mathcal{P}, f, \alpha) &= \frac{\alpha(b) - \alpha(a)}{n} \sum_{i=1}^{n} (f(x_i) - f(x_{i-1}) \\ &= \frac{\alpha(b) - \alpha(a)}{n} \left(\cancel{f(x_1)} - f(x_0) + \cancel{f(x_2)} - \cancel{f(x_1)} + \cancel{f(x_3)} - \cancel{f(x_2)} + \cdots + f(x_n) - \cancel{f(x_{n-1})}\right) \\ &= \left(\frac{\alpha(b) - \alpha(a)}{n}\right) (f(x_n) - f(x_0)) \\ &= \left(\frac{\alpha(b) - \alpha(a)}{n}\right) (f(a) - f(b)) \\ &= \frac{1}{n} \underbrace{(\alpha(b) - \alpha(a))\,(f(b) - f(a))}_{\in \mathbb{R}}, \end{aligned}$$

so given $\varepsilon > 0$, pick $n \in \mathbb{N}$ such that

$$\left| \frac{1}{n} \left(\alpha(b) - \alpha(a)\right) \left(f(b) - f(a)\right) \right| < \varepsilon.$$

For such a function, $|U(\mathcal{P}, f, \alpha) - L(\mathcal{P}, f, \alpha)| < \varepsilon$. By theorem 6.6, $f \in \mathcal{R}_\alpha[a, b]$. □

Note (f monotone decreasing)**.** In this case, the proof is pretty much the same; not tricky to work out the details.

Theorem: Baby Rudin 6.10

Let $f : [a, b] \to \mathbb{R}$ be bounded and continuous at all but finitely many points. Let $\alpha : [a, b] \to \mathbb{R}$ be monotone increasing, and continuous at every point where f is not continuous. Then, $f \in \mathcal{R}_\alpha[a, b]$.

Quotes of the day: Dr. Joshua Zahl 01/19/2024

No quotes today :(

We will now prove theorem 6.10:

Proof. Let $N := \sup\limits_{x \in [a,b]} |f|$; this is finite since f is bounded. Let $\mathcal{E} := \{e_1, \dots, e_k\}$ be the set of points where f is discontinuous.

Let $\varepsilon_1 > 0$. Since α is continuous at each $e_i \in \mathcal{E}$, we can pick $u_i < e_i < v_i$, where $u_i, v_i \in [a, b]$, such that $0 \leq \alpha(v_i) - \alpha(u_i) < \varepsilon_1$. The inequalities can be equality if $e_i = a$ or $e_i = b$.

Let $\mathcal{K} := [a, b] \setminus \bigcup\limits_{i=1}^{k} (u_i, v_i)$. Since $\mathcal{K}$ is closed and bounded, it is compact. Furthermore, since f is continuous on $\mathcal{K}$, it is uniformly continuous on $\mathcal{K}$: for all $x, y \in \mathcal{K}$ with $|x - y| < \delta$, we have $|f(x) - f(y)| < \varepsilon_1$.

Let $\{y_i\} \subseteq \mathcal{K}$ be a set of points such that for every $x \in \mathcal{K}$, there is an index i such that $y_i \leq x \leq y_{i+1}$, and $0 < y_{i+1} - y_i < \delta$. Also, let $\mathcal{P} := \{u_i, v_i\}_{i=1}^k \cup \{y_i\} \cup \{a, b\}$ (might have to re-order to pout these in increasing order). Hence,

$$\begin{aligned} 0 \leq U(\mathcal{P}, f, \alpha) - L(\mathcal{P}, f, \alpha) &= \sum_{i=1}^{n} (M_i - m_i) \Delta\alpha_i \text{ (for interval } [x_{i-1}, x_i]) \\ &\leq \underbrace{k(2N)\varepsilon_1}_{[u_i, v_i] \text{ intervals}} + \underbrace{\varepsilon_1 \left(\alpha(b) - \alpha(a)\right)}_{[y_{i-1}, y_i] \text{ intervals}}. \end{aligned}$$

Given $\varepsilon > 0$, choose ε_1 such that

$$k(2N)\varepsilon_1 + \varepsilon_1 \left(\alpha(b) - \alpha(a)\right) < \varepsilon;$$

we use the partition $\mathcal{P}$. Therefore, we have shown that $f \in \mathcal{R}_\alpha[a, b]$. □

Food for thought 4. What if f and α are both discontinuous at a common point? If $f \in \mathcal{R}_\alpha[a, b]$ always? Does it depend on f and α? Or is this never true?

Solution. Consider the case

$$f(x) = \alpha(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}.$$

Let $\mathcal{P} = \{-1 = x_0, x_1, x_2, \dots, x_n = 1\}$ be the partition on the interval $[-1, 1]$. There are two cases that we need to consider here: if we look at

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

the only interesting term is the one about the point of discontinuity (the origin); we can choose our partition to be such that the origin in between two points of the partition, but if we work this out, we get that $M_k - m_k = 1 - 0 = 1$, and $\alpha_k - \alpha_{k-1} = 1$, so

$f \notin \mathcal{R}_\alpha[a,b]$. In this case that zero is one of the partition points, we end up getting the same thing, so this function turns out to not be Riemann-Stieltjes integrable with this integrator.

However, while keeping f the same, if we slightly change α to be

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

and we let the origin be one of the partition points, we see that over the interval $[s,0]$, where $s \in \mathcal{P}$,

$$U(\mathcal{P}, f, \alpha) - L(\mathcal{P}, f, \alpha) = \sum_{i=1}^{n} (M_i - m_i)\Delta\alpha_i = (1-0)(0-0) = 0,$$

so this in fact is now Riemann-Stieltjes integrable. It is still Riemann-Stieltjes integrable if we consider an interval of the form $[0,t]$ for $t \in \mathcal{P}$.

This is particularly interesting because if we compute the integral $\int_{-1}^{1} f\, d\alpha$, we get that it evaluates to zero, which means in this case, even though it is integrable, the integrator was unable to detect the step up in the function. So we conclude that if the function and the integrator share a point of discontinuity, then sometimes the function is still Riemann-Stieltjes integrable. However, funny things happen in such situations. $\square$

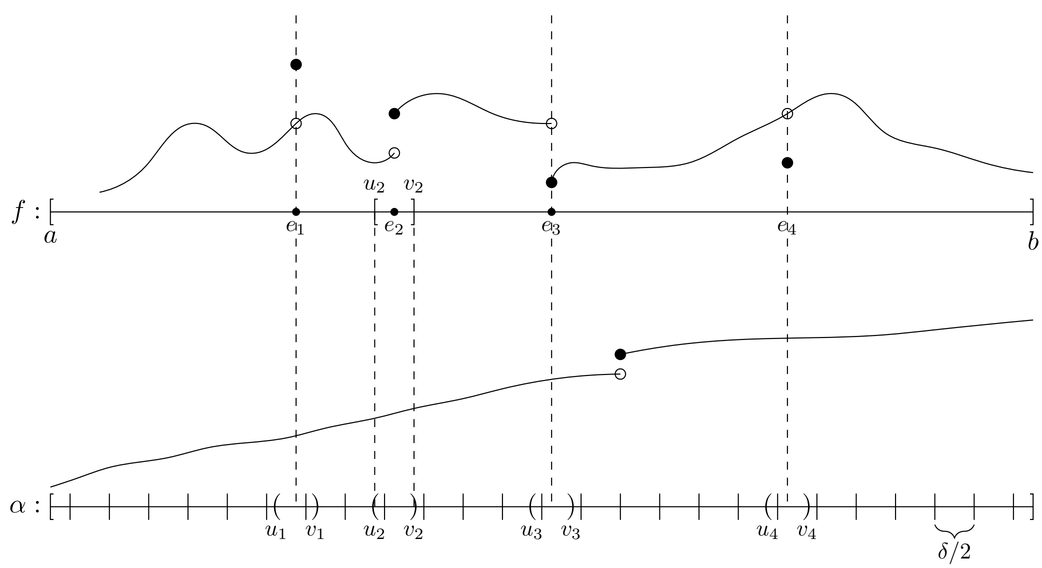

Figure 4: “Proof by picture” for the theorem.

Theorem: Baby Rudin 6.11

Let $f : [a,b] \to \mathbb{R}$ be bounded, and $\alpha : [a,b] \to \mathbb{R}$ be monotone increasing. Suppose $f \in \mathcal{R}_\alpha[a,b]$. Suppose $m \leq f(x) \leq M$ for all $x \in [a,b]$. Let $\varphi : [m,M] \to \mathbb{R}$ be continuous; then $\varphi \circ f \in \mathcal{R}_\alpha[a,b]$.

Proof. Given in Rudin. $\square$

Quotes of the day: Dr. Joshua Zahl 01/22/2024

No quotes today :(

Theorem: Properties of the Riemann-Stieltjes integral (Baby Rudin 6.12)

Let $\alpha : [a, b] \to \mathbb{R}$ be monotonically increasing, and $f, f_1, f_2 : [a, b] \to \mathbb{R}$ be functions satisfying $f, f_1, f_2 \in \mathcal{R}_\alpha[a, b]$.

a) Linearity: $f_1 + f_2 \in \mathcal{R}_\alpha[a, b]$ and $\int_a^b (f_1 + f_2)\, d\alpha = \int_a^b f_1\, d\alpha + \int_a^b f_2\, d\alpha$. For $c \in \mathbb{R}$, $cf \in \mathcal{R}_\alpha[a, b]$ and $\int_a^b cf\, d\alpha = c\int_a^b f\, d\alpha$.

b) Weak positivity/non-negativity: If $f(x) \geq 0$ for all $x \in [a, b]$, then $\int_a^b f\, d\alpha \geq 0$.

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 $c \in [a, b]$, $f \in \mathcal{R}_\alpha[a, c]$ 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 $\left|\int_a^b f\, d\alpha\right| \leq M\left(\alpha(b) - \alpha(a)\right)$.

e) Let $\alpha_1, \alpha_2 : [a, b] \to \mathbb{R}$ br monotone increasing, and $f : [a, b] \to \mathbb{R}$ satisfying $f \in \mathcal{R}_{\alpha_1}[a, b]$ 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 $c \in \mathbb{R}$, $f \in \mathcal{R}_{c\alpha_1}[a, b]$, and $\int_a^b f\, d(c\alpha_1) = c\int_a^b f\, d\alpha_1$.

Proof. The proof is given on page 128 of Baby Rudin; it's not very involved, so can be treated as an exercise as well. □

Recall $\mathscr{C}([a, b])$, the space of continuous functions $f : [a, b] \to \mathbb{R}$. Define $\|f\|_{\mathscr{C}([a,b])} = \sup_{x\in[a,b]} |f(x)|$. Hence, the metric is $d(f, g) = \|f - g\|_{\mathscr{C}([a,b])}$. We say that the pair $\left(\mathscr{C}([a, b]), \|\cdot\|_{\mathscr{C}([a,b])}\right)$ is a *normed vector space*.

Property a) of theorem 6.12 says: If $\alpha : [a, b] \to \mathbb{R}$ is monotone increasing, then the function $T(f) = \int_a^b f\, d\alpha$ is a linear function from the vector space $\mathscr{C}([a, b])$ to $\mathbb{R}$. Hence,

$$T(f + g) = T(f) + T(g)$$
$$T(cf) = cT(f).$$

Property d) says that T is bounded, i.e., $|T(f)| \leq (\alpha(b) - \alpha(a))\, \|f\|_{\mathscr{C}([a,b])}$.

Notation 2. People sometimes write Tf instead of $T(f)$, however it's the same thing. For example, in linear algebra, we write Mv where M is a matrix and v is a vector, but this is technically $M(v)$.

Property b) says that T is non-negative, i.e., if $f \in \mathscr{C}([a,b])$ with $f(x) \geq 0$ for all $x \in [a,b]$. Then $Tf \geq 0$.

In functional analysis (MATH 421), and more generally in Physics, we want to study linear functions whose domain is $\mathscr{C}([a,b])$ (or more general), and whose co-domain is $\mathbb{R}$ (or more often $\mathbb{C}$). Functions of this type are called "linear operators" or "linear functionals".

Theorem: Riesz Representation Theorem 1.0

Let $T : \mathscr{C}([a,b]) \to \mathbb{R}$ be linear, bounded, and non-negative. Then, there exists a unique monotone increasing $\alpha : [a,b] \to \mathbb{R}$, such that $Tf = \displaystyle\int_a^b f\,d\alpha$.

We want to find a better version of the theorem where we can drop the non-negative hypothesis:

Theorem: Riesz Representation Theorem 2.0

Let $T : \mathscr{C}([a,b]) \to \mathbb{R}$ be linear and bounded. Then, there exist two monotone increasing functions $\alpha, \beta : [a,b] \to \mathbb{R}$ such that

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

Note (Extension of the definition of the Riemann-Stieltjes integral)**.** Note that for monotone increasing α, β, $\alpha - \beta$ is not necessarily monotonically increasing, so we would have to change the definition of the Riemann-Stieltjes integral from monotonically increasing α to α that is the difference of monotonically increasing functions. However, we don't really need to get into that since we can just write it as the first equality shown above.

Quotes of the day: Dr. Joshua Zahl 01/24/2024

"I love nitpicking, because math is meant to be precise."

Theorem: Baby Rudin 6.13

Let $f, g \in \mathcal{R}_\alpha[a,b]$. Then

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

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

(a) *Proof.* By theorem 6.11, $\alpha(x) = x^2$, $(f+g)^2, (f-g)^2 \in \mathcal{R}_\alpha[a,b]$. By theorem 6.12(a), $(f+g)^2 - (f-g)^2 = 4fg \in \mathcal{R}_\alpha[a,b]$. Finally, by theorem 6.12(a), for $c = \dfrac{1}{4}$, $fg \in \mathcal{R}_\alpha[a,b]$. $\square$

(b) *Proof.* By theorem 6.11, $\alpha(x) = |x|$, $|f| \in \mathcal{R}_\alpha[a,b]$. Let $c = \operatorname{sgn} \displaystyle\int_a^b f\,d\alpha$, so

$$\left|\int_a^b f\,d\alpha\right| = c\int_a^b f\,d\alpha = \int_a^b cf\,d\alpha \leq \int_a^b |f|\,d\alpha.$$

(by theorem 6.12(a)). $\square$

Theorem: Baby Rudin 6.15

Let $f : [a, b] \to \mathbb{R}$ 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 $\displaystyle\int_a^b f\, d\alpha = f(s)$.

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

$$\begin{aligned} \sup_{x \in [x_1, x_2]} f(x) \leq & f(x_1) + \frac{\varepsilon}{2} \Rightarrow M_2 \leq f(x_1) + \frac{\varepsilon}{2} \\ \inf_{x \in [x_1, x_2]} f(x) \geq & f(x_1) - \frac{\varepsilon}{2} \Rightarrow m_2 \geq f(x_1) - \frac{\varepsilon}{2}. \end{aligned}$$

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

Food for thought 5. How would we change the proof if $f(x) = 1$ or $x \geq s$? We would get the same result, but would need to change the roles of x_1, x_2. If defined at neither, then probably $s \in [x_1, x_2]$. This step function is quite important in electrical engineering; it even has a special name:

Definition: Heavyside step function

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

Theorem: Baby Rudin 6.16

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

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

Note. We know that both the integral and sum exist because the sum converges, and f is bounded.

Proof. Let $R_N = \displaystyle\int_a^b f\, d\alpha - \sum_{n=1}^{N} c_n f(s_n)$. Our goal is to show that for all $\varepsilon > 0$, there exists N_0 such that for all $n \geq N_0$, $|R_N| < \varepsilon$. So, fix N, let $\alpha_1 = \displaystyle\sum_{n=1}^{N} c_n I(x - s_n)$, $\alpha_2 = \displaystyle\sum_{n=N+1}^{\infty} c_n I(x - s_n)$. By theorem 6.12(c), $\displaystyle\int_a^b f\, d\alpha = \int_a^b f\, d\alpha_1 + \int_a^b f\, d\alpha_2$, and

$f \in \mathcal{R}_{\alpha_1}[a,b]$, $f \in \mathcal{R}_{\alpha_2}[a,b]$ since f is continuous. Hence,

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

by theorem 6.15. Therefore, $R_N = \displaystyle\int_a^b f\,d\alpha_2$. Let $K = \displaystyle\sup_{x\in[a,b]} |f|$. By theorem 6.12(b),

$$\int_a^b f\,d\alpha_2 \le K \int_a^b 1\,d\alpha_2 = K[\alpha_2(b)-\alpha_2(a)] = K \sum_{n=N+1}^{\infty} c_n \to 0 \text{ as } N \to \infty.$$

In summary, we wrote the difference as the sum of a main term and a tail term, and showed that the tail term goes to zero. We could make this more formal using ε's, but we were out of time. □

Theorem: Baby Rudin 6.17

Let $f : [a,b] \to \mathbb{R}$ be bounded, and $\alpha : [a,b] \to \mathbb{R}$ be 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

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

Proof. **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, and the second one is left as an exercise.

Step-2: Since $\alpha' \in \mathcal{R}[a,b]$ for all $\varepsilon > 0$, there exists $\mathcal{P}$ (of $[a,b]$) such that $U(\mathcal{P},\alpha') - L(\mathcal{P},\alpha') < \varepsilon$. This inequality continues to hold for every refinement $\mathcal{P}'$ of $\mathcal{P}$.

We have $U(\mathcal{P},\alpha') - L(\mathcal{P},\alpha') = \displaystyle\sum_{i=1}^{n}(A_i - a_i)\Delta x_i$, where $A_i := \sup\{\alpha'(x) : x \in [x_{i-1},x_i]\}$, and $a_i := \inf\{\alpha'(x) : x \in [x_{i-1},x_i]\}$. By the Mean Value Theorem for each $i = 1,\dots,n$, there exists $t_i \in [x_{i-1},x_i]$ such that $\Delta\alpha_i = \alpha'(t_i)\Delta x_i$. Now, this suggests $\displaystyle\overline{\int_a^b} f\,d\alpha = \overline{\int_a^b} f\alpha'\,dx$, but we need to be careful:

For every $s_i \in [x_{i-1},x_i]$, we have $|\alpha'(s_i) - \alpha'(t_i)| \le A_i - a_i$, so $\displaystyle\sum_{i=1}^{n} |\alpha'(s_i)-\alpha'(t_i)|\Delta x_i \le \sum_{i=1}^{n}(A_i-a_i)\Delta x_i$ for *every* choice of $s_i \in [x_{i-1},x_i]$, $1 \le i \le n$. Let $K := \displaystyle\sup_{x\in[a,b]} |f|$, then

$$\sum_{i=1}^{n} |f(s_i)\alpha'(s_i)\Delta x_i - f(s_i)\alpha'(t_i)\Delta x_i| \le K\varepsilon. \tag{1}$$

Hence,

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

Recall that if $|a+b| \le c$, then $a \le b+c$ and $b \le a+c$.

Note. Whenever we have an inequality, we might wonder whether it is "sharp" or "tight", meaning it is equality or the closest to equality as possible.

Taking the supremum of $s_i \in [x_{i-1}, x_i]$, $i = 1, \ldots, n$, we conclude

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

Hence,

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

i.e., for every ε, we found some partition $\mathcal{P}$ that makes the inequality hold. Recall that it also holds for every refinement $\mathcal{P}'$ of $\mathcal{P}$.

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

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

Since we are considering a more strict set of partitions, will this give us the infimum we want? The answer is yes: for any non-refinement partition, we union it with $\mathcal{P}$ to get a refinement, i.e., for all $\varepsilon > 0$, $\overline{\int_a^b} f\, d\alpha \le \overline{\int_a^b} f\alpha'\, dx + K\varepsilon \Rightarrow \overline{\int_a^b} f\, d\alpha \le \overline{\int_a^b} f\alpha'\, dx$. The other three inequalities can be done as an exercise after the following note:

Note. For the inequality

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

from eq. (1), we get

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

and so on.

□

Theorem: Baby Rudin 6.19

Let $\varphi : [A, B] \to [a, b]$ be a strictly increasing, surjective, and continuous function.

Let $\alpha : [a, b] \to \mathbb{R}$ be our monotone increasing integrator, and let $f \in \mathcal{R}_\alpha[a, b]$.

Define $g := f \circ \varphi : [A, B] \to \mathbb{R}$, and $\beta := \alpha \circ \varphi : [A, B] \to \mathbb{R}$. Hence, $g \in \mathcal{R}_\beta[A, B]$ and

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

Example 4. Let $\alpha(x) = x$, and φ is differentiable. Then $d\beta = \varphi'(x)\, dx$, i.e.,

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

Proof. Partitions $\mathcal{P} := \{a = x_0, x_1, \ldots, x_n = b\}$ of $[a, b]$, and partitions $\mathcal{Q}$ of $[A, B]$ are in 1-1 correspondence via $x_i = \varphi(y_i)$.

We have $\alpha(x_i) = \alpha \circ \varphi(y_i) = \beta(y_i)$, and

$$\{f(x) : x \in [x_{i-1}, x_i]\} = \{g(y) : y \in [y_{i-1}, y_i]\}.$$

Hence, $U(\mathcal{P}, f, \alpha) = U(\mathcal{Q}, g, \beta)$ and $L(\mathcal{P}, f, \alpha) = L(\mathcal{Q}, g, \beta)$. For all $\varepsilon > 0$, since $f \in \mathcal{R}_\alpha[a, b]$ there exists a partition $\mathcal{P}$ of $[a, b]$ such that $U(\mathcal{P}, f, \alpha) - L(\mathcal{P}, f, \alpha) < \varepsilon$. Therefore, $U(\mathcal{Q}, f, \alpha) - L(\mathcal{Q}, f, \alpha) < \varepsilon$, and $g \in \mathcal{R}_\beta[A, B]$.

Finally,

$$\int_A^B g\,d\beta = \inf_{\mathcal{Q}} U(\mathcal{Q}, g, \beta) = \inf_{\mathcal{P}} U(\mathcal{P}, f, \alpha) = \int_a^b f\,d\alpha.$$

$\square$

Note (About the properties of φ)**.** Without φ being strictly increasing, surjective, and continuous in the theorem hypothesis, we won't get a 1-1 correspondence between the partition $\mathcal{P}$ and the partition $\mathcal{Q}$.

Theorem: Baby Rudin 6.20

Let $f \in \mathcal{R}[a, b]$. For $x \in [a, b]$, define $F(x) := \displaystyle\int_a^x f(t)\,dt$, $F(a) := 0$. Then, F is continuous on $[a, b]$. If $c \in [a, b]$, and f is continuous at c, then F is differentiable at c, and the derivative of $F'(c) = f(c)$.

Proof. Continuity: Let $K = \sup\limits_{t\in[a,b]} |f(t)|$. By theorem 6.12(c), for $a \le x \le y \le b$,

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

Thus,

$$|F(y) - F(x)| = \left|\int_x^y f(t)\,dt\right| \le \int_x^y |f(t)|\,dt \le \int_x^y K\,dt = K(y - x).$$

Hence, for every $\varepsilon > 0$, select $\delta = \dfrac{\varepsilon}{K}$ (or $\delta = \varepsilon$ if $K = 0$); if $|x - y| < \delta$, $|f(x) - f(y)| < \varepsilon$.

Differentiability at c: Suppose $c \ne b$, i.e., $c \in [a, b)$. Let us compute $\lim\limits_{h\searrow 0} \dfrac{F(c + h) - F(c)}{h}$.

For $h > 0$, we have

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

Here we exploit a trick, where we write $f(c) = \dfrac{1}{h}\displaystyle\int_c^{c+h} f(c)\,dt$. Hence, we have

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

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

$$\left|\frac{1}{h}\left(F(c + h) - F(c)\right) - f(c)\right| < \varepsilon \Rightarrow \lim_{h\searrow 0} \frac{F(c + h) - F(c)}{h} = f(c).$$

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

Food for thought 6. If f is *not* continuous at c, is F

(a) *Never* differentiable at c.

(b) Maybe differentiable (depends on f and c).

(c) Always differentiable.

The answer to this should be (b) Maybe differentiable, since we could have a removable discontinuity, which the Riemann integral cannot see, so it will be just fine: If $f(x) = g(x)$ *except* at one point, then $\int_a^b f(x)\,dx = \int_a^b g(x)\,dx$. In contrast if it was even a jump discontinuity, f fails to be continuous, and hence it does not work.

Quotes of the day: Dr. Joshua Zahl 01/31/2024

No quotes today :(

We showed last time that if $f : [a, b] \to \mathbb{R}$ continuous, and $F(x) = \int_a^b f(t)\,dt$, then $F'(x) = f(x)$ for all $x \in [a, b]$.

Theorem: Fundamental theorem of Calculus (Baby Rudin 6.21)

Let $f \in \mathcal{R}[a, b]$, let $F : [a, b] \to \mathbb{R}$ 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 MV, for any partition $P = \{x_0, \dots, x_n\}$ there are numbers $t_i \in [x_{i-1}, x_i]$ for $1 \le i \le n$ such that $F'(t_i) = (F(x_i) - F(x_{i-1}))/\Delta x_i$. So,

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

Then,

$$\left| \int_a^b f\,dx - (F(b) - F(a)) \right| \le U(P, f) - L(P, f).$$

Since $f \in \mathcal{R}[a, b]$, then for all $\varepsilon > 0$, there exists a partition P such that $U(P, f) - L(P, f) < \varepsilon$, and therefore,

$$\left| \int_a^b f\,dx - (F(b) - F(a)) \right| < \varepsilon.$$

□

This sets us up for proving things we know to be true about integration. We start by integration parts:

Theorem: Integration by parts (Baby Rudin 6.22)

Let $F, G : [a, b] \to \mathbb{R}$ 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

$$H(b) - H(a) = \int_a^b H'(x)\,dx$$

i.e.,

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

$\square$

In both os these results, we have this hypothesis that $f, g \in \mathcal{R}[a,b]$.

Food for thought 7. If $F; [a,b] \to \mathbb{R}$ is differentiable and $F' = f$, do we need repeat $f \in \mathcal{R}[a,b]$, or does this hold automatically, i.e., is $F' \in \mathcal{R}[a,b]$ for every $F : [a,b] \to \mathbb{R}$ differentiable?

If we ask that there exists $F : [a,b] \to \mathbb{R}$ differentiable, so that F' is discontinuous at every $x \in [a,b]$? The professor noted that "we've replaced a hard question with a harder question." We won't be doing this in class, but the answer to this question is *no*.

It is an interesting question: which sets can be the set of discontinuities of a derivative? We get that $S \subseteq [0,1]$, so can we find an F' that is discontinuous at S (where $F : [0,1] \to \mathbb{R}$ is differentiable). These are called $F - \delta$ sets.

Perhaps we wish for the derivative to blow up, but then it isn't Riemann integrable; here is a function that is worth remembering:

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

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

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

We will explore some definitions:

Definition: Absolute convergence of an integral

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

$$\int_a^\infty f(x)\,dx = \lim_{b\to\infty} \int_a^b f(x)\,dx.$$

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

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

If $f : \mathbb{R} \to \mathbb{R}$ and both $\displaystyle\int_0^\infty f(x)\,dx$ and $\displaystyle\int_{-\infty}^0 f(x)\,dx$ 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 $\displaystyle\int_{-\infty}^\infty f(x)\,dx$ converges (absolutely).

Food for thought 8. Can we construct a function that converges absolutely?

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

Quotes of the day: Dr. Joshua Zahl 02/02/2024

No quotes today :(

Definition: Riemann-Stieltjes integrability of complex valued functions

Let $f : [a, b] \to \mathbb{C}$, and $\alpha : [a, b] \to \mathbb{R}$ be monotone increasing. We say that $f \in \mathcal{R}_\alpha[a, b]$ if $\operatorname{Re} f \in \mathcal{R}_\alpha[a, b]$ and $\operatorname{Im} f \in \mathcal{R}_\alpha[a, b]$, and if so we define

$$\int_a^b f \, d\alpha := \int_a^b \operatorname{Re} f \, d\alpha + i \int_a^b \operatorname{Im} f \, d\alpha.$$

This definition can be extended for $\int_a^\infty f \, d\alpha$ and $\int_{-\infty}^\infty f \, d\alpha$ as well, and also for $f : [a, b] \to F^n$, where F is a field (generally just $\mathbb{R}$ or $\mathbb{C}$).

Food for thought 9. What happens if $\alpha : [a, b] \to \mathbb{C}$? We cannot make these "monotone increasing" because there is no order in the complex plane.

Answer. If α is differentiable, we can still try to use the Riemann integral, i.e.,

$$\int f \, d\alpha = \int f \alpha' \, dx.$$

We might also try to define what a "complex function of bounded variation" looks like. However, in practice, this is not really used since by the time we get around to thinking about complex integrators, we are usually working with Lebesgue integration.

3 Sequences and series of functions

3.1 The setup

Let $\mathcal{E}$ be a set ($\mathcal{E} = [a, b]$, $\mathcal{E} = \mathbb{R}$). Let $\{f_n\}_{n \in \mathbb{N}}$ be a sequence of functions $f_n : \mathcal{E} \to \mathcal{S}$ – usually $\mathcal{S} = \mathbb{R}$, $\mathcal{S} = \mathbb{C}$, or some metric space $(\mathcal{M}, d)$ – and let $f : \mathcal{E} \to \mathcal{S}$.

What does it mean for f_n to converge to f, i.e., $f_n \to f$, and the functions f_n all have some property P (e.g. all continuous, all integrable, etc.), must it be true that f also has P?

3.2 Point-wise convergence

Definition: Pointwise convergence

Let $\{f_n\}_{n \in \mathbb{N}}$ be a sequence of functions $F_n : \mathcal{E} \to \mathcal{M}$. If the sequence $\{f_n(x)\}_{n \in \mathbb{N}}$ converges for every $x \in \mathcal{E}$, then we say $\{f_n\}_{n \in \mathbb{N}}$ ***converges point-wise*** (on $\mathcal{E}$).

Note that there are two definitions that are used for convergence of sequence of functions:

- **Uniform convergence**: For all $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $n > N$ and for all $x \in \mathcal{E}$, $d(f_n(x), f(x)) < \varepsilon$.
- **Point-wise convergence**: For all $x \in \mathcal{E}$, given $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $n > N$, $d(f_n(x), f(x)) < \varepsilon$.

Uniform convergence is something we will cover later on in the course.

For point-wise convergence, what properties hold?

(a) Continuity: $\mathcal{E} = \mathbb{R}$. If each f_n is continuous and $f_n \to f$ point-wise, must it be true that f is continuous?

This is not necessarily true; consider

$$f_n = e^{-nx^2}, \quad \text{and} \quad f = \begin{cases} 1, & x = 0 \\ 0, & x \neq 0 \end{cases},$$

but $f_n \to f$. Also,

$$f_n = x^n, \quad \text{and} \quad f = \begin{cases} 1, & x = 1 \\ 0, & x \neq 1 \end{cases},$$

where $\mathcal{E} = [0, 1]$, but $f_n \to f$. In fact, we can make our points of discontinuity arbitrary where we want.

(b) Boundedness: If each f_n is bounded and $f_n \to f$ point-wise, must f be bounded?

This is not necessarily true; consider for $\mathcal{E} = (0, 1)$, $f_n = \dfrac{1}{x - a_n}$ such that $a_n \to 0^-$. Also, for $\mathcal{E} = \mathbb{R}$,

$$f_n = \begin{cases} x, & x \in [-n, n] \\ n, & x \in [n, \infty) \\ -n, & x \in (-\infty, n] \end{cases},$$

where $f_n \to f(x) = x$, but f_n is bounded for each $n \in \mathbb{N}$, whereas $f(x) = x$ is unbounded.

(c) Quantitative boundedness: If each f_n is bounded by 1, i.e., $|f_n(x)| \leq 1$, for all $x \in \mathcal{E}$, and $f_n \to f$ point-wise, must it be true that $|f(x)| \leq 1$ for all $x \in \mathcal{E}$?

This is in fact true: for the sake of contradiction, assume $|f(x')| > 1$ for some $x' \in \mathcal{E}$. Hence, $|f(x')| = 1 + \varepsilon > 1$, there must be some $f_n(x') > 1$.

If we change the condition in the hypothesis to be strictly less than 1, this is no longer true; we can get $|f(x)| = 1$.

(d) Riemann integrability: Given $f_n \in \mathcal{R}[0, 1]$, $f_n \to f$ point-wise, must it be true that $f \in \mathcal{R}[0, 1]$?

This is not necessarily true; we don't know many functions that fail to be integrable, but one example is

$$f(x) = \begin{cases} 1, & x \in \mathbb{Q} \\ 0, & x \notin \mathbb{Q} \end{cases}, \quad \text{and} \quad f_n(x) = \begin{cases} 1, & x \in \{q_1, q_2, \ldots, q_n\} \\ 0, & \text{otherwise} \end{cases},$$

where $\{q_n\}_{n \in \mathbb{N}}$ is an enumeration of the rationals.

Food for thought 10. Assuming f_n and f are both Riemann integrable, must $\displaystyle\int_a^b f_n(x)\,dx$ converge to $\displaystyle\int_a^b f(x)\,dx$?

Quotes of the day: Dr. Joshua Zahl 02/06/2024

No quotes today :(

We continue with solution of the problem from last time.

Solution. Essentially, we are asking that if $f_n - f \to 0$ point-wise, does it mean that $\displaystyle\lim_{n\to\infty} \int_0^1 g_n\,dx = 0$? This is in fact not necessarily true: consider

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

We get $\int_0^1 g_n\,dx = 1$. An even more extreme counter-example

$$g_n(x) = \begin{cases} n^2, & 0 < x < \frac{1}{n} \\ 0, & \text{otherwise} \end{cases}, \quad \text{then} \quad \int_0^1 g_n\,dx = n.$$

□

Food for thought 11. We have seen examples of $f_n \to f$ point-wise, where f_n are continuous or integrable and the limit is not. Why does this happen?

Suppose $f_n \to f$, $f_n : \mathbb{R} \to \mathbb{R}$ and f_n are continuous at $c \in \mathbb{R}$. Is f continuous at c?

Solution. f is continuous at c if $\lim_{x\to c} f(x) = f(c)$, so f is continuous at c if

$$\lim_{x\to c} f(x) = \lim_{x\to c}\lim_{n\to\infty} f_n(x) = \lim_{n\to\infty}\lim_{x\to c} f_n(x).$$

Similarly, $f \in \mathcal{R}[0,1]$ if $\lim_{m\to\infty}[U(P_m, f) - L(P_m, f)] = 0$.

If $f_n \to f$ point-wise, is it true that

$$\lim_{m\to\infty} U\left(P_m, \lim_{n\to\infty} f_n(x)\right) = \lim_{n\to\infty}\lim_{m\to\infty} U(P_m, f_n)$$
$$\lim_{m\to\infty} L\left(P_m, \lim_{n\to\infty} f_n(x)\right) = \lim_{n\to\infty}\lim_{m\to\infty} L(P_m, f_n)?$$

This is true for a sequence of Lipschitz continuous functions, however the limit does not have to be Lipschitz continuous (due to failure to interchange limits).

Our final example showing that we *cannot* (in general) interchange limits. Consider $a_{n,m} = \dfrac{n}{n+m}$, where $n, m \in \mathbb{N}$; we get that

$$\lim_{n\to\infty} a_{n,m} = 1 \Rightarrow \lim_{m\to\infty}\lim_{n\to\infty} a_{n,m} = 1.$$
$$\lim_{m\to\infty} a_{n,m} = 0 \Rightarrow \lim_{n\to\infty}\lim_{m\to\infty} a_{n,m} = 0.$$

□

Definition: Uniform convergence of a sequence of functions

Let $\mathcal{E}$ be a set. For a metric space $(\mathcal{M}, d)$ (i.e., $\mathbb{R}$ or $\mathbb{C}$), let $\{f_n\}$ be a sequence of functions $f_n : \mathcal{E} \to \mathcal{M}$ and $f : \mathcal{E} \to \mathcal{M}$. We say $f_n \to f$ ***uniformly*** is:

$$\text{For all } \varepsilon > 0, \text{ there exists } N \in \mathbb{N} \text{ such that, for all } n > N, x \in \mathcal{E}, d\left(f_n(x), f(x)\right) < \varepsilon.$$

Theorem: Cauchy criteria for uniform convergence (Baby Rudin 7.8)

Let $\mathcal{E}$ be a set, $(\mathcal{M}, d)$ a *complete* metric space. Then $\{f_n\}$ converges uniformly (to same $f : \mathcal{E} \to \mathcal{M}$) iff

$$\text{For all } \varepsilon > 0, \text{ there exists } N \in \mathbb{N} \text{ such that, for all } m, n > N, \text{ for all } x \in \mathcal{E}, d\left(f_n(x), f_m(x)\right) < \varepsilon,$$

which is the Cauchy criterion for uniform convergence.

Proof. ($\Rightarrow$) Suppose $f_n \to f$ uniformly. Let $\varepsilon > 0$; there exists $N \in \mathbb{N}$ such that for all $n > N$, for all $x \in \mathcal{E}$, $d(f_n(x), f(x)) < \dfrac{\varepsilon}{2}$. So for all $m, n > N$, for all $x \in \mathcal{E}$,

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

($\Leftarrow$) Suppose $\{f_n(x)\}$ satisfies the Cauchy criteria. For each $x \in \mathcal{E}$, $\{f_n(x)\}$ is a Cauchy sequence in $\mathcal{M}$, and $(\mathcal{M}, d)$ is complete, so $\{f_n(x)\}$ converges, i.e., $\lim_{n\to\infty} f_n(x)$ exists.

Define $f(x) = \lim_{n\to\infty} f_n(x)$. Let $\varepsilon > 0$; since $\{f_n(x)\}$ satisfies Cauchy criteria, there exists $N \in \mathbb{N}$ such that for all $m, n > N$, for all $x \in \mathcal{E}$, we have

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

Hence,

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

Since $f_m(x) \to f(x)$, we can select $m > N$ such that $d(f_m(x), f(x)) < \frac{\varepsilon}{2}$. Therefore,

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

$\square$

Quotes of the day: Dr. Joshua Zahl 02/08/2024

No quotes today :(

Theorem: Baby Rudin 7.11

Let $(\mathcal{M}_1, d_1)$ and $(\mathcal{M}_2, d_2)$ be metric spaces with $(\mathcal{M}_2, d_2)$ complete, i.e., $\mathbb{R}$ or $\mathbb{C}$. Let $\mathcal{E} \subseteq \mathcal{M}_1$, and let $\{f_n\}$ be a sequence of functions $f_n : \mathcal{E} \to \mathcal{M}_2$, and suppose $f_n \to f$ uniformly on $\mathcal{E}$.

Let $x \in \mathcal{M}_1$ be a limit point of $\mathcal{E}$. Suppose $\lim_{t\to x} f_n(t) = y_n$ exists for each n; $\{y_n\}$ is a convergent sequence, i.e., $y_n \to y \in \mathcal{M}_2$, and $\lim_{t\to x} f(t) = y$, i.e.,

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

Proof. **Step-1:** Show that $\{y_n\}$ converges.

It suffices to show that $\{y_n\}$ is Cauchy. Let $\varepsilon > 0$ be given. Choose N such that for all $m, n > N$, for all $t \in \mathcal{E}$, $d_2(f_n(t), f_m(t)) < \frac{\varepsilon}{3}$, and thus

$$\begin{aligned} d_2(y_n, y_m) \leq& d_2(y_n, f_n(t)) + d_2(f_n(t), y_m) \\ \leq& d_2(y_n, f_n(t)) + d_2(f_n(t), f_m(t)) + d_2(f_m(t), y_m). \end{aligned}$$

We can choose t such that the above is at most $< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon$; call this the "$\frac{\varepsilon}{3}$ trick".

In conclusion, for all $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $m, n > N$, $d_2(y_n, y_m) < \varepsilon$, i.e., $\{y_n\}$ is Cauchy, and by completeness of $(\mathcal{M}_2, d_2)$, hence convergent.

Step-2: Prove that $f(t) \to y$ as $t \to x$.

For all $t \in \mathcal{E}$ and n,

$$d_2(f(t), y) \leq d_2(f(t), f_n(t)) + d_2(f_n(t), y_n) + d_2(y_n, y). \tag{$\star$}$$

Let $\varepsilon > 0$; since $f_n \to f$ uniformly, there exists $N_1 \in \mathbb{N}$ such that for all $n > N_1$, for all $t \in \mathcal{E}$,

$$d_2(f(t), f_n(t)) < \frac{\varepsilon}{3}.$$

Since $y_n \to y$, there exists $N_2 \in \mathbb{N}$ such that for all $n > N_2$, $d_2(y_n, y) < \frac{\varepsilon}{3}$. Let $N := \max\{N_1, N_2\}$. Applying eq. ($\star$) with this choice of N, we have

$$d_2(f(t), y) \leq \frac{\varepsilon}{3} + d_2(f_N(t), y_N) + \frac{\varepsilon}{3}.$$

Since $\lim_{t \to x} f_N(t) = y_N$, there exists $\delta > 0$ such that for all $t \in \mathcal{E}$, $d_1(t, x) < \delta$, we have $d_2(f_N(t), y_N) < \frac{\varepsilon}{3}$. Hence, for all $t \in \mathcal{E}$, for all x obeying $d_1(t, x) < \delta$, we have

$$d_2(f(t), y) < \varepsilon.$$

□

Corollary: Baby Rudin 7.12

Let $(\mathcal{M}_1, d_1)$, $(\mathcal{M}_2, d_2)$, $\{f_n\}$, f, and $\mathcal{E}$ be as before. If each f_n is continuous on $\mathcal{E}$, and $f_n \to f$ uniformly, then f is continuous on $\mathcal{E}$.

Effectively, "the uniform limit of continuous functions is continuous."

Proof. f is always continuous at isolated points, so we only need to consider limit points, $x \in \mathcal{E} \cap \mathcal{E}'$, For every such x, theorem 7.11 implies

$$f(x) = \lim_{n \to \infty} f_n(x) = \lim_{n \to \infty} \lim_{t \to x} f_n(t) = \lim_{t \to x} \lim_{n \to \infty} f_n(t) = \lim_{t \to x} f(t).$$

□

3.3 Series of functions

Definition: Convergence of a series of functions to a function

Let $\mathcal{E}$ be a set, let $\{f_n\}$ be a sequence of functions, $f_n : \mathcal{E} \to \mathbb{R}$ or $\mathcal{E} \to \mathbb{C}$, and let $g : \mathcal{E} \to \mathbb{R}$ or $\mathcal{E} \to \mathbb{C}$. We say $\sum_{n \in \mathbb{N}} f_n$ converges point-wise (uniformly) to g is the sequence $S_n := \sum_{i=1}^{n} f_i$ converges point-wise (uniformly) to g.

Example 5. The series $1 + \sum_{n \in \mathbb{N}} \frac{x^n}{n!}$ converges to $g(x) = e^x$

- point-wise on $\mathbb{R}$.
- uniformly on any bounded set $\mathcal{E} \subseteq \mathbb{R}$, or any compact set $\mathcal{K} \subseteq \mathbb{R}$.

Theorem: Weierstraß M-test (Baby Rudin theorem 7.10)

Let $\mathcal{E}$ be a set, $f_n : \mathcal{E} \to \mathbb{R}$ or $\mathcal{E} \to \mathbb{C}$. If $|f_n(x)| \leq M_n$ for all $n > N_0 \in \mathbb{N}$, for all $x \in \mathcal{E}$, and if $\sum_{n=N_0}^{\infty} M_n < \infty$, then $\sum_{n \in \mathbb{N}} f_n$ converges uniformly.

Quotes of the day: Dr. Joshua Zahl 02/09/2024

No quotes today :(

Theorem: Dini's uniform convergence theorem (Baby Rudin 7.13)

Let $(\mathcal{M}, d)$ be a compact metric space (i.e., $[a, b]$), $\{f_n\}$ a sequence of functions, $f_n : \mathcal{M} \to \mathbb{R}$. Suppose that

(a) Each f_n is continuous.

(b) f_n converges *point-wise* to some continuous $f : \mathcal{M} \to \mathbb{R}$.

(c) $f_n(x) \geq f_{n+1}(x)$ for each $x \in \mathcal{M}$, $n \in \mathbb{N}$.

Then, $f_n \to f$ *uniformly* on $\mathcal{M}$.

Proof. Let $g_n = f_n - f$. Then, (a) g_n is continuous, (b) $g_n \to 0$ point-wise, (c) $g_n(x) \geq g_{n+1}(x) \geq 0$ for all $n \in \mathbb{N}$.

Goal: Prove $g_n \to 0$ uniformly, i.e.,

$$\text{For all } \varepsilon > 0, \text{ there exists } N \in \mathbb{N} \text{ such that for all } n > N, x \in \mathcal{M}, 0 \leq g_n(x) < \varepsilon.$$

Since g_n is monotonically decreasing, it is sufficient to show for all $x \in \mathcal{M}$, $g_n(x) < \varepsilon$.

Let $\mathcal{K}_n = g_n^{-1}([\varepsilon, \infty))$, $\mathcal{K}_n$ is closed, hence compact ($\mathcal{M}$ compact). Since $\{g_n\}$ is decreasing, $\mathcal{K}_n$ are nested, i.e., $\mathcal{K}_{n+1} \subseteq \mathcal{K}$. Since $g_n \to 0$ point-wise, for each $x \in \mathcal{M}$, there exists n such that $g_n(x) < \varepsilon \Rightarrow x \notin \mathcal{K}_n$. Since x was arbitrary, $\bigcap_{n=1}^{\infty} \mathcal{K}_n = \emptyset$. By Baby Rudin theorem 2.36, there exists $N \in \mathbb{N}$ such that $\mathcal{K}_N = \emptyset$, i.e.,

$$\begin{aligned} &g_N(x) < \varepsilon \quad \text{for all } x \in \mathcal{M} \\ \Rightarrow &g_n(x) < \varepsilon \quad \text{for all } x \in \mathcal{M}, n \geq N \\ \Rightarrow &|g_n(x)| < \varepsilon \quad \text{for all } x \in \mathcal{M}, n \geq N. \end{aligned}$$

□

Definition: Supremum norm

Let $(\mathcal{X}, d)$ be a non-empty metric space. Define

$$\mathscr{C}(\mathcal{X}) = \{f : \mathcal{X} \to \mathbb{C} : f \text{ is bounded and continuous}\}.$$

For each $f \in \mathscr{C}(\mathcal{X})$, define the "supremum norm"

$$\|f\| = \sup_{x \in \mathcal{X}} |f(x)|, \text{ for } f \in \mathscr{C}(\mathcal{X}), \|f\| < \infty.$$

Note. If $\mathcal{X}$ is compact in the above definition, f being bounded is superfluous.

Notation 3 (Alternate notation)**.** Some other notation for the supremum norm is: $\|f\|_{\mathscr{C}(\mathcal{X})}$, $\|f\|_{\mathscr{C}^0(\mathcal{X})}$, $\|f\|_\infty$, where the first one is probably the best one.

Note that $\mathscr{C}(\mathcal{X})$ is a vector space over $\mathbb{C}$, with $\|\cdot\|$ as the norm. For this, we have

1. $\|f\| \geq 0$, $\|f\| = 0$ iff $f(x) = 0$ for all $x \in \mathcal{X}$, i.e., $f = 0$.
2. For $\lambda \in \mathbb{C}$, $\|\lambda f\| = \lambda \|f\|$.
3. $|f(x) + g(x)| \leq |f(x)| + |g(x)| \leq \|f\| + \|g\| \Rightarrow \|f + g\| \leq \|f\| + \|g\|$.

Thus, if we define $\varrho(f, g) = \|f - g\|$, then ϱ is a metric, and $(\mathscr{C}(\mathcal{X}), \varrho)$ is a metric space. Therefore,

$$\begin{aligned} f_n \to f \text{ uniformly} &\iff \text{for all } \varepsilon > 0, \text{ there exists } N \in \mathbb{N} \text{ such that for all } x \in \mathcal{X}, \text{ for all } n > N, |f_n(x) - f(x)| < \varepsilon \\ &\iff \text{for all } \varepsilon > 0, \text{ there exists } N \in \mathbb{N} \text{ such that for all } n > N, \|f - f_n\| < \varepsilon \\ &\iff f_n \to f \text{ in the metric space } \mathscr{C}(\mathcal{X}). \end{aligned}$$

Theorem: Baby Rudin 7.15

$\mathscr{C}(\mathcal{X})$ is complete.

Proof. Let $\{f_n\}$ be a Cauchy sequence (in $\mathscr{C}(\mathcal{X})$). By theorem 7.8 (Cauchy's criteria), $f_n \to f$ uniformly for some $f : \mathcal{X} \to \mathbb{C}$. by corollary 7.12, f is continuous, since it is the uniform limit of a continuous function. Finally, f is bounded, and $f_n \to f$ uniformly, so there exists $N \in \mathbb{N}$ such that $|f(x) - f_N(x)| < 1$ for all $x \in \mathcal{X}$, so

$$\begin{aligned} &|f(x)| < |f_N(x)| + 1 \leq \|f_N\| + 1 \\ \Rightarrow\ &\|f\| < \|f_N\| + 1 < \infty, \end{aligned}$$

so f is bounded, and hence $f \in \mathscr{C}(\mathcal{X})$. ☐

Quotes of the day: Dr. Joshua Zahl 02/12/2024

No quotes today :(

We continue the talk about $\mathscr{C}(\mathcal{X})$. What about the metric space $\mathscr{C}(\mathcal{X}, \mathcal{Y})$, the bounded continuous functions $f : \mathcal{X} \to \mathcal{Y}$ ($\mathcal{X}, \mathcal{Y}$ are metric spaces) (bounded here means $f(\mathcal{X})$ is contained in some r-ball in $\mathcal{Y}$). Our metric is $d(f, g) = \sup\limits_{x \in \mathcal{X}} d_{\mathcal{Y}}(f(x), g(x))$ (check this is a metric). Is $\mathscr{C}(\mathcal{X}, \mathcal{Y})$ complete? Yes, this is true if and only if $\mathcal{Y}$ is complete; this has the same proof as before: consider constant function of Cauchy sequence that doesn't converge.

Theorem: Baby Rudin 7.16

Let $\alpha : [a, b] \to \mathbb{R}$ be a monotone increasing function. Let $\{f_n\}$ be a sequence $f_n \in \mathcal{R}_\alpha[a, b]$. Let $f : [a, b] \to \mathbb{R}$ and suppose $f_n \to f$ uniformly. Then $f \in \mathcal{R}_\alpha[a, b]$, and $\lim\limits_{n\to\infty} \int_a^b f_n \, d\alpha = \int_a^b f \, d\alpha$.

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

$$|f(x)| \leq |f(x) - f_N(x)| + |f_N(x)| < K + 1.$$

For all $\varepsilon_1 > 0$, since $f_n \to f$ uniformly, there exists $N \in \mathbb{N}$ such that for all $n > N$, for all $x \in [a, b]$, we have $|f(x) - f_n(x)| \leq \varepsilon_1$. Note that $f_n(x) - \varepsilon_1 \leq f(x) \leq f_n(x) + \varepsilon_1$, so

$$\int_a^b (f_n - \varepsilon_1) \, d\alpha \leq \underline{\int_a^b} f \, d\alpha \leq \overline{\int_a^b} f \, d\alpha \leq \int_a^b (f_n + \varepsilon_1) \, d\alpha;$$

since each m_i is the greatest lower bound, it is greater than $f_n - \varepsilon$ on each interval for any partition, so $\underline{\int_a^b} (f_n - \varepsilon_1) \, d\alpha \leq \underline{\int_a^b} f \, d\alpha$, but the upper and lower integrals converge because f_n is Riemann integrable. This is only the first inequality, but the others follow in the same manner. Rearranging, we get

$$\overline{\int_a^b} f \, d\alpha - \underline{\int_a^b} f \, d\alpha \leq \int_a^b (f_n + \varepsilon_1) \, d\alpha - \int_a^b (f_n - \varepsilon_1) \, d\alpha = \int_a^b 2\varepsilon_1 \, d\alpha = 2\varepsilon_1[\alpha(b) - \alpha(a)] := \varepsilon.$$

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

Now, we need to show that the integral agrees with the limit of the integrals of f_n, which is not something we get to assert for free as we could in the point-wise case. Given $\varepsilon_2 > 0$, we have

$$\int_a^b f_n \, d\alpha - \int_a^b \varepsilon_2 \, d\alpha \leq \int_a^b f \, d\alpha \leq \int_a^b f_n \, d\alpha + \int_a^b \varepsilon_2 \, d\alpha,$$

and thus,

$$\left| \int_a^b f \, d\alpha - \int_a^b f_n \, d\alpha \right| < \varepsilon_2[\alpha(b) - \alpha(a)] := \varepsilon.$$

□

Corollary

If $f_n \in \mathcal{R}_\alpha[a,b]$, and $\sum_{i=1}^{\infty} f_i$ 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 sketch. Let $g_n := \sum_{i=1}^{n} f_i$, and $g_n \to f$ uniformly. Apply theorem 7.16, the rest of the proof is left as an exercise. The idea is to move an infinite sum inside an integral, which we cannot normally do, but the point of theorem 7.16 is to characterize when we can do this. □

The next theorem we look at is a bit tricky to prove. Note that we can have a series of differentiable functions that converge uniformly to some function which 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 $\sum_{i=1}^{\infty} f_i$, which converges uniformly – by the Weierstraß M-test – is not differentiable (in fact this is not differentiable anywhere, but this is hard to show.) The derivative of f_n is $\frac{4^{n^2}}{2^n}\cos(4^n x)$, and basically show that the cosine terms cancel out in such a way that it cannot happen.

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

Theorem: Baby Rudin 7.17

Let $\{f_n\}$ be a sequence of functions $f_n : [a,b] \to \mathbb{R}$. Suppose

(a) Each $\{f_n\}$ is differentiable on $[a,b]$.

(b) There exists $x_0 \in [a,b]$ such that $\{f_n(x_0)\}$ converges.

(c) f_n' converge uniformly on $[a,b]$.

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

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

We won't do the proof in this lecture, but we will get started with an important estimate:

Step-1: Show that $f_n \to f$ uniformly.

Let $\varepsilon > 0$; let $N \in \mathbb{N}$ be large enough such that $|f_m(x_0) - f_n(x_0)| < \varepsilon$ for all $m, n > N$ (convergent sequences are Cauchy) and $|f'_m(x) - f'_n(x)| < \varepsilon$ for all $m, n > N$, for all $x \in [a, b]$. Here is the crucial idea of this proof: we apply MVT to the difference $f_m - f_n$. For $x, t \in [a, b]$, $x \neq t$, there exists $c \in [x, t]$ such that

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

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

1. $|(f_n(x) - f_m(x)) - (f_n(t) - f_m(t))| < \varepsilon|b - a|$; this is useful for uniform convergence.
2. $\dfrac{|(f_n(x) - f_m(x)) - (f_n(t) - f_m(t))|}{|x - t|} < \varepsilon$; this is useful for differentiability.

Quotes of the day: Dr. Joshua Zahl 02/16/2024

"It's like looking for hay in a haystack, but we never get any hay." - On trying to write down a function that is continuous everywhere, but differentiable nowhere.

We pick up the proof of theorem 7.17 from last time:

Proof. Till now, we have fixed an $\varepsilon > 0$, and found an $N \in \mathbb{N}$ such that for all $m, n > N$,

$$|f_n(x_0) - f_m(x_0)| < \varepsilon \quad \text{and} \quad |f'_n(x) - f'_m(x)| < \varepsilon, \text{ for all } x \in [a, b].$$

Furthermore, using MVT, we get that for $x, t \in [x, b]$ ($x \neq t$):

1. $|[f_n(x) - f_m(x)] - [f_n(t) - f_m(t)]| < \varepsilon|b - a|$; this is useful for uniform convergence.
2. $\dfrac{|[f_n(x) - f_m(x)] - [f_n(t) - f_m(t)]|}{|x - t|} < \varepsilon$; this is useful for differentiability.

The goal here is to prove this in steps:

1. Show that there exists some f to which f_n uniformly converges.
2. Prove the statement of the theorem for the derivative.

By (1), for $t = x_0$, consider $m, n > N$, $x \in [a, b]$, which gives us

$$|f_n(x) - f_m(x)| \leq |(f_n(x) - f_m(x)) - (f_n(x_0) - f_m(x_0))| + |f_n(x_0) - f_m(x_0)| \leq \varepsilon_1(b - a) + \varepsilon_1.$$

Hence, let $\varepsilon := \varepsilon_1(b - a + 1)$, and $\{f_n\}$ satisfies the Cauchy criterion for uniform convergence.

We move on to step-2: showing that $f' = \lim\limits_{n\to\infty} f'_n$. Fix $x \in [a, b]$; we first need to show that f is differentiable at x. Since the derivative involves limits, this will involve a careful interchange of limits. Define

$$\varphi_n(t) := \frac{f_n(t) - f_n(x)}{t - x}, \quad \varphi(t) := \frac{f(t) - f(x)}{t - x}.$$

Hence, $f'_n = \lim\limits_{t\to x} \varphi_n(t)$, and similarly we get f', if it exists. Inequality (2) says that for $m, n > N$,

$$|\varphi_n(t) - \varphi_m(t)| \leq \varepsilon,$$

i.e., $\{\varphi_n\}$ satisfies Cauchy criterion for uniform convergence, which implies φ_n converges uniformly on the domain $[a, b]\backslash\{x\}$; here x is a limit point, which is something that comes up later. This says *something* about φ_n, but we don't know if it converges to $\varphi(t)$ at any fixed t. Showing point-wise is enough: fix $t \in [a, b]\backslash\{x\}$. Hence,

$$\varphi_n(t) - \varphi(t) = \left|\frac{(f_n(t) - f_n(x)) - (f(t) - f(x))}{t - x}\right| \leq \underbrace{\left|\frac{f_n(t) - f(t)}{t - x}\right| + \left|\frac{f_n(x) - f(x)}{t - x}\right|}_{\to 0 \text{ as } n\to\infty}.$$

In conclusion, $\varphi_n \to \varphi$ point-wise, hence uniformly on $[a, b]\backslash\{x\}$. Finally, since x is a limit point of $[a, b]\backslash\{x\}$, and $\varphi_n \to \varphi$ uniformly on $[a, b]\backslash\{x\}$, we apply theorem 7.11 to conclude that

$$f'(x) = \lim_{t\to x} \varphi(t) = \lim_{t\to x} \lim_{n\to\infty} \varphi_n(t) = \lim_{n\to\infty} \lim_{t\to x} \varphi_n(t) = \lim_{n\to\infty} f'_n(x)$$

when the limit exists, which it does in this case. $\square$

3.4 The Weierstraß function

Theorem: Baby Rudin 7.18

The exists a continuous function $f : \mathbb{R} \to \mathbb{R}$ such that $f'(x)$ does not exist for any $x \in \mathbb{R}$.

One way to do this is construct such a function. Note that if we pick a function from $\mathcal{C}(\mathbb{R})$ at random, it will almost surely be a function that is not differentiable anywhere. However, if we try to write a function down, it is usually differentiable (this could be seen as an analogue to the fact that most of the numbers that exist are transcendental, but if we think of a real number, it will almost always be algebraic.) This happens often in math: we can prove that almost always functions have a certain property, but it's significantly harder (if even feasible) to write an example down.

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

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

For large n, note that $f(x)$ is very small, but very spiky. However, note that the 4^n is getting large faster than $\left(\frac{3}{4}\right)^n$ is getting small, i.e., if we multiply them, we get 3^n, which still blows up. Hence, we get that $(f(x)$ is 3^n-Lipschitz.

Fix $x \in \mathbb{R}$. We wish to show that $f(x)$ is not differentiable at x. It suffices to find $\delta_m \searrow 0$ such that

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

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

The full proof is given in Baby Rudin, or can be treated as an exercise. ☐

Quotes of the day: Dr. Joshua Zahl 02/26/2024

"It is nice when sets are compact."

4 Equicontinuity

Definition: Equicontinuity

Let $(\mathcal{X}, d)$ be a metric space, let $\mathcal{E} \subseteq \mathcal{X}$, and let $\mathscr{F}$ be a family (i.e., a set) of functions $f : \mathcal{E} \to \mathbb{C}$. We say that $\mathscr{F}$ is equicontinuous if

$$\text{for all } \varepsilon > 0, \text{ there exists } \delta > 0 \text{ such that for all } x, y \in \mathcal{E} \text{ with } d(x,y) < \delta, \text{ for all } f \in \mathscr{F}, \ |f(x) - f(y)| < \varepsilon.$$

Note (Co-domain)**.** Note that the co-domain of these functions can be generalized to any complete metric space (or maybe any metric space?), and this works just fine.

Note (Remarks about equicontinuous functions)**.** Note the following about equicontinuous functions:

1. If $\mathscr{F}$ is equicontinuous, each $f \in \mathscr{F}$ is uniformly continuous.

2. The converse of (1) is false:

   (a) Consider $\mathcal{X} = [0, 1]$, $\mathscr{F} = \{f_n\}_{n=1}^{\infty}$, and $f_n(x) = x^n$.

   (b) Consider $f_n(x) = nx$ in the same metric space.

3. If $\mathscr{F}$ is finite, and each $f \in \mathscr{F}$ is uniformly continuous, then $\mathscr{F}$ is equicontinuous.

Theorem: Baby Rudin 7.24

Let $(\mathcal{K}, d)$ be a compact metric space, and let $f_n : \mathcal{K} \to \mathbb{C}$ be continuous functions. Suppose $\{f_n\}_{n=1}^{\infty}$ converge uniformly on $\mathcal{K}$. Then, $\{f_n\}$ is equicontinuous.

Note (Notation)**.** We acknowledge slight abuse of notation in the theorem statement: we have defined equicontinuity for a family of functions, but a sequence may have repeats, so it isn't exactly a family. However, this is fine because we can just let the family be the set of sequence elements, which will never be empty; it would, however, funnily enough be fine if it was the empty set, since by our definition the empty set is equicontinuous.

Proof. We use the "$\varepsilon/3$ argument".

Let $\varepsilon > 0$; since $\{f_n\}$ converges uniformly, there exists $N \in \mathbb{N}$ such that for all $m, n \geq N$, for all $x \in \mathcal{K}$,

$$|f_n(x) - f_m(x)| < \frac{\varepsilon}{3}.$$

Since $\mathcal{K}$ is compact, each f_n is uniformly continuous, the family $\{f_1, \ldots, f_N\}$ is equicontinuous, i.e., there exists $\delta > 0$ such that for all $x, y \in \mathcal{K}$ with $d(x, y) < \delta$, for all $n \leq N$,

$$|f_n(x) - f_n(y)| < \frac{\varepsilon}{3}.$$

Now, if $n > N$ and $x, y \in \mathcal{K}$ with $d(x, y) < \delta$,

$$|f_n(x) - f_n(y)| \leq \underbrace{|f_n(x) - f_N(x)|}_{<\varepsilon/3} + \underbrace{|f_N(x) - f_N(y)|}_{<\varepsilon/3} + \underbrace{|f_N(y) - f_n(y)|}_{<\varepsilon/3} < \varepsilon,$$

by the Cauchy criterion for uniform convergence. ∎

Theorem: Baby Rudin problem 7.16

Let $\mathcal{K}$ be a compact metric space, $\{f_n\}$ an equicontinuous family of functions, $f_n : \mathcal{K} \to \mathbb{C}$. If $\{f_n\}_{n=1}^{\infty}$ converges point-wise on $\mathcal{K}$, then $\{f_n\}_{n=1}^{\infty}$ converges uniformly.

Proof. We once again do an "$\varepsilon/3$ argument".

Let $\varepsilon > 0$; select $\delta > 0$ such that for all $x, y \in \mathcal{K}$ with $d(x, y) < \delta$, for all f_n,

$$|f_n(x) - f_n(y)| < \frac{\varepsilon}{3}.$$

Since $\mathcal{K}$ is compact, the open cover $\{\mathcal{N}(\delta; x)\}_{x \in \mathcal{K}}$ has a finite subcover, $\mathcal{N}(\delta; x_1), \mathcal{N}(\delta; x_2), \ldots, \mathcal{N}(\delta; x_l)$.

Thus, given $x \in \mathcal{K}$, there exists x_j such that $d(x, x_j) < \delta$. So, for $m, n \in \mathbb{N}$,

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

Since $\{f_n\}$ converges point-wise, for each $j = 1, \ldots, l$, there exists N_j such that for all $m, n \geq N_j$,

$$|f_n(x_j) - f_m(x_j)| < \frac{\varepsilon}{3}.$$

Let $N := \max\{N_1, \dots, N_l\}$; then for all $m, n \geq N$, for all $j \in \{1, \dots, l\}$, we have

$$|f_n(x_j) - f_m(x_j)| < \frac{\varepsilon}{3}.$$

Therefore, we conclude that

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

□

Note (General math advice from the professor)**.** As we get better at math (especially analysis), it is almost necessary to remember the proofs of the theorems, because while problem statements might not be the exact same, sometimes the same proof techniques are used. However, it is obviously not feasible to memorize every single theorem's proof (unless you are actually capable of that), so it is worth abstracting it to something like "$\varepsilon/3$ argument".

Quotes of the day: Dr. Joshua Zahl 02/28/2024

No quotes today :(

Definition: Boundedness

Let $(\mathcal{X}, d)$ be a metric space, $\mathcal{E} \subseteq \mathcal{X}$, and $\mathscr{F}$ be a family of functions $f : \mathcal{E} \to \mathbb{C}$. We say $\mathscr{F}$ is ***point-wise bounded*** if there exists $\varphi : \mathcal{E} \to \mathbb{R}$ such that $|f(x)| \leq \varphi(x)$ fr every $x \in \mathcal{E}$, and $f \in \mathscr{F}$. We say $\mathscr{F}$ is ***uniformly bounded*** if there exists $M \in \mathbb{R}$ such that $|f(x)| \leq M$ for all $x \in \mathcal{E}$ and $f \in \mathscr{F}$.

How would we generalize this, i.e., for $f : \mathcal{E} \to \mathcal{Y}$, where $(\mathcal{Y}, \rho)$ is a metric space? Point-wise is harder to talk about, but in the uniform case, we say that $f(\mathcal{E})$ is contained with a bounded set; here our bounded set is centred at zero, but could be anywhere. However, we will just stick with looking at functions to $\mathbb{C}$.

Theorem: Baby Rudin 7.23

Let $\mathcal{X}$ be a metric space, $\mathcal{E} \subseteq \mathcal{X}$, $\mathcal{E}$ countable. Let $f_n : \mathcal{E} \to \mathbb{C}$, and suppose $\{f_n\}$ is point-wise bounded on $\mathcal{E}$. Then there exists a subsequence that converges point-wise on $\mathcal{E}$.

Proof sketch. We will do a diagonalization argument. The idea here is that we get a sub-sequence that would for one element of $\mathcal{E}$, then a sub-sequence that works for two elements of $\mathcal{E}$, etc. □

Proof. Let $\mathcal{E} := \{x_1, x_2, \dots\}$. We know $\{f_n(x_1)\}_{n=1}^{\infty}$ is a bounded sequence of complex numbers, and hence has a convergent sub-sequence $\{f_{1,k}\}_{k=1}^{\infty}$; note that this is a sequence of functions, it is *not* evaluation at x_1. We will construct successive such sub-sequences.

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

where $\{f_{i,k}\}_{k=1}^{\infty}$ is a convergent sub-sequence of $\{f_{i-1,k}\}_{k=1}^{\infty}$. Consider the diagonal sequence $f_{i,i}$; this converges at x_j for every j. Because $\{f_{i,i}\}$ is a sub-sequence of $\{f_{j,k}\}_{k=1}^{\infty}$, except possibly for the first $j-1$ elements, we conclude the proof. □

Theorem: Arzelá-Ascoli theorem (Baby Rudin 7.25)

Let $\mathcal{K}$ be a compact metric space, $\{f_n\} \subseteq \mathscr{C}(\mathcal{K})$ be equicontinuous and point-wise bounded. Then

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

(b) $\{f_n\}$ has a uniformly convergent sub-sequence.

Before we prove this, we recall the definition of sequential compactness (Bolzano-Weierstraß): every convergent sequence has a convergent subsequence. However, in this case we cannot really say that this statement means "compact", because it is a bit vague. We need to specify uniform convergence here, because if it's point-wise convergence, then this is not a sequence in $\mathscr{C}(\mathcal{K})$ anymore, but in $\mathbb{C}$ (or $\mathbb{R}$); this doesn't really help us comment on the compactness of $\{f_n\}$.

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

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

where x_i is a point with $x \in \mathcal{N}(\delta; x_i)$. Hence, we have uniform boundedness.

(b) *Step-1:* Let $\mathcal{E}$ be a countable dense subset of $\mathcal{K}$. The existence of this is a straightforward exercise using covers of balls having radii equal to $\dfrac{1}{n}$ for all $n \in \mathbb{N}$ and compactness. By theorem 7.23, there exists a sub-sequence of $\{f_n\}$ that converges point-wise on $\mathcal{E}$; let this sequence be $\{g_i\}$. We show that $\{g_i\}$ satisfies the Cauchy criterion for convergence:

Step-2: Let $\varepsilon > 0$. By equicontinuity of $\{g_i\}$, there exists $\delta > 0$ such that for all $x, y \in \mathcal{K}$ having $d(x, y) < \delta$, for all $i \in \mathbb{N}$, $|g_i(x) - g_i(y)| < \varepsilon/3$. Since $\mathcal{E} \subseteq \mathcal{K}$ is dense, $\{\mathcal{N}(\delta; y)\}_{y \in \mathcal{K}}$ is a cover of $\mathcal{K}$, thus there exists a finite subcover $\{\mathcal{N}(\delta; y_1), \ldots, \mathcal{N}(\delta; y_l)\}$, where $y_i \in \mathcal{E}$ for all $1 \leq i \leq l$. For all $x \in \mathcal{K}$, there exists s such that $d(x, y_s) < \delta$.

Step-3: Now all that remains to do is an $\varepsilon/3$ argument. For $x \in \mathcal{K}$, $i, j \in \mathbb{N}$, we have

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

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

$\square$

Note (Importance of compactness)**.** We used compactness in an essential manner in parts (a) and (b); a good exercise is to find out how this fails when $\mathcal{K}$ is not compact.

Quotes of the day: Dr. Joshua Zahl 03/01/2024

"Not sure if heavyside was a person...might've been."

5 Stone-Weierstraß theorem

5.1 Weierstraß approximation

Definition: Convolution

Let $f, g : \mathbb{R} \to \mathbb{R}$ or $\mathbb{R} \to \mathbb{C}$ be Riemann integrable over all of $\mathbb{R}$, i.e., the following integrals exist:

$$\lim_{z \to -\infty} \int_z^0 f(x)\, dx \quad \text{and} \quad \lim_{z \to \infty} \int_0^z f(x)\, dx.$$

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

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

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

Note. For most examples that we care about, the domain is $\mathbb{R}$.

Exercise 1. If $f * g(x)$ exists, then $g * f(x)$ exists and $f * g(x) = g * f(x)$.

Definition: Approximate identity

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

(a) $\displaystyle\int_{-\infty}^{\infty} f_n(t)\,dt = 1$ for all n.

(b) There exists $M \geq 0$ such that

$$\int_{-\infty}^{\infty} |f_n(t)|\,dt \leq M, \text{ for all } n \in \mathbb{N}.$$

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

(c) For all $\delta > 0$,

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

Example 6. Consider the function $f_n(t) = nf(nt)$, such that

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

Also, the limit of this sequence obeys

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

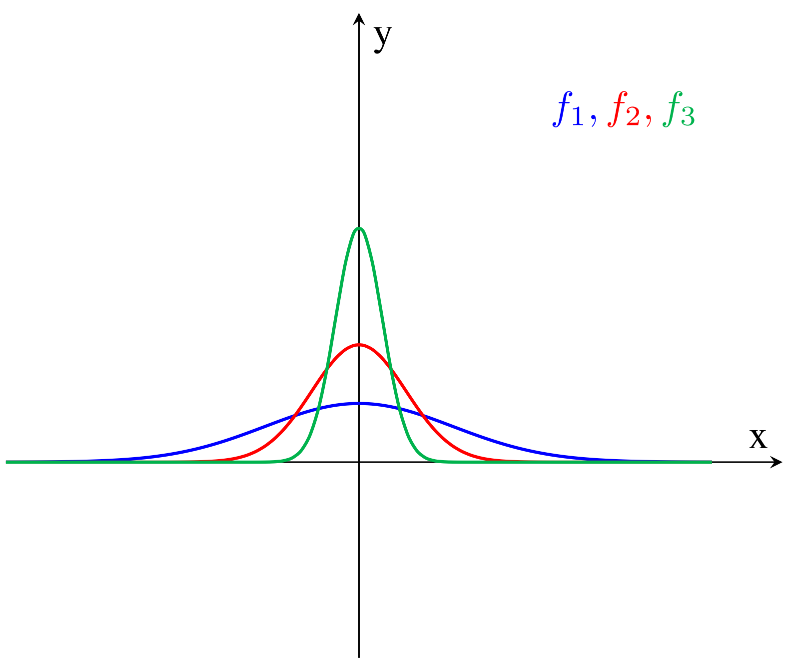

Figure 5: Plot of $f_n(x)$ for $n = 1, 2, 3$. The trend continues, and the functions become narrower and taller for larger n.

We can see that we're slowly approaching the Dirac-δ "function".

We will now build on the example further:

Example 7. Consider the function

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

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

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

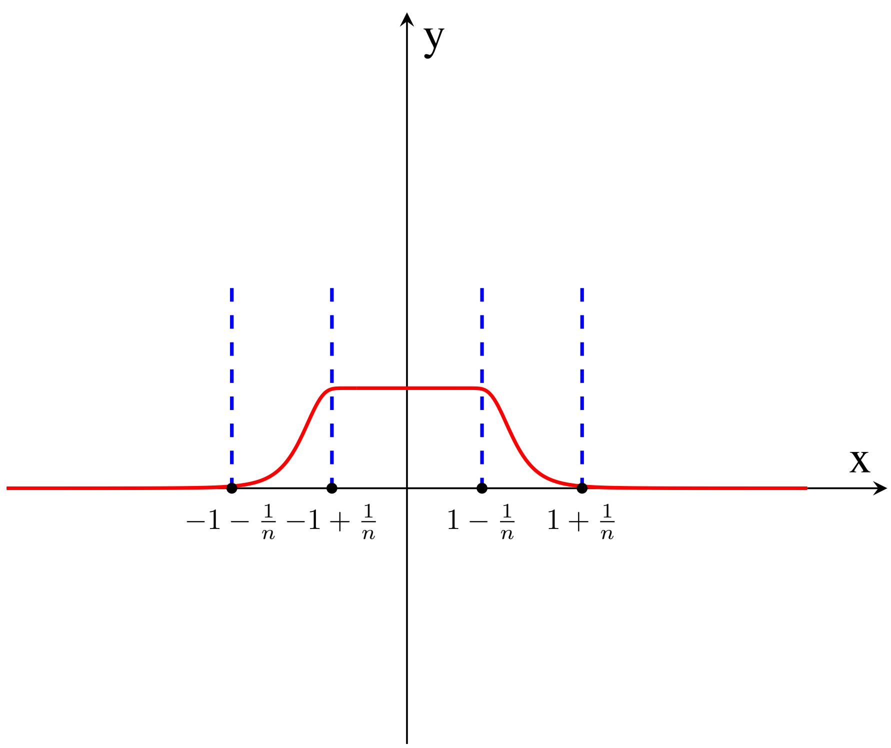

Figure 6: How the plot would look for a convolution of $g(x)$ with $f_n(x)$ at some n.

Quotes of the day: Dr. Joshua Zahl 03/04/2024

"What is this strange light?" - On sunlight coming through the window.

"The human eye is surprisingly bad at distinguishing between C^∞ and twice-differentiable functions."

Example 8. Consider the function

$$f_n(x) = \begin{cases} c_n(1-x^2)^n & -1 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}.$$

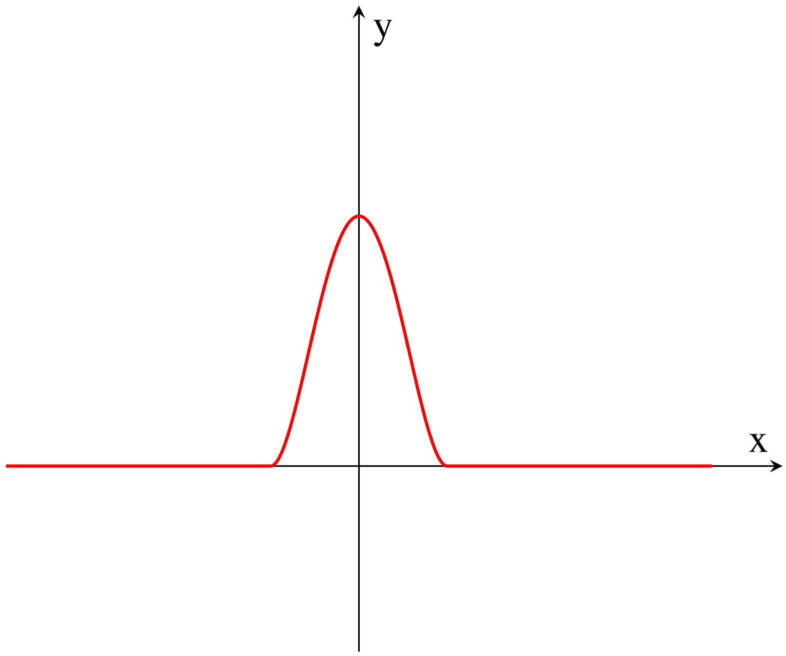

Figure 7: Example of how $f_n(x)$ would look; this plot is for $c_n = 2\sqrt{5}$.

Note that this function looks quite a bit similar to the previous example, but this is because "the human eye is surprisingly bad at distinguishing between C^∞ and twice-differentiable functions."

Choose c_n such that

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

Is $\{f_n\}$ here an approximate identity? Note that

(a) $\displaystyle\int f_n = 1$ is true by definition.

(b) $\displaystyle\int |f_n| = \int f_n = 1$ also holds, so it is bounded.

(c) Finally, we require that $\displaystyle\int_{-\infty}^{-\delta} |f_n| = \int_{\delta}^{\infty} |f_n| = 0$; this is just a calculation.

Note that $f_n(x) \geq \dfrac{1}{2}$ on the interval $\left[-\dfrac{1}{2\sqrt{n}}, \dfrac{1}{2\sqrt{n}}\right]$. Hence,

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

Thus, $c_n \leq 2\sqrt{n}$ (note that $100n^{100}$ would also be fine; what we care about is the bound).

Hence, for $\delta > 0$,

$$c_n(1-x^2)^n \leq \underbrace{2\sqrt{n}(1-\delta^2)^n}_{\to 0} \text{ on } (-\infty, -\delta) \cup (\delta, \infty).$$

Therefore,

$$\int_{-\infty}^{-\delta} |f_n(x)|\,dx = \int_{-1}^{-\delta} f_n(x)\,dx \leq \int_{-1}^{-\delta} 2\sqrt{n}(1-\delta^2)\,dx \leq 2\sqrt{n}(1-\delta^2)^n,$$

which goes to zero, as $n \to \infty$. The proof is the same for $\int_{\delta}^{\infty} |f_n(x)|\, dx$.

Note (Bounding strategy used above)**.** This is a standard strategy while analyzing integrals; when we wish to bound an integral (with a non-negative integrand), it is always a good idea to ask ourselves over what interval we know the function to be large/small. This also illustrates that it is always worth asking where you need finesse with a bound, and where a crude bound will do the exact same thing.

Theorem: A

Let $\{f_n\}$ be an approximate identity. Let $g : \mathbb{R} \to \mathbb{C}$ be uniformly bounded and uniformly continuous. Then, $f_n * g \to g$ uniformly on $\mathbb{R}$.

Notation 4. From now onward, theorems labelled in capital roman letters will be no-name theorems that are not stated in Baby Rudin.

Proof. Select M_1 such that $\int_{-\infty}^{\infty} |f_n| \le M_1$ for all n, and M_2 such that $|g(x)| \le M_2$ for all $x \in R$.

Given $\varepsilon > 0$, there exists $\delta > 0$ such that for all $x, y \in \mathbb{R}$ having $|x - y| < \delta$, we have $|g(x) - g(y)| < \varepsilon$.

For $x \in \mathbb{R}$, we have

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

The second integral on the RHS converges because $g(x)$ is a fixed complex number, and multiplying by a complex does not affect convergence (exercise). However, the third integral is a bit more tricky. While $f_n(x)$ is an integrable function, multiplying it by a bounded function can be troublesome: just consider the series analogue where the sum of $\frac{(-1)^n}{n}$ is convergent, but its absolute value isn't. However, we are in luck here, since $f_n(x)$ is absolutely convergent, and thus, we quickly bound it as follows to confirm convergence:

$$\int_{-\infty}^{\infty} |f_n(t) g(x-t)|\, dt \le \int_{-\infty}^{\infty} M_2 |f_n(t)|\, dt \le M_1 M_2.$$

Therefore, by applying theorem 6.12, which also holds for absolute values of functions (exercise), we get

$$\begin{aligned}
|f_n * g(x) - g(x)| &= \left| \int_{-\infty}^{\infty} f_n(t) \left(g(x-t) - g(x) \right)\, dt \right| \\
&\le \int_{-\infty}^{\infty} \left| f_n(t) \left(g(x-t) - g(x) \right) \right| dt \\
&= \int_{-\infty}^{-\delta/2} \left| f_n(t) \left(g(x-t) - g(x) \right) \right| dt + \int_{-\delta/2}^{\delta/2} \left| f_n(t) \left(g(x-t) - g(x) \right) \right| dt \\
&\quad + \int_{\delta/2}^{\infty} \left| f_n(t) \left(g(x-t) - g(x) \right) \right| dt \\
&\le 2M_2 \int_{-\infty}^{-\delta/2} |f_n(t)|\, dt + \varepsilon \int_{-\delta/2}^{\delta/2} |f_n(t)|\, dt + 2M_2 \int_{\delta/2}^{\infty} |f_n(t)|\, dt \\
&\le 2M_2 \left(\int_{-\infty}^{-\delta/2} |f_n(t)|\, dt + \varepsilon_1 \int_{-\delta/2}^{\delta/2} |f_n(t)|\, dt + 2M_2 \int_{\delta/2}^{\infty} |f_n(t)|\, dt \right) + \varepsilon_1 \int_{-\infty}^{\infty} |f_n(t)|\, dt.
\end{aligned}$$

Select N large enough such that for all $n \ge N$, we have

$$|f_n * g(x) - g(x)| \le \varepsilon_1 (4M_2 + M_1) := \varepsilon.$$

$\square$

Note (Convergence of integrals)**.** Note that we have to be careful about the convergence of integrals in the proof above. We cited many Rudin theorems that we have proved for proper integrals, but we are using them for improper integrals; they are true, but using them without proving them is obviously, pedagogically, a lacking choice. However, we did this since it is not worth proving them all, since they are fairly simple exercises, and acknowledging that this could be an issue is a second best thing.

Note (History of approximate identities)**.** Approximate identities didn't really come about as these set of axioms; instead people used families of functions which had these properties to do proofs, and realized that they are doing the same proof.

Quotes of the day: Dr. Joshua Zahl 03/06/2024

No quotes today :(

Definition: Compact support

Let $(\mathcal{X}, d)$ be a metric space ($\mathcal{X} = \mathbb{R}$), and $f : \mathcal{X} \to \mathbb{C}$ (or $\mathcal{X} \to \mathbb{R}$).

We say that f has ***compact support*** if there is a compact set $\mathcal{K}$ such that $f(x) = 0$ for all $x \in \mathcal{X} \setminus \mathcal{K}$.

Example 9. Looking at this in the case of $\mathcal{X} = \mathbb{R}$, we see that $f : \mathbb{R} \to \mathbb{C}$ has compact support iff there exists $R \in \mathbb{R}$ such that $f(x) = 0$ for all $x \notin [-R, R]$.

Example 10. Recall the convolution examples that we looked at; we defined

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

Here, $f(x) = 0$ and $g(x) = 0$ if $x \notin [-R, R]$ for some R.

Lemma

Let $f : \mathbb{R} \to \mathbb{C}$ or $f : \mathbb{R} \to \mathbb{R}$ be compactly supported and integrable. Let $q : \mathbb{R} \to \mathbb{C}$ be a polynomial function; then $f * q$ is a polynomial function. Additionally, *if* $f : \mathbb{R} \to \mathbb{R}$, and q has real co-efficients, then $f * q$ has real co-efficients.

Proof. Write $q(x) := \sum_{k=0}^{n} a_k x^k$, where $x^0 = 1$. Hence, we have

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

where we used the binomial theorem. Now, by Baby Rudin theorem 6.12, we have

$$f * q(x) = \sum_{k=0}^{n}\sum_{l=0}^{k}\left(\underbrace{\int_{-R}^{R} f(t)a_k\binom{k}{l}(-t)^{k-l}x^l\,dt}_{I}\right),$$

where $I \in \mathbb{C}$; $I \in \mathbb{R}$ if $f : \mathbb{R} \to \mathbb{R}$ and all $a_k \in \mathbb{R}$. $\square$

Theorem: Weierstraß approximation theorem (Baby Rudin 7.26)

Let $f : [a,b] \to \mathbb{C}$ or $f : [a,b] \to \mathbb{R}$ be continuous. Then, there exists a sequence of polynomial functions $\{p_n\}$ such that $p_n \to f$ uniformly on $[a,b]$.

If $f : [a,b] \to \mathbb{R}$, then $\{p_n\}$ can be chosen to havea real co-efficients.

Proof step-1 and step-2. We want to reduce the statement of theorem 7.26 to a special case: the interval $[a,b] = [0,1]$, $f(0) = 0$, and $f(1) = 0$. Suppose the theorem is true for such functions; let $g : [a,b] \to \mathbb{C}$ be continuous. Let $f_1(x) := g\left(a + x(b-a)\right)$, and $f_2(x) := f_1(x) - f_1(0)(1-x) - f_1(1)x$. Hence, if $q_n \to f_2$ uniformly, let $x' := \dfrac{x-a}{b-a}$; so we have

$$p_n(x) = q_n(x') + f_1(0)(1-x') + f_1(1)x'.$$

In conclusion, it suffices to prove theorem 7.26 for $f : [0,1] \to \mathbb{C}$, with $f(0) = f(1) = 0$. We will extend $f : \mathbb{R} \to \mathbb{C}$ by setting $f(x) = 0$ for $x \notin [0,1]$. This function is uniformly continuous and bounded, by theorem A, if $\{\tilde{q}_n\}$ is an approximate identity, then $\tilde{q}_n * f \to f$ uniformly. Here, we let

$$\tilde{q}_n(x) = \begin{cases} c_n(1-x^2)^n & -1 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}, \quad q_n(x) = c_n(1-x^2)^n.$$

□

For the last step of the proof, we have to show $\tilde{q}_n * f = q_n * f$.

Quotes of the day: Dr. Joshua Zahl 03/08/2024

"I think this looks cooler." - when asked why he only drew zig-zags on a function when some simple lines would've also worked.

Theorem: B

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

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

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

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

□

5.2 Stone's generalization of the Weierstraß approximation theorem

Definition: Algebra

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

(a) $f + g \in \mathscr{A}$.

(b) $f \cdot g \in \mathscr{A}$.

(c) $cf \in \mathscr{A}$.

Example 11. A few examples of algebras are:

(a) $\mathscr{A}$: polynomial functions $f : \mathbb{R} \to \mathbb{C}$.

(b) $\mathscr{A}$: $\mathscr{C}(\mathbb{R})$, which are the bounded continuous functions $f : \mathbb{R} \to \mathbb{C}$.

(c) $\mathscr{A}$: trigonometric polynomial functions, which are polynomials of the form

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

(d) $\mathscr{A}$: symmetric polynomial functions.

(e) $\mathscr{A}$: piecewise polynomial functions.

(f) $\mathscr{A}$: functions of the form

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

(g) $\mathscr{A}$: functions of the form

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

(h) $\mathscr{A}$: holomorphic functions over $\mathbb{C}$ (or over simply connected subsets of $\mathbb{C}$).

Definition: Uniformly closed

We say $\mathscr{A}$ is ***uniformly closed*** if: for all uniformly convergent sequences $\{f_n\} \subseteq \mathscr{A}$, we have $\lim f_n \in \mathscr{A}$.

Definition: Uniform closure

Let $\mathscr{A}$ be an algebra, and

$$\begin{aligned}\mathscr{B} :=& \{f : \mathcal{E} \to \mathbb{C} : \text{there exist } \{f_n\} \subseteq \mathscr{A} \text{ such that } f_n \rightrightarrows f\} \\ =& \text{“Set of limit points of uniformly convergent sequences in” } \mathscr{A}.\end{aligned}$$

$\mathscr{B}$ is called the uniform closure of $\mathscr{A}$, which we will denote by $\mathrm{Cl}_{\mathrm{u}}(\mathscr{A})$.

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

Notation 5 (Double right arrows)**.** We acknowledge the introduction of new notation $f_n \rightrightarrows f$, which is defined to mean “f_n converges to f uniformly”.

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

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

Quotes of the day: Dr. Joshua Zahl 03/11/2024

"What does obstruction mean? I looked it up in the dictionary before coming to class." - On the idea for Stone-Weierstraß having an obstruction.

Theorem: Baby Rudin 7.29

Let $\mathscr{A}$ be an algebra of bounded functions. Then, the uniform closure $\mathrm{Cl}_\mathrm{u}(\mathscr{A})$ is a uniformly closed algebra.

Proof. Recall $(\mathcal{X}, d)$ is a metric space of bounded functions $\mathcal{E} \to \mathbb{C}$ (or $\mathcal{E} \to \mathbb{R}$), $\mathrm{Cl}_\mathrm{u}(\mathscr{A})$ is the closure of $\mathscr{A}$ in $(\mathcal{X}, d)$, and $\mathrm{Cl}_\mathrm{u}(\mathscr{A})$ is obviously closed. Hence, we have shown that $\mathrm{Cl}_\mathrm{u}(\mathscr{A})$ is uniformly closed. We now verify that this is an algebra.

Let $f, g \in \mathrm{Cl}_\mathrm{u}(\mathscr{A})$, $c \in \mathbb{C}$ (or $c \in \mathbb{R}$). Let $\{f_n\} \subseteq \mathscr{A}$, such that $f_n \to f$ uniformly, $\{g_n\} \subseteq \mathscr{A}$, such that $g_n \to g$ uniformly. All we need to show is that

$$\begin{aligned} f_n + g_n &\to f + g \text{ uniformly} \\ f_n \cdot g_n &\to f \cdot g \text{ uniformly} \\ cf_n &\to cf \text{ uniformly.} \end{aligned}$$

These are straightforward and left as an exercise. Note that your proof here should use the fact that these functions are uniformly bounded, because otherwise we can run into trouble; consider the following example: For $\mathcal{E} = \mathbb{R}$, $f_n(x) = x$, $g_n(x) = \frac{1}{n}$, we have $f_n(x)g_n(x) = \frac{x}{n}$. □

5.3 Idea for Stone-Weierstraß

Let $\mathcal{K}$ be a compact set, $\mathscr{A}$ an (real) algebra of continuous function $f : \mathcal{K} \to \mathbb{R}$. Then, $\mathrm{Cl}_\mathrm{u}(\mathscr{A})$ is the algebra of *all* continuous functions $f : \mathcal{K} \to \mathbb{R}$. This is true *unless* there is an *obvious* obstruction; spoiler, there is. We will now acknowledge and destroy these obstructions.

Definition: Separates points

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

Example 12. Connsider $\mathcal{E} = [-1, 1]$ or $\mathcal{E} = \mathbb{R}$. If $\mathscr{A}$ is all polynomial functions, then we are fine, but if $\mathscr{A}$ is the set of all even polynomial functions, then this does not separate points, since $f(x) = f(-x)$ for all $x \in \mathcal{E}$.

Failure to separate points is an obstruction to Stone-Weierstraß .

Definition: Vanishes at no point of $\mathcal{E}$

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

Example 13. The previous example $\mathcal{E} = [-1, 1]$ or $\mathcal{E} = \mathbb{R}$, $\mathscr{A}$ is all polynomial functions; this works just fine once again. However, now if $\mathscr{A}$ is the set of all odd polynomial functions, then we infer from $f(-x) = -f(x)$ for all $x \in \mathcal{E}$, that $f(0) = 0$. So this does not satisfy the definition.

Vanishing at a point of $\mathcal{E}$ is an obstruction.

Digression. If $\mathscr{A}$ is the set of all odd polynomials, it is clearly not an algebra, so what is the smallest algebra that contains all odd polynomials? This is the ideal generated by x, (x) over the ring of polynomials $\mathbb{C}[x]$ (or $\mathbb{R}[x]$). A natural question to ask ourselves now in this context is what is the uniform closure of this? This is not something we have answered so far, but is something worth thinking about.

Having acknowledged these obstructions, we can now theorise:

Theorem: Stone-Weierstraß theorem for reals (Baby Rudin 7.32)

Let $\mathcal{K}$ be a compact metric space, $\mathscr{A}$ an algebra of continuous functions $f : \mathcal{K} \to \mathbb{R}$. Suppose that $\mathscr{A}$ separates points and vanishes at no point of $\mathcal{K}$. Then, the uniform closure $\mathrm{Cl}_{\mathrm{u}}(\mathscr{A})$ is the algebra of all continuous functions $f : \mathcal{K} \to \mathbb{R}$.

The proof has 3 main steps, that we will deal with as lemmas.

Lemma 1. Let $\mathcal{K}$ and $\mathscr{A}$ as above. If $f \in \mathrm{Cl}_{\mathrm{u}}(\mathscr{A})$, then $|f| \in \mathrm{Cl}_{\mathrm{u}}(\mathscr{A})$.

Proof. Let $f \in \mathrm{Cl}_{\mathrm{u}}(\mathscr{A})$, $M := \sup\{|f(x)| : x \in \mathcal{K}\}$, $M < \infty$. Let $\varepsilon > 0$; we wish to find $g\,\mathrm{Cl}_{\mathrm{u}}(\mathscr{A})$ such that

$$\sup_{x \in \mathcal{K}} \big||f(x)| - g(x)\big| < \varepsilon,$$

which would tell us $|f| \in \mathrm{Cl}_{\mathrm{u}}(\mathscr{A})$.

By the Weierstraß approximation theorem (or an explicit computation), there exists a polynomial function q such that $|q(y)-|y|| < \frac{\varepsilon}{2}$ for all $y \in [-M, M]$.

Let $p(y) = q(y) - q(0) \Rightarrow p(0) = 0$. Hence,

$$\begin{aligned}\big|p(y) - |y|\big| =& \big|q(y) - q(0) - |y| - 0\big| \\ \leq& \big|q(y) - |y|\big| + \big|q(0) - 0\big| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon, \quad \text{for all } y \in [-M, M].\end{aligned}$$

Write $p(x) = \sum_{k=1}^{n} c_k x^k$, and since $\mathrm{Cl}_{\mathrm{u}}(\mathscr{A})$ is an algebra, we define

$$p(f) := \underbrace{\sum_{k=1}^{n} c_k f^k}_{g} \in \mathop{\mathrm{Cl}}_{\mathrm{u}}(\mathscr{A}).$$

For $x \in \mathcal{K}$,

$$\begin{aligned}\big|g(x) - |f(x)|\big| =& \big|p(f)(x) - |f(x)|\big| \\ =& \big|p(f(x)) - |f(x)|\big| < \varepsilon.\end{aligned}$$

□

Note (Usage of Weierstraß approximation in the proof)**.** Recall that we said that Stone-Weierstraß would imply the Weierstraß approximation, but then using the approximation theorem in the proof seems a bit circular. It is not, because we are using the approximation theorem to approximate a very specific function by a polynomial function, which can also be achieved by an explicit computation; we are just making our life easier.

Quotes of the day: Dr. Joshua Zahl 03/13/2024

"Man...just alphabet salad today!" - on mixing up some letters.

Lemma 2. Let $\mathcal{K}$ and $\mathscr{A}$ as before. Let $f_1, \ldots, f_n \in \mathscr{A}$; then $\max\{f_1, \ldots, f_n\} \in \mathrm{Cl_u}(\mathscr{A})$, and $\min\{f_1, \ldots, f_n\} \in \mathrm{Cl_u}(\mathscr{A})$.

Proof. We show this by induction on n.

The case $n = 1$ is trivial. Assume that $g := \max\{f_1, \ldots, f_k\} \in \mathrm{Cl_u}(\mathscr{A})$. Hence, for $f_1, \ldots, f_{k+1} \in \mathrm{Cl_u}(\mathscr{A})$, we have

$$\max\{f_1, \ldots, f_{k+1}\} = \max\{g, f_{k+1}\},$$

where both $g, f_{k+1} \in \mathrm{Cl_u}(\mathscr{A})$. Thus,

$$\max\{f, g\} = \frac{1}{2}(f + g) + \frac{1}{2}|f - g| \in \underset{\mathrm{u}}{\mathrm{Cl}}(\mathscr{A}),$$

since $f + g \in \mathrm{Cl_u}(\mathscr{A})$ and $f - g \in \mathrm{Cl_u}(\mathscr{A}) \Rightarrow |f - g| \in \mathrm{Cl_u}(\mathscr{A})$. $\square$

Note. We define the max of finitely many functions to just literally be the function we obtain by picking the max at every point x of their shared domain.

Lemma 3. Let $\mathcal{K}$ and $\mathscr{A}$ as before. Let $x, y \in \mathcal{K}$, $x \neq y$, $c, d \in \mathbb{R}$, then there exists $f \in \mathscr{A}$ such that $f(x) = c$ and $f(y) = d$.

Note. If $x = y$, then this is true only if $c = d$; this is obvious because otherwise we are dealing with multifunctions.

Proof. Since $\mathscr{A}$ *separates points* and *vanishes at no point*, there exists $g, h, k \in \mathscr{A}$ such that

$$g(x) \neq g(y), \quad h(x) \neq 0, \quad k(y) \neq 0.$$

Let

$$u(z) = g(z)k(z) - \underbrace{g(x)}_{\in\mathbb{R}} k(z) = (g(z) - g(x))k(z) \in \mathscr{A}.$$

We might be tempted to say that since we are taking the product of two things that are in the algebra (in the last equality above), the product is in the algebra, but we cannot assert a priori that $g(z) - g(x) \in \mathscr{A}$, since $g(x) \in \mathscr{A}$ is not guaranteed. Hence, we have to show the first equality above.

Similarly, let

$$v(z) = g(z)h(z) - \underbrace{g(y)}_{\in\mathbb{R}} h(z) = (g(z) - g(y))h(z) \in \mathscr{A}.$$

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

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

Hence, $f(x) = c$ and $f(y) = d$.

The case where $x = y$ and $c = d$, the proof is left as an exercise. $\square$

Notation 6. We define $\mathbb{N}_n := \{i \in \mathbb{N} : 1 \leq i \leq n\}$.

Lemma 4. Let $\mathcal{K}$ and $\mathscr{A}$ as before. Let $f : \mathcal{K} \to \mathbb{R}$ be continuous, $x \in \mathcal{K}$, and $\varepsilon > 0$. Then, there exists $g \in \mathscr{A}$ such that

(a) $g(x) = f(x)$.

(b) $g(t) - f(t) \geq -\varepsilon$ for all $t \in \mathcal{K}$.

Proof. For each $y \in \mathcal{K}$, use lemma 3 to find $g_y \in \mathscr{A}$ such that $g_y(x) = f(x)$, $g_y(y) = f(y)$. Since $g_y \in \mathscr{A} \Rightarrow g_y$ is continuous, and f is continuous by our hypothesis, so $g_y - f$ is continuous. Here, $g_y(x) - f(x) = 0$, and $(g_y - f)(y) = g_y(y) - f(y) = 0$. By definition of continuity, there exists an open set $\mathcal{U}_u \subseteq \mathcal{K}$ such that $g_y(t) - f(t) > -\varepsilon$ for all $t \in \mathcal{U}_y$. Note that $\{\mathcal{U}_y\}_{y\in\mathcal{K}}$ is a cover for $\mathcal{K}$, and thus we extract a finite subcover $\{\mathcal{U}_{y_i}\}_{i\in\mathbb{N}_n}$. Let $g := \max\{g_{y_i}\}_{i\in\mathbb{N}_n} \in \mathrm{Cl_u}(\mathscr{A})$ (by lemma 2). We wish to verify that

$$g(x) = \max\{g_{y_i}\}_{i\in\mathbb{N}_n} = \max\{\underbrace{f(x), f(x), \ldots, f(x)}_{n \text{ times}}\} = f(x).$$

For $t \in \mathcal{K}$, t is contained in some $\mathcal{U}_{y_j}$ $(1 \leq j \leq n)$, which tells us that $g(t) \geq g_{y_i}(t)$. Hence,

$$g(t) - f(t) \geq g_{y_j}(t) - f(t) > -\varepsilon.$$

□

Lemma 5. Let $\mathcal{K}$ and $\mathscr{A}$ as before. Let $f : \mathcal{K} \to \mathbb{R}$ be continuous. Then, there exists $h \in \mathrm{Cl_u}(\mathscr{A})$ such that

$$|h(t) - f(t)| \leq \varepsilon \quad \text{for all } t \in \mathcal{K}.$$

Proof. For $x \in \mathcal{K}$, let $g_x \in \mathrm{Cl_u}(\mathscr{A})$ be as in lemma 4 (same ε). Note that in this case $g_x - f$ is continuous, and $(g_x - f)(x) = 0$. Hence, there exists an open set $\mathcal{U}_x$ containing x such that

$$g_x(t) - f(t) < \varepsilon \quad \text{for all } t \in \mathcal{U}_x.$$

Extract a finite subcover $\{\mathcal{U}_{x_i}\}_{i \in \mathbb{N}_m}$. Let $h := \max\{g_{y_i}\}_{i \in \mathbb{N}_m} \in \mathrm{Cl_u}(\mathscr{A})$ (by lemma 2). By the same argument as lemma 4, we have

$$\begin{aligned} h(t) - f(t) <& \varepsilon \quad \text{for all } t \in \mathcal{K} \\ h(t) - f(t) \geq& -\varepsilon \quad \text{for all } t \in \mathcal{K}, \end{aligned}$$

where we the second inequality since it is true for each g_{y_i}. □

Quotes of the day: Dr. Joshua Zahl 03/15/2024

"I hope they're still teaching about holomorphic functions in complex analysis, cause if they're not, what are they talking about?"

Finally, we conclude the proof for Stone-Weierstraß :

Proof of Stone-Weierstraß . Let $f : \mathcal{K} \to \mathbb{R}$ be continuous. For each n, select $g_n \in \mathrm{Cl_u}(\mathscr{A})$ such that

$$\|f - g_n\|_\infty < \frac{1}{2n}.$$

Additionally, select $f_n \in \mathscr{A}$ such that

$$\|f_n - g_n\|_\infty < \frac{1}{2n}.$$

Therefore, we conclude that

$$\|f - f_n\|_\infty < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

i.e., $f_n \to f$ uniformly. □

5.4 Stone-Weierstraß for complex functions

Conjecture 1. The Stone-Weierstraß theorem is true even when $\mathscr{C}_{\mathbb{R}}(\mathcal{K})$ (continuous functions of the form $f : \mathcal{K} \to \mathbb{R}$) is replaced with $\mathscr{C}(\mathcal{K})$ (continuous functions of the form $f : \mathcal{K} \to \mathbb{R}$ or $f : \mathcal{K} \to \mathbb{C}$).

This is in fact false: the following is a counter-example.

Let $\mathcal{K} = S' =$ unit circle. We define this as

$$\begin{aligned} S' :=& \{z \in \mathbb{C} : |z| = 1\} \\ =& \{e^{it} : t \in [0, 2\pi]\}. \end{aligned}$$

Any $f : S' \to \mathbb{C}$ can be represented as $f(z)$, or $f(e^{it})$, where $t \in [0, 2\pi]$ and $f(e^{i0}) = f(e^{i2\pi})$. Let $\mathscr{A}$ be an algebra of polynomial functions in complex co-efficients:

$$f(z) = \sum_{k=0}^{n} c_k z^k, \quad c_k \in \mathbb{C}$$
$$f(e^{it}) = \sum_{k=0}^{n} d_k e^{kit} \quad d_k \in \mathbb{C}.$$

$\mathscr{A}$ separates points, and vanishes at no point. Let $g(z) = z \in \mathscr{C}(\mathcal{K})$. What we do now is inspired by the contour integral from complex analysis, in particular the key fact that the contour integral of a function that is holomorphic over the interior of the contour, is just zero (Residue theorem).

Let $p \in \mathscr{A}$ is a polynomial function, written as $p(z) = \sum_{k=0}^{n} c_k z^k$. We compute

$$\begin{aligned}\int_0^{2\pi} p(e^{it})e^{it}\, dt &= \int_0^{2\pi} \sum_{k=0}^{n} c_k e^{i(k+1)t}\, dt \\ &= \sum_{k=0}^{n} c_k \int_0^{2\pi} e^{i(k+1)t}\, dt \\ &= \sum_{k=0}^{n} c_k \int_0^{2\pi} \left[\cos\left[(k+1)t\right] + i \sin\left[(k+1)t\right]\right]\, dt \\ &= 0.\end{aligned}$$

Now, consider $g(e^{it}) = e^{-it}$. Hence,

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

If there existed $\{p_n\} \subseteq \mathscr{A}$ such that $p_n \to g$ uniformly (on $\mathcal{K}$), by Baby Rudin theorem 7.16, we get

$$\underbrace{\int_0^{2\pi} p_n(e^{it})e^{it}\, dt}_{=0} \rightrightarrows \int_0^{2\pi} g(e^{it})e^{it}\, dt = 2\pi,$$

which is absurd. In conclusion, $g(z) = \overline{z} \notin \mathscr{A}$, where this denotes the complex conjugate function.

While this is not the *only* counter-example, the obstruction that this one suggests is the only one we need to destroy. Recall that functions vanishing at a point was an obstruction, because that was a fact we "couldn't escape", meaning this property would always exist under the operations of the algebra. This is the same problem with the algebra being holomorphic: we can never escape this fact. How do we remedy this? Well, we just saw that the complex conjugation is not a holomorphic function, so if we just include that our algebra cannot be holomorphic.

Definition: Self-adjoint

Let $\mathcal{K}$ be a compact metric space, $\mathscr{A} \subseteq \mathscr{C}(\mathcal{K})$ is an algebra that separates points, and vanishes at no point. We say that $\mathscr{A}$ is ***self-adjoint*** if for $f \in \mathscr{A}$, $\overline{f} \in \mathscr{A}$, where $\overline{f}(z) = \overline{f(z)}$.

We can now prove the theorem.

Quotes of the day: Dr. Joshua Zahl 03/18/2024

"Maybe this will make it into the next version of Rudin."

Proof. Given that $\mathscr{A} \subseteq \mathscr{C}(\mathcal{K})$ (where $\mathcal{K}$ is compact), sucht that $\mathscr{A}$ separates points, vanishes at no point, and is self-adjoint, then $\mathrm{Cl_u}(\mathscr{A})$. We will prove this using Stone-Weierstraß for the reals.

Step-1. Let $\mathscr{A}_{\mathbb{R}}$ is an algebra of real values functions in $\mathscr{A}$, i.e., $\{f \in \mathscr{A} : \mathrm{Im}(f) \equiv 0\}$.

Let $f = u + iv \in \mathscr{A}$, then $u = \mathrm{Re}(f) = \dfrac{1}{2}(f + \overline{f}) \in \mathscr{A}$. Hence, $\mathrm{Re}(f) \in \mathscr{A}_{\mathbb{R}}$. Similarly, $v = \mathrm{Im}(f) = \dfrac{1}{2i}(f - \overline{f}) \in \mathscr{A}$, and thus $\mathrm{Im}(f) \in \mathscr{A}_{\mathbb{R}}$. Note that this clearly does not work if the algebra isn't self adjoint, so this fact is absolutely crucial.

Step-2. Let $x, y \in \mathcal{K}$ $(x \neq y)$. By Lemma 4, there exists $f \in \mathscr{A}$ such that $f(x) = 0$, $f(y) = 1$. Hence, $\mathrm{Re}(f)(x) = 0$ and $\mathrm{Re}(f)(y) = 1$. These are both in $\mathscr{A}_{\mathbb{R}}$, and hence $\mathscr{A}_{\mathbb{R}}$ separates points.

Note (Reliance on real algebras)**.** Proof step-2 does not actually rely on the algebra being real. This is why Rudin does this separately before the proof of either the real or complex version of Stone-Weierstraß . We did it in the proof of the real version of Stone-Weierstraß because we wanted to go step by step, so we undo this unnecessary reliance by acknowledging it here.

Step-3. Let $x \in \mathcal{K}$. Since $\mathscr{A}$ vanishes at no point, there exists $f \in \mathscr{A}$ such that $f(x) \neq 0$. At least one of $\mathrm{Re}(f)(x)$ or $\mathrm{Im}(f)(x)$ is not zero. Essentially, the argument using $f\overline{f} = |f|^2$ also works.

Note (Rudin's way of showing this)**.** Rudin does this in a slightly different way which is worth knowing: because $f(x) = re^{i\theta}$, we have $e^{-i\theta} f(x) = r \in \mathscr{A}$.

Step-4. By Stone-Weierstraß , $\mathrm{Cl_u}(\mathcal{A}_{\mathbb{R}}) = \mathscr{C}_{\mathbb{R}}(\mathcal{K})$, which is the space of continuous, bounded, real-valued functions $f : \mathcal{K} \to \mathbb{R}$.

Step-5. Let $f = u + iv \in \mathscr{C}(\mathcal{K})$, $u, v \in \mathscr{C}_{\mathbb{R}}(\mathcal{K})$. This is because functions from $\mathbb{R}^n \to \mathbb{R}^n$ are continuous only if the function is continuous component wise. The proof for this is given in Baby Rudin, but is also not very involved, so can be treated as an exercise.

Hence, there exists $g, h \in \mathscr{A}_{\mathbb{R}}$ such that $\|u - g\|_\infty < \dfrac{\varepsilon}{2}$ and $\|v - h\|_\infty < \dfrac{\varepsilon}{2}$. We have then that $g, ih \in \mathscr{A}$, so clearly $g + ih \in \mathscr{A}$ and

$$\|f - (g + ih)\|_\infty \leq \|u - g\|_\infty + \|iv - ih\|_\infty < \varepsilon.$$

Uniform limit of continuous functions is also continuous, which gives us the set inclusion the other way.

□

6 Fourier Analysis

Fourier analysis can be described as "glorified linear algebra", but it is different from linear algebra because this is in an infinite-dimensional vector space. So before we get into it, we have to talk a bit about linear algebra.

6.1 Interlude: Linear Algebra

Let $\mathcal{F}$ be a field (we generally only care about $\mathbb{R}$ or $\mathbb{C}$) here. A vector space over $\mathcal{F}$ (called an $\mathcal{F}-$vector space) is a set $\mathcal{V}$ along with two operations:

1. Vector addition: for $u, v \in \mathcal{V}$, vector addition is $u + v$.
2. Scalar multiplication: for $v \in \mathcal{V}$, $\lambda \in \mathcal{F}$, scalar multiplication is λv.

These are essentially maps $\mathcal{V} \times \mathcal{V} \to \mathcal{V}$ and $\mathcal{F} \times \mathcal{V} \to \mathcal{V}$ respectively. These must satisfy the vector space axioms:

1. Associativity of addition: For $u, v, w \in \mathcal{V}$, $u + (v + w) = (u + v) + w$.
2. Commutativity of addition: $u + v = v + u$.
3. There is an additive identity $0 \in \mathcal{V}$ such that $u + 0 = 0 + u = u$.
4. Every element has an additive inverse, i.e., for all $v \in \mathcal{V}$, there exists $u \in \mathcal{V}$ such that $v + u = u + v = 0$. We refer to u as "$-v$", which is the "inverse" of v.

5. Scalar multiplication and field element multiplication is compatible: for $a, b \in \mathcal{F}$ and $v \in \mathcal{V}$, we have $(ab)v = a(bv)$.

6. Scalar multiplication respects the field identity: $1v = v$ where 1 is the multiplicative identity of the field, and $v \in \mathcal{V}$.

7. Distributivity: $a(u+v) = au + av$ and $(a+b)v = av + bv$.

Depending on the size of u, v, w, associativity could break; for example the octonions don't have associativity. This is a very canonically algebraic construction done with the purpose of it being a counterexample (in some sense), so this is bound to be a bit confusing. Additionally, not all operations are commutative, the simplest example being matrix multiplication.

Some examples of vector spaces:

- $\mathbb{R}^d$, $\mathbb{C}^d$ of $\mathbb{R}$ and $\mathbb{C}$ respectively, with operations:

$$\begin{aligned}(x_1, x \ldots, x_n) + (y_1, \ldots, y_n) =& (x_1 + y_1, \ldots, x_n + y_n)\\ a(x_1, \ldots, x_n) =& (ax_1, \ldots, ax_n).\end{aligned}$$

  Verification of the axioms is left as an exercise.

- Sequences $\{x_n\}_{n \in \mathbb{N}}$, where $x_i \in \mathbb{R}$ or $\mathbb{C}$. Component wise operations as before; this is just as infinite dimensional version of the on we had before.

- Sequences $\{x_n\}_{n \in \mathbb{N}}$ where $x_n = 0$ for all $n \geq N$. "Sequences that are eventually null."

- Sequences $\{x_n\}_{n \in \mathbb{N}}$ such that $\sum_{n \in \mathbb{N}} |x_n| < \infty$. To generalize, when $1 \leq p < \infty$,

$$\left(\sum_{n \in \mathbb{N}} |x_n|^p\right)^{1/p} < \infty.$$

  Addition works fine, since $|x_n + y_n|^p \leq 2^{p-1}(|x_n|^p + |y_n|^p)$.

- $\mathcal{V} = \mathscr{C}_{\mathbb{R}}(\mathbb{R})$, $\mathscr{C}_{\mathbb{R}}(\mathcal{K})$, or $\mathscr{C}(\mathcal{K})$.

- $\mathcal{R}([a, b])$.

Quotes of the day: Dr. Joshua Zahl 03/20/2024

No quotes today :(

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

Definition: Norm

A ***norm*** is a function $\|\cdot\| : \mathcal{V} \to \mathbb{R}$ that satisfies the following properties:

(a) $\|x + y\| \leq \|x\| + \|y\|$, for all $x, y \in \mathcal{V}$ (triangle inequality.)

(b) $\|ax\| = a\,\|x\|$, where $a \in \mathcal{F}$, $x \in \mathcal{V}$.

(c) $\|x\| \geq 0$, and $\|x\| = 0$ iff $x = 0$.

Example 14. $\mathcal{V} = \mathbb{R}^d$ or $\mathbb{C}^d$, where we define for all $1 \leq p < \infty$ real:

$$\|x\|_p = \left(\sum_{j=1}^{n} |x_i|^p\right)^{1/p}.$$

Example 15. For all $1 \leq p < \infty$ real, for the vector space

$$V = \left\{ a = (a_n)_{n\in\mathbb{N}\ (\mathbb{Z})} : \left(\sum_{n\in\mathbb{N}\ (\mathbb{Z})} |a_n|^p \right)^{1/p} < \infty \right\} = \ell^p(\mathbb{N})\ (\ell^p(\mathbb{Z})),$$

we define

$$\|a\|_p = \left(\sum_{n\in\mathbb{N}} |a_n|^p \right)^{1/p}.$$

Note (Indexing using $\mathbb{Z}$)**.** We index can in the integers, because this is a countable set. The reason we might prefer this over the natural numbers is because sometimes it is notationally more natural to use the integers, and if we restrict indexing the natural numbers, then everything will be indexed in bijections to the natural numbers, which makes everything clunky.

Example 16. For $\mathcal{V} = \mathscr{C}([0,1])$, we define

$$\|f\| = \int_0^1 |f(t)|\, dt.$$

Example 17. For $\mathcal{V} = \mathscr{C}([0,1])$, for all $1 \leq p < \infty$ real, we define

$$\|f\|_{L^p} = \left(\int_0^1 |f(t)|^p\, dt \right)^{1/p}.$$

A special case of this which we showed on the homework is

$$\|f\|_{L^\infty} = \|f\|_\infty = \sup_{x\in[0,1]} |f(x)|.$$

Example 18. For $\mathcal{V} = \mathscr{R}([0,1])$, for all $1 \leq p < \infty$ real, we define

$$\|f\|_{L^p} = \left(\int_0^1 |f(t)|^p\, dt \right)^{1/p}.$$

However this time, there is an issue here: since we extended to Riemann integrable functions, we can have something like

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

which has $\|f\|_{L^p} = 0$ for all $1 \leq p < \infty$, but $f(x) \neq 0$. This is problematic, and we deal with this in two possible ways: we could restrict the functions to only continuous functions – which is what Rudin does, and what we will do in some cases – but otherwise we do this in another manner, which builds towards the Lebesgue integral.

Note ($p = \infty$ case.)**.** In the case $p = \infty$, we look at the supremum norm, under which this problem will not exist, since if the supremum of a function is zero on a set, it better be identically zero (under absolute values). Hence, we acknowledge that we really only need the requirement for the function to be bounded for the supremum norm to be well defined, and none of what follows is required for it.

6.2 Fix for example 18

We define an equivalence relation $\sim$ on $\mathcal{R}[0,1]$ as follows:

Definition

We define $f \sim g$ iff

$$\int_0^1 |f(t) - g(t)|\, dt = 0.$$

It is not hard to verify that this is an equivalence relation, and is left as an exercise.

Let $\mathcal{V} = \mathcal{R}([0,1])/\sim$. If $\sim$ is an equivalence relation on $\mathcal{S}$, for $x \in \mathcal{S}$,

$$[x] = \{y \in S : y \sim x\}.$$

Thus, $\mathcal{V} = \{[f] : f \in \mathcal{R}([0,1])\}$. An example is $[0] = \left\{ f \in \mathcal{R}([0,1]) : \int_0^1 |f(t)|\,dt = 0 \right\}$. Now, we can define

$$\|[f]\|_p = \left(\int_0^1 |f(t)|^p \right)^{1/p}.$$

The final issue now is to verify whether this is well defined: verify that if $[f] = [g]$, then

$$\left(\int_0^1 |f(t)|^p \right)^{1/p} = \left(\int_0^1 |g(t)|^p \right)^{1/p}.$$

This is a standard verification that this is not a multi-function, and left as an exercise. Note that this was a homework problem for us, so while this is simple, it is not necessarily easy. Also, verifying it for any p (we had $p = 2$) works, because the proof will be the same for every p.

6.3 Setup for Fourier analysis

Definition: Normed vector space

A pair $(\mathcal{V}, \|\cdot\|)$ is called a ***normed vector space***. This induces a metric $d(x,y) = \|x - y\|$.

If $(\mathcal{V}, d)$ is complete, then we call $(\mathcal{V}, \|\cdot\|)$ a ***Banach space***.

Example 19. $(\mathbb{R}^d, \|\cdot\|_p)$ is a Banach space.

Example 20. $(\mathscr{C}([0,1]), \|\cdot\|_{L^p})$ for $p < \infty$ is *not* a Banach space.

Definition: Inner product space

Let $\mathcal{V}$ be a vector space, $\langle \cdot,\cdot \rangle$ an inner product. Then, $(\mathcal{V}, \langle \cdot,\cdot \rangle)$ is called an ***inner product space***.

A complete inner product space is called a ***Hilbert space***.

Note. An inner product induces a norm:

$$\|x\| = (\langle x,x \rangle)^{1/2}.$$

Example 21. For $\mathcal{V} = \ell^2(\mathbb{N})$ (or $\ell^2(\mathbb{Z})$), for $a = (a_n)_{n \in \mathbb{N}}$ and $b = (b_n)_{n \in \mathbb{N}}$,

$$\langle a,b \rangle = \sum_{n \in \mathbb{N}} a_n \overline{b_n},$$

where $\bar{\cdot}$ is the complex conjugate as before. Here, we note that $\langle a,b \rangle = \overline{\langle b,a \rangle}$.

Example 22. For $\mathcal{V} = \mathscr{C}([0,1])$,

$$\langle f,g \rangle = \int_0^1 f(t)\overline{g(t)}\,dt.$$

Similarly, for $\mathcal{V} = \mathscr{C}([-\pi,\pi])$,

$$\langle f,g \rangle = \int_{-\pi}^{\pi} f(t)\overline{g(t)}\,dt.$$

Inner product spaces are of particular importance to us in Fourier analysis.

Quotes of the day: Dr. Joshua Zahl 03/22/2024

"The truth is a little bit unpleasant, so we ignore it." - on abuse of notation.

Before we move no we note that the only vector spaces that we will talk about will be $\mathscr{C}([0,1])$ and $\mathcal{R}[0,1]$ (or $\mathcal{R}[a,b]$).

Notation 7 (Abuse of notation)**.** Whenever we talk about $\mathcal{R}[0,1]$ (or $\mathcal{R}[a,b]$), the inner product should look something like

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

because the functions are equivalence classes. However, this is clunky, so we will just write

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

This is obviously abuse of notation, but this makes our life easier, and it is pretty much almost always obvious from the context which vector space we are talking about.

Definition: Orthonormal system

If $\{\varphi_n\}_{n\in\mathcal{C}}$ (where $\mathcal{C}$ is some countable set) is a sequence of Riemann integrable functions on $[a,b]$ that are orthonormal, i.e.,

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

then we call $\{\varphi_n\}$ a "system of orthonormal functions" or an "***orthonormal system***" on $[a,b]$.

Note. We used a countable set $\mathcal{C}$ instead of something familiar like $\mathbb{N}$ or $\mathbb{Z}$ because it might be convenient to switch between them, so leaving it as ambiguous is more general. Now onward we will just omit the set over which it is indexed, because it will always either be $\mathbb{N}$ or $\mathbb{Z}$.

Note (Notation overlap with Linear algebra)**.** Joseph Fourier intially invented linear algebra as a way of studying the heat equation. While this certainly was not the birth of linear algebra, it predated linear algebra in some sense that after this was when people started using linear algebra in other fields of math, and it evolved to be what it is today. Hence, there is some notational overlap with linear algebra.

Example 23. Some examples of orthonormal systems on $[0,1]$ are:

- Consider $\{e^{2\pi inx}\}_{n\in\mathbb{Z}}$ such that

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

- For the family $\{\sin(n\pi x)\}_{n\in\mathbb{N}} \cup \{\cos(n\pi x)\}_{n\in\mathbb{N}}$. We get something like the following picture

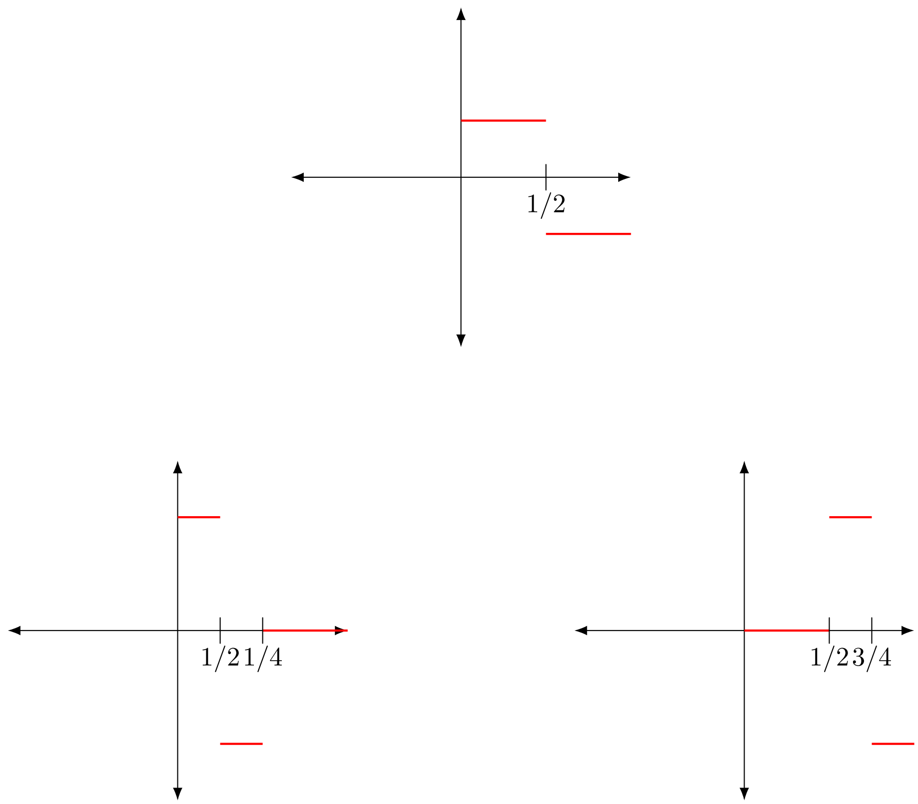

Figure 8: Figure showing how wavelets “evolve”. The top wavelet is called the “mother wavelet”, and the “daughter wavelets” are the smaller constituents we get from it. The smaller ones can be divided into smaller waveletes in the same manner.

Wavelets are like loading images on very slow internet: they are blurry at the start, but as more show up, it becomes clearer.

Definition: Fourier coefficient

If $\{\varphi_n\}$ is an orthnormal system on $[a, b]$ and if $c_n = \langle f, \varphi_n \rangle$, then we call c_n the n^{th} ***Fourier coefficient*** of f relative to $\{\varphi_n\}$.

If $[a, b] = [0, 1]$ and $\varphi_n = e^{2\pi i n x}$, we write $c_n = \hat{f}(n)$.

Food for thought 12. What is the geometric interpretation of $c_n \varphi_n(x)$?

Answer. Consider an orthonormal system $\{\varphi_n\}$, $v \in \mathcal{V}$, let us look at $\langle v, \varphi_n \rangle \varphi_n$ in a few situations:

- Let $\mathcal{V} = \mathbb{R}^2$, and φ_1 a unit vector in $\mathbb{R}^2$:

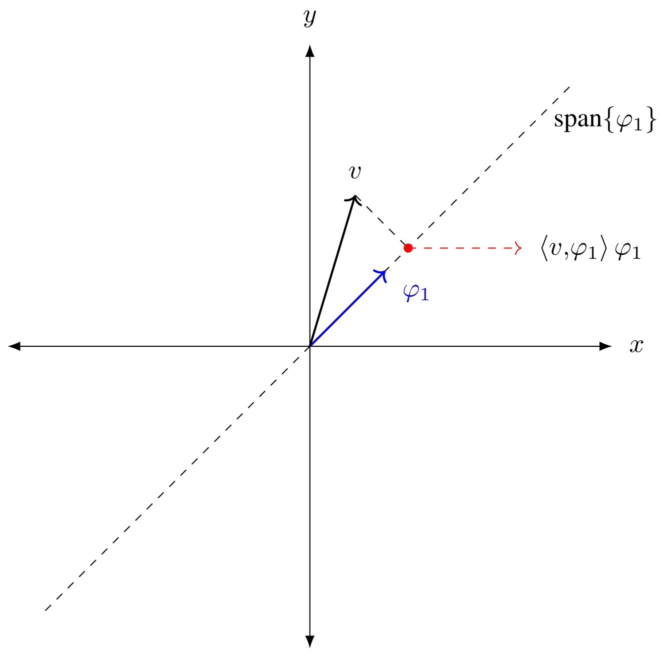

Figure 9: Visual representation of the example.

It is the unique point on the span of φ_1, which minimizes the distance between the vectors.

- Let $\mathcal{V} = \mathbb{R}^3$, and let $\{\varphi_1, \varphi_2\}$ span the xy-plane.

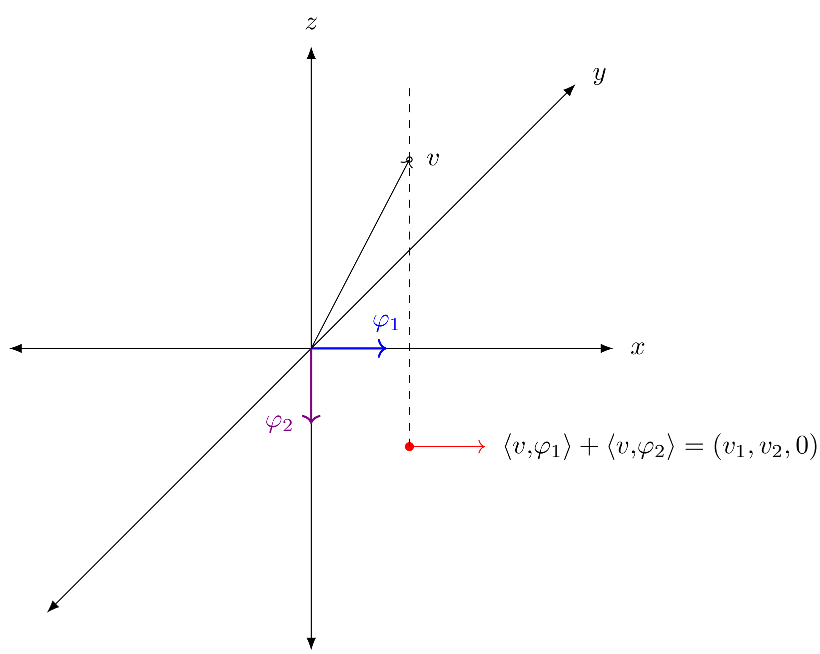

Figure 10: Visual representation of the example.

In this case, we are minimizing the distance from the vector to the plane xy.

Note. In the previous example, note that while the point which minimizes the distance is unique, the points mapping to this point unedr the projection map are many: all the points in $\mathbb{R}^3$ that project onto the same point is the line above that point; this is called the fiber of that point under the pre-image of the projection map.

6.4 Fourier series

Definition: Fourier series

If $\{c_n\}_{n\in\mathbb{N}}$ are the Fourier co-efficients of $f \in \mathcal{V} = \mathcal{R}[a,b]/\sim$ relative to $\{\varphi_n\}_{n\in\mathbb{N}}$, then $\sum_{n=1}^{\infty} c_n\varphi_n$ is called the ***Fourier series*** of f relative to $\{\varphi_n\}$, and we write $f \sim \sum_{n=1}^{\infty} c_n\varphi_n$.

Note. The $\sim$ in this definition has nothing to do with the equivalence relation defined before. This is just an unfortunate coincidence because of typewriters not having enough math symbols. In this case, this just means that $\{c_n\}_{n\in\mathbb{N}}$ are the fourier coefficients of $\{\varphi_n\}_{n\in\mathbb{N}}$ with respect to f.

Theorem: Baby Rudin 8.11

Let $\{\varphi_n\}_{n\in\mathbb{N}}$ be an orthonormal system on $[a,b]$, and let $f \in \mathcal{R}[a,b]$, with Fourier coefficients $\{c_n\}_{n\in\mathbb{N}}$. Let

$$S_n(x) = \sum_{j=1}^{n} c_j\varphi_j(x);$$

this is called the n^{th} partial Fourier sum of f.

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

$$\|f - S_n\|_{L^2} \le \|f - t_n\|_{L^2},$$

with equality iff $s = t$, i.e., $d_j = c_j$ for all $j = 1,\ldots,n$.

Notation 8 (Specifying the norm)**.** Now onward, there will be no subscript on the norm, because it is always clear from context.

Quotes of the day: Dr. Joshua Zahl 03/26/2024

"This is a projection onto a single vector; you could imagine a projection onto the span of multiple vectors if you were better at drawing."

Proof. Note that

$$\begin{aligned}
\langle f, t_n\rangle &= \int_a^b f(x)\overline{t_n(x)}\,dx \\
&= \int_a^b f(x)\sum_{j=1}^{n}\overline{d_j\varphi_j(x)}\,dx \\
&= \int_a^b \sum_{j=1}^{n}\overline{d_j}\,\underbrace{f(x)\overline{\varphi_j(x)}}_{\langle f,\varphi_j\rangle}\,dx \\
&= \sum_{j=1}^{n} c_j\overline{d_j}.
\end{aligned}$$

Similarly, $\langle t_n, f\rangle = \sum_{j=1}^{n} d_j\overline{c_j}$. Now, we make a series of computations:

$$\begin{aligned}
\|t_n\|^2 &= \int_a^b t_n\overline{t_n}\,dx \\
&= \int_a^b \left(\sum_{j=1}^{n} d_j\varphi_j\right)\left(\sum_{k=1}^{n} \overline{d_k\varphi_k}\right)\,dx \\
&= \sum_{j=1}^{n} |d_j|^2,
\end{aligned}$$

and thus,

$$\begin{aligned}
\|f - t_n\|^2 &= \int_a^b |f - t_n|^2\,dx \\
&= \int_a^b |f|^2\,dx - \int_a^b f\overline{t_n}\,dx - \int_a^b \overline{f}t_n\,dx + \int_a^b |t_n|^2\,dx \\
&= \int_a^b |f|^2\,dx - \sum_{j=1}^{n} c_j\overline{d_j} - \sum_{j=1}^{n} d_j\overline{c_j} + \sum_{j=1}^{n} d_j\overline{d_j}\,dx \\
&= \int_a^b |f|^2\,dx - \sum_{j=1}^{n} |c_j|^2 + \sum_{j=1}^{n} |d_j - c_j|^2\,dx \\
&\geq \int_a^b |f|^2\,dx - \sum_{j=1}^{n} |c_j|^2 \quad \text{with equality iff } d_j = c_j \text{ for all } j.
\end{aligned}$$

If $d_j = c_j$ for all j, then

$$\|f - s_n\|^2 = \|f\|^2 + \|s_n\|^2\,. \tag{2}$$

Hence,

$$\|f - t_n\|^2 = \int_a^b |f - t_n|^2\,dx \geq \|f\|^2 - \|s_n\|^2 = \|f - s_n\|^2\,,$$

with equality iff $d_j = c_j$ for all j. □

As a consequence of eq. (2),

$$\sum_{j=1}^{n} |c_n|^2 = \|s_n\|^2 \leq \|f\|^2 \Rightarrow \sum_{j=1}^{\infty} |c_n|^2 \leq \|f\|^2\,. \tag{3}$$

This is called the *Bessel inequality*. We get equality if $f \in \operatorname{span}\{\varphi_n\}_{n=1}^{\infty}$. From now onward, we let $[a, b] = [0, 1]$, $\{\varphi_n\}_{n\in\mathbb{Z}}$ such that $\varphi_n(x) = e^{2\pi i n x} := e_n(x)$.

Definition: L-periodic

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

Example 24. $e_n(x)$ is 1 periodic for all $n \in \mathbb{Z}$.

Let $\mathcal{V} := \{f : \mathbb{R} \to \mathbb{C} : f \text{ is 1 periodic and integrable on } [0, 1]\}/\sim$. On $\mathcal{V}$, we define

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

Note (Another way of thinking about Fourier analysis)**.** Initially we though about these functions as eigenfunctions for our system from the heat equation, or something else modelled using a PDE of a similar sort. But there is a different way to think about Fourier analysis in terms of groups (this is outside the scope of this course):

Consider $f : \mathbb{R} \to \mathbb{C}$, 1-periodic; these have a 1-1 correspondence with functions of the form $f : \underbrace{\mathbb{R}/\mathbb{Z}}_{\mathcal{G}} \to \mathbb{C}$. Note that $\mathcal{G}$ is the set of equivalence classes such that $x \sim y$ iff $x - y \in \mathbb{Z}$. Here, $\mathcal{G}$ is an abelian group having elements $e_n : \mathcal{G} \to \mathbb{C}$ (or $e_n : \mathcal{G} \to$ complex number of magnitude 1). The group multiplication here is $e_n e_m = e_{n+m}(x)$. Most things we do can be written in this more abstract setting, where we have a function that maps from groups to complex numbers, and instead of a basis, we have characters which map from $\mathcal{G}$ to complex numbers of magnitude 1. These characters form the "dual group".

Example 25. We have $\mathbb{R}/\mathbb{Z} \leftrightarrow \mathbb{Z}$ ($\mathbb{Z}$ here is the dual group), but also $\mathbb{Z}_p \leftrightarrow \hat{\mathbb{Z}}_p$ ($\mathbb{Z}_p := \mathbb{Z}/p\mathbb{Z}$) and $\mathbb{R} \leftrightarrow \mathbb{R}$. Finally, for $f : \mathbb{R} \to \mathbb{C}$, we have $\hat{f} = \displaystyle\int e^{2\pi i x}$.

Quotes of the day: Dr. Joshua Zahl 03/27/2024

"The bulk Fourier analysis is starting at a picture and figuring out what I what it tells you."

Definition

The N^{th} partial Fourier series of f is

$$S_N(x) := \sum_{n=-N}^{N} c_n e_n(x).$$

Lemma

For $x \in \mathbb{R}$, $S_N(x) = \displaystyle\int_0^1 f(x-t) D_N(t)\, dt$, where

$$D_N(t) = \sum_{n=-N}^{N} e^{2\pi i n t} = \frac{\sin(2\pi(N+\frac{1}{2})t)}{\sin(\pi t)}.$$

D_N is called the Dirichlet kernel.

If we graph this function, it looks like a sine curve bound by an asymptotic envelope on all sides, and it oscillates very quickly (looking at the Wikipedia page is the best way of visualizing it). It oscillates very quickly.

Proof step-1. Note that

$$\begin{aligned}
D_N(t) =& e^{-2\pi i N t} \sum_{k=0}^{2N} e^{2\pi i n t} \\
=& e^{-2\pi i N t} \left(\frac{e^{2\pi i (2N+1)t} - 1}{e^{2\pi i t} - 1} \right) \frac{e^{-\pi i t}}{e^{-\pi i t}} \\
=& \frac{e^{2\pi i (N+\frac{1}{2})t} - e^{-2\pi i (N+\frac{1}{2})t}}{e^{\pi i t} - e^{-\pi i t}} \\
=& \frac{\sin(2\pi(N+\frac{1}{2})t)}{\sin(\pi t)}.
\end{aligned}$$

□

Proof step-2. Now,

$$\begin{aligned}
S_N(x) &= \sum_{x=-N}^{N} c_n e^{2\pi inx} \\
&= \sum_{x=-N}^{N} \langle f_n, e_n \rangle \, e^{2\pi inx} \\
&= \sum_{x=-N}^{N} \int_0^1 f(t) e^{2\pi in(-t)} e^{2\pi inx} \, dt \\
&= \sum_{x=-N}^{N} \int_0^1 f(t) e^{2\pi in(x-t)} \, dt \\
&= \int_x^{x+1} f(x-s) D_N(s) \, ds \qquad (\dagger) \\
&= \int_0^1 f(x-s) D_N(s) \, ds
\end{aligned}$$

where eq. (†) is where we use the fact that the function is $1-$periodic. □

Definition: Lipschitz continuous at a point

Let $f : \mathbb{R} \to \mathbb{C}$, $x \in \mathbb{R}$. We say f is Lipschitz continuous at x if there exists $\delta > 0$ and $L > 0$ such that for all $y \in \mathbb{R}$ having $|x - y| < \delta$, we have $|f(x) - f(y)| \leq L|x - y|$.

Alternatively for all $t \in \mathbb{R}$ having $|t| < \delta$, $|f(x + t) - f(x)| \leq L|t|$.

Theorem: Baby Rudin 8.14

Let $f : \mathbb{R} \to \mathbb{C}$ be 1-periodic and integrable on $[0, 1]$. Suppose f is Lipschitz continuous at x. Then, $\lim\limits_{N\to\infty} S_N(x) = f(x)$.

Recall we have already shown this result if $\{D_N(x)\}$ was an approximate identity. It is not one, but it has properties that resemble it. The proof should looks similar.

Proof setup. To begin we compute a few things that will make our lives easier:

1. Note that
$$\int_0^1 D_N(t) \, dt = \langle D_N, 1 \rangle = \langle D_N, e_0(x) \rangle = 1. \qquad (\star_1)$$

2. Also,
$$\begin{aligned}
D_N(t) &= \frac{1}{\sin(\pi t)} [\sin(\pi t) \cos(2\pi Nt) + \sin(2\pi Nt) \cos(\pi t)] \\
&= \cos(2\pi Nt) + \cot(\pi t) \sin(2\pi Nt). \qquad (\star_1)
\end{aligned}$$

3. If g is 1-periodic and integrable on $[0, 1]$,
$$\sum_{n=-\infty}^{\infty} |c_n|^2 \leq \int_0^1 |g|^2 \, dx < \infty \Rightarrow c_n \to 0,$$
as $n \to \infty$ or $n \to -\infty$. This is called the *Riemann-Lebesgue* lemma (or maybe a stronger version of this is called that.)

   As a consequence,
$$\lim_{N\to\infty} \langle g, \sin(2\pi Nx) \rangle \to 0. \qquad (\star_3)$$

Therefore,

$$
\begin{aligned}
|\langle g, \sin(2\pi N x)\rangle| =& |\langle g, \sin(2\pi N x)\rangle| \\
\le& \frac{1}{2}|\langle g, e^{2\pi i n x}\rangle| + \frac{1}{2}|\langle g, e^{-2\pi i n x}\rangle| \\
=& \frac{1}{2}(|c_n| + |c_n|) = |c_n| \to 0.
\end{aligned}
$$

□

The intuition behind is that Dirichlet kernel is oscillating very quickly; we say that it is oscillating roughly in a "mean zero way", meaning that it has mean zero almost everywhere.

Proof. Note that

$$
\begin{aligned}
f(x) - S_N(x) =& f(x)\int_0^1 D_N(t)\,dt - \int_0^1 f(x-t)D_N(t)\,dt \\
=& \int_0^1 [f(x) - f(x-t)]D_N(t)\,dt.
\end{aligned}
$$

Using eq. ($\star_1$), we get

$$
f(x) - S_N(x) = \int_0^1 [f(x) - f(x-t)]\cos(2\pi N t)\,dt + \int_0^1 [f(x) - f(x-t)]\cot(\pi t)\sin(2\pi N t)\,dt.
$$

Writing these back as inner products, we get

$$
f(x) - S_N(x) = \underbrace{\langle f(x) - f(x-t), \cos(2\pi N t)\rangle}_{:=A} + \underbrace{\langle (f(x) - f(x-t))\cot(\pi t), \sin(2\pi N t)\rangle}_{:=B}.
$$

Since A is 1-periodic and integrable, we use eq. ($\star_3$) to conclude that $A \to 0$ as $N \to \infty$. B however, is trickier because we cannot be sure that it is bounded, and hence cannot be sure that it is integrable. Hence, we cannot use eq. ($\star_3$). Note that it is clearly 1-periodic.

Let

$$
|h(t)| = \left|(f(x) - f(x-t))\frac{\cos(\pi t)}{\sin(\pi t)}\right|.
$$

We are concerned with the boundedness of this function on the interval $\left[-\frac{1}{2}, \frac{1}{2}\right]$. There exists $\delta > 0$ such that for all $t \in [-\delta, \delta]$,

$$
\begin{aligned}
|h(t)| =& \left|(f(x) - f(x-t))\frac{\cos(\pi t)}{\sin(\pi t)}\right| \\
\le& \frac{L|t||\cos(\pi t)|}{\sin(\pi t)} = L|\cos(\pi t)| \cdot \frac{t}{\sin(\pi t)} \le L,
\end{aligned}
$$

where we get $L \in \mathbb{R}$ from Lipschitz continuity of f. For $t \in \left[-\frac{1}{2}, \frac{1}{2}\right] \setminus [-\delta, \delta]$,

$$
|h(t)| \le 4\sup_{z\in\mathbb{R}}|f(z)|\frac{1}{\pi\delta} < \infty.
$$

Therefore, we conclude that $h(t)$ is bounded.

Since h is continuous and bounded on $\left[-\frac{1}{2}, \frac{1}{2}\right]$. Now, from the Lipschitz continuity of f, there exists $\delta > 0$, there exists M such that $|h(t)| \le M$, for all $|t| \le \delta$. So we construct partitions such that h is "nice" on $\left[-\frac{1}{2}, -\delta\right]$ and $\left[\delta, \frac{1}{2}\right]$; we can make δ small enough here (from Lipschitz continuity as shown), such that we don't have to worry about h misbehaving at 0.

Hence, now we use eq. ($\star_3$) to conclude that $\langle h(t), \sin(Nt)\rangle \to 0$ as $N \to \infty$, giving us the desired result. □

Corollary: Consequences of Baby Rudin 8.14

a) If $f(x) = 0$ for all x in some open interval $\mathcal{I}$, then $S_N(x) \to 0$ for all $x \in \mathcal{I}$. Hence a Fourier series is able to deal with a function that is zero on some interval $\mathcal{I}$ but badly behaved outside of that; this is unlike a Taylor series.

b) If $f(x) = g(x)$ on an interval $\mathcal{I}$, then $S_N(f; x) - S_N(g; x) \to 0$ as $N \to \infty$.

Note that b) here is kind of just a re-skin of a) in the most natural sense.

Theorem: Baby Rudin 8.15

Let $f : \mathbb{R} \to \mathbb{C}$ be 1-periodic and continuous. Given $\varepsilon > 0$, there exists a trigonometric polynomial function $\displaystyle\sum_{n=-N}^{N} a_n e^{2\pi i n x}$ ($a_n \in \mathbb{C}$), such that

$$\sup_{x \in \mathbb{R}} |f(x) - P(x)| < \varepsilon.$$

Proof sketch. We want to apply Stone-Weierstraß. Our metric space is $\mathcal{S}^1 := \{z \in \mathbb{C} : |z| = 1\}$. Define $F : \mathcal{S}^1 \to \mathbb{C}$ such that

$$F(e^{2\pi i t}) = f(t);$$

this is well defined since f is 1-periodic. Define

$$\mathscr{A} := \left\{ \sum_{n=-N}^{N} a_n z^n : a_n \in \mathbb{C}, N \in \mathbb{N} \right\}.$$

Easy to check that it separates no points and vanishes nowhere. Additionally,

$$\overline{\sum_{n=-N}^{N} a_n z^n} = \sum_{n=-N}^{N} \overline{a_n} z^n \in \mathscr{A},$$

so this algebra is also self-adjoint, completing the pre-requisites for complex Stone-Weierstraß. Therefore, we are done. $\square$

Quotes of the day: Dr. Joshua Zahl 04/06/2024

No quotes today :(

Note (Remarks)**.** We note the following things:

1. $S_N(f; x) = \displaystyle\sum_{n=-N}^{N} c_n e^{2\pi i n x}$ might not be the polynomial "found" by Stone-Weierstraß. We might hope to get lucky here so that the $S_N(f; x)$ we get is the polynomial we get, but this is *not* necessarily true.

2. There exists continuous, 1-periodic functions where $S_n(f)$ *does not* converge pointwise to f.

   We can even make it such that $S_N(f)$ does not converge pointwise at countably many points. A natural question then we can ask ourselves is that how big can we make a set such that $S_N(f)$ still fails to converge pointwise to f, for all the points in the set. This has occupied people studying harmonic analysisfrom its inception until today. However, there was important progress on this during the 80s. Even if we write down such a function, it is difficult to prove, but it is worth knowing that there is such a thing.

3. There exist continuous, 1-periodic functions f where $S_n(f) \to f$ pointwise, but not uniformly.

   Things get worse if we expand what it means to be integrable: there exist functions that are not Riemann integrable but are Lebesgue integrable, and there are functions that are not Lebesgue integrable, but not Riemann integrable (I know!). We can think of Lebesgue integrable functions as monotone limits of step functions. We cab find Lebesgue integrable functions that fail to converge to f to almost everywhere.

   If we have that the functions are square integrable, we are working with something much nicer: $S_N \to f$ pointwise almost everywhere. This is why we often work with L^2. For bad L^1 functions, we can have $\int_0^1 |f| < \infty$, but $\int_0^1 |f|^2 = \infty$ (To know more about this read Carleson's theorem).

4. Baby Rudin problem 8.15 describes an explicit sequence of trigonometric polynomial functions that converge uniformly to f:

$$\sigma_N = \frac{S_0 + S_1 + \cdots + S_N}{N+1} \quad \text{(Cesàro mean)},$$

   where S_i is the i^{th} partial Fourier sum.

Food for thought 13 (Importance of Fourier series)**.** If the Cesàro mean is so good, why do we study S_N? More generally, given the shortcomings of S_N, why do we study Fourier series?

Answer. S_N still has some nice properties. Recalling what we talked about with L^2 before: our orthonormal system is, in some way, as close as you can get to this. If we have a trigonometric polynomial function that is close to f in the supremum norm, then it is clearly close to f in the L^2 norm. However, even if S_N is not this trigonometric polynomial, it is closerto f in the L^2 norm: we formulate this as a theorem. □

Theorem C: Plancherel theorem/Parseval-Plancherel identity

Let $f : \mathbb{R} \to \mathbb{C}$ be 1-periodic and integrable on $[0, 1]$. Then $\lim_{N\to\infty} \|f - S_N\|_2 = 0$, i.e., $S_N \to f$ in $(L^2([0, 1]), \|\cdot\|_2)$.

Digression (ℓ^2 and L^2)**.** We can unify ℓ^2 and L^2 since a sequence indexed by real numbers is basically a function from $\mathbb{R}$ to $\mathbb{C}$ (there is some abuse of notation here), and each element gets the same "weight". This is like the Riemann-Stieltjes integral, in terms of weighting; we are alluding to $L^2(\mathcal{X}, \mu)$, where μ is a "measure" on $\mathcal{X}$. All of these definitions fall into this kind of framework, so we can prove all of these results in a unified manner. For example, the Schwartz inequlity in ℓ^2 the same as proving it in $L^2(\mathbb{Z}, \mu)$, because these are the same spaces (take MATH 420).

Proof. Since $\|\cdot\|_2$ is a metric, we have $\|f + g\|_2 \le \|f\|_2 + \|g\|_2$ (Minkowski's identity as well). For $f \in \mathcal{R}[0, 1]$ and $\varepsilon > 0$, there exists $g : [0, 1] \to \mathbb{C}$ continuous so that $\|f - g\|_2 < \varepsilon$ (Baby Rudin problem 6.12).

Given $\varepsilon > 0$, select some continuous $g : [0, 1] \to \mathbb{C}$ such that $\|f - g\|_2 \le \frac{\varepsilon}{3}$. Hence, we have

$$\|S_N(f) - f\|_2 \le \underbrace{\|S_N(f) - S_N(g)\|_2}_{:(A)} + \underbrace{\|S_N(g) - g\|_2}_{:(B)} + \underbrace{\|g - f\|_2}_{<\varepsilon/3}. \tag{$\spadesuit$}$$

Here,

$$A : \|S_N(f) - S_N(g)\|_2 = \|S_N(f - g)\|_2 \le \|f - g\|_2 < \frac{\varepsilon}{3}.$$

For (B), recall that by theorem 8.15, there exists a trigonometric polynomial function p having degree N_0, such that $\|g - p\|_\infty < \varepsilon/3$; hence, $\|g - p\|_2 < \varepsilon/3$. Thus, for all $n > N_0$, by theorem 8.11,

$$\|S_N(g) - g\|_2 \le \|p - g\|_2 < \frac{\varepsilon}{3}.$$

Therefore, resolving these in eq. ($\spadesuit$), we get

$$\|S_N(f) - f\| < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon,$$

concluding the proof. □

Note. A way to think about this proof more abstractly is: this is an $\varepsilon/3$ proof. We have seen this many times, but we are actually trying to understand S_N – a linear operator. We understand this on continuous functions, which is a subspace of L^2, and wish to understand on L^2. What this theorem says specifically is that continuous functions are dense in L^2. This is a trick that is worth knowing: when we want to understand a linear operator, we can do it by controlling it on a dense subspace, so that we can control the error.

The reason is that being a Banach space is very important for Bounded Linear Extension theorem, which says that if a function is bounded on a dense subspace, it is bounded on the whole space. We don't get this for L^2 because it is not a Banach space, but we get something similar, which we explore futher in MATH 420.

Quotes of the day: Dr. Joshua Zahl 04/08/2024

No quotes today :(

Notation 9. Let

$$\hat{f}(n) = \int_0^1 f(x)e^{-2\pi inx}\,dx,$$

where $\hat{f} : \mathbb{Z} \to \mathbb{C}$.

Theorem: Baby Rudin theorem 8.16

Let $f, g : \mathbb{R}^2 \to \mathbb{C}$ be 1-periodic functions that are integrable on $[0, 1]$. Let $\{c_n\}_{n\in\mathbb{Z}}$ and $\{d_n\}_{n\in\mathbb{Z}}$ be their Fourier coefficients with respect to the orthonormal basis $\{e^{2\pi inx}_{n\in\mathbb{Z}}\}$. Then,

$$\int_0^1 f(x)\overline{g(x)}\,dx = \sum_{n\in\mathbb{Z}} c_n\overline{d_n}.$$

We can restate this as $\langle f,g\rangle = \left\langle \hat{f},\hat{g}\right\rangle = \sum_{n\in\mathbb{Z}} \hat{f}(n)\overline{\hat{g}(n)}$, where we first have the inner product on $L^2([0, 1])$ and then the inner product on $\ell^2(\mathbb{Z})$.

Before we prove the theorem, it is worth noting that we have seen a few version of Cauchy-Schwartz before:

$$\left|\sum_{k=1}^{n} a_k\overline{b_k}\right| \le \left(\sum_k |a_k|^2\right)^{1/2} \left(\sum_k |b_k|^2\right)^{1/2},$$

which in terms of the inner product on $\ell^2(\mathbb{Z})$ is $|\langle a,b\rangle| \le (\langle a,a\rangle)^{1/2}(\langle b,b\rangle)^{1/2}$. We also have a version for the inner product on $L^2([0, 1])$: $|\langle f,g\rangle| \le \|f\|_2\,\|g\|_2$, i.e.,

$$\left|\int_0^1 f(x)\overline{g(x)}\,dx\right| \le \left(\int_0^1 |f(x)|^2\,dx\right)^{1/2} \left(\int_0^1 |g(x)|\,dx\right)^{1/2}.$$

This is a special case of the more general Hölder's inequality: $|\langle f,g\rangle| \le \|f\|_p\,\|g\|_p$ for all $1 \le p, q \le \infty$, where $\dfrac{1}{p} + \dfrac{1}{q} = 1$. Written out, this says

$$\left|\int_0^1 f(x)\overline{g(x)}\,dx\right| \le \left(\int_0^1 |f(x)|^p\,dx\right)^{1/p} \left(\int_0^1 |g(x)|^q\,dx\right)^{1/q}.$$

We now do the proof for Parseval's theorem.

Proof. Note that

$$\begin{aligned}
\langle S_N(f), g\rangle &= \int_0^1 \sum_{n=-N}^{N} c_n e^{2\pi i n x}\overline{g(x)}\, dx \\
&= \sum_{n=-N}^{N} c_n \overline{\int_0^1 e^{-2\pi i n x} g(x)\, dx} \\
&= \sum_{n=-N}^{N} c_n \overline{d_n}.
\end{aligned}$$

Using Cauchy Schwartz inequality, we geometric

$$\left|\langle f, g\rangle - \sum_{n=-N}^{N} c_n \overline{d_n}\right| \leq |\,\langle f, g\rangle - \langle S_N(f), g\rangle\,| = |\,\langle f - S_N, g\rangle\,| \leq \underbrace{\|f - S_N(f)\|_2}_{\to 0} \|g\|_2 \,.$$

Hence, we have shown that

$$\lim_{N\to\infty} \sum_{n=-N}^{N} c_n \overline{d_n} = \langle f, g\rangle \,.$$

We are not yet done because we have proven this for the symmetric sum, but this is a very specific order of summing, and we are not able to switch the order of summing yet. To say every sum of this kind converges to this, we need to show absolute convergence. Therefore, once again by Cauchy-Schwartz,

$$\begin{aligned}
\sum_{n=-N}^{N} |c_n \overline{d_n}| &\leq \left(\sum_{n=-N}^{N} |c_n|^2\right)^{1/2} \left(\sum_{n=-N}^{N} |d_n|^2\right)^{1/2} \qquad (\clubsuit) \\
&\leq \|f\|_2 \, \|g\|_2
\end{aligned}$$

where we use Bessel's inequality in eq. ($\clubsuit$) to get the next inequality. $\square$