Baby Analysis*

Salman A. Faris

4 November 2019

Contents

*most of the content of this notes are based on King's College London's 5CCM221A Real Analysis Fall Term 2019, lectured by Prof. Simon Scott and Prof. Alexander Pushnitski.

1 Preliminaries

1.1 A short notational comment

On proof notations

1. We say a proof is *proof-isomorphic* (or just *isomorphic*) if the steps to reach the conclusion of a particular proof is very similar to another proof. Whenever a proof is isomorphic but bears some importance, we will still write down the proof with clear references to which steps we will omit. We say such steps that are omitted as *step-isomorphic*.

2. By ^+A-manipulating or ^+A-manipulation we mean the trick of making the existence of A from nothingness (identity) with respect to the operation $+$. *Remark: Here A is not restricted to be the real number. It can be applied to functions, expressions, little-oh notations etc.*

2 Continuous Functions I: Basics

3 Continuous Functions II: Further properties

3.1 Continuous Functions on Compact Intervals

We focus our attention to continuous functions on **closed and bounded** intervals, otherwise known as **compact intervals**.

Remark 3.1.1

*When dealing with compact intervals, **Bolzano-Weierstrass is your best friend**.*

Theorem 3.1.2 (Local to Global Property)

If $f : [a, b] \to \mathbb{R}$ is continuous, then f is bounded.

Theorem 3.1.3 (Maxima-Minima Theorem)

*If $f : [a, b] \to \mathbb{R}$ is continuous, then it attains a **maximum** and **minimum** value.*

Theorem 3.1.4 (Intermediate Value Theorem)

Let $f : [a, b] \to \mathbb{R}$ be continuous. Define $M = \max f$ and $m = \min f$. Then:

$$\forall d \in [m, M], \ \exists\, c \in [a, b] \text{ such that } f(c) = d$$

Proof. Consider $g(x) = f(x) - d$. By assumption, 0 lies between $g(a)$ and $g(b)$. WLOG, assume $g(a) < 0$ and $g(b) > 0$.

Goal: *Find $c \in [a, b]$ such that $g(c) = 0$*

Not too hard. Consider $g(\frac{a+b}{2})$. Set $a_0 = a$ and $b_0 = b$. We would have two cases.

$$\text{Two cases} = \begin{cases} g(\frac{a+b}{2}) \geqslant 0, & \text{let } a_1 = a \text{ and } b_1 = (a+b)/2 \\ g(\frac{a+b}{2}) \leqslant 0 & \text{let } a_1 = (a+b)/2 \text{ and } b_1 = b \end{cases}$$

In both cases, $g(a_1) \leqslant 0$ and $g(b_1) \geqslant 0$. Continue this procedure as in Bolzano-Weierstrass. We will obtain the sequences: $\{a_n\}_{n\in\mathbb{N}}, \{b_n\}_{n\in\mathbb{N}}$ with the properties:

(1) $a_n \leqslant b_n$

(2) $b_n - a_n = \dfrac{b-a}{2^n}$

(3) $g(a_n) \leqslant 0$ and $g(b_n) \geqslant 0$

Use these properties with the fact that a_n and b_n are monotone bounded sequences.

■

Combining all the three theorems above, Theorem 3.1.2, 3.1.3 and 3.1.4, we obtain the following corollary.

Theorem 3.1.5 (TFAE Uniform Continuity)

Let $f : I \to \mathbb{R}$. Then the following are equivalent:

1. $\forall \varepsilon > 0, \exists \delta > 0$ *such that* $\forall x, y \in I, |x - y| < \delta \implies |f(x) - f(y)| < \varepsilon$
2. *There exists a function* $\omega_f(\delta) \to 0$ *as* $\delta \to 0$ *such that:*

$$|f(x) - f(y)| \leqslant \omega_f(\delta)$$

Comment(s):

1. The function is **Lipschitz continuous** if: $\omega_f(\delta) = C\delta$
2. The function is **Holder continuous** if: $\omega_f(\delta) = C\delta^\alpha$ for some $\alpha \in (0, 1)$.

Remark 3.1.6

For functions over a compact interval, we have the following chain of (strict) inclusions:

$$\text{Lipschitz} \subset \alpha\text{-Holder} \subset \text{Uniformly continuous}$$

This simply means:

$$\text{Lipschitz} \implies \alpha\text{-Holder} \implies \text{Uniformly continuous}$$

but this is **not a two-way street**!

It is quite immediate (after some series of practice) to prove that some function is continuous (or uniform continuous) especially if it is an exercise and it you are required to prove the function's continuity. But how about we go of proving something is not continuous or uniform continuous? This is something what I call breaking the epsilon-delta rectangle.

Example 3.1.7 (*Break the Epsilon-Delta Rectangle*)

Our claim is that $f(x) = \sin(x^2)$ is continuous on $\mathbb{R}$ but not uniformly continuous.

In layman's term, we are claiming that we can draw f without lifting the pencil, but as the function can very large, we will reach a point where we cannot draw it in such a way anymore.

Proof. Now, f is obviously continuous as the composition of two continuous functions on $\mathbb{R}$ is continuous there. Now, to show f is not uniformly continuous. Fix $\varepsilon = 1$, we are going to show that for any choice of δ, our function f will break the rectangle (dependent on ε, δ) that we forced f to be into. In other words, we will show that for any δ, despite $|x - y| < \delta$, we are able to make $|f(x) - f(y)| \geqslant 1$.

Since $x, y \in \mathbb{R}$ is arbitrary, consider $x_n = \sqrt{2\pi n}$ and $y_n = \sqrt{2\pi n + \pi/2}$ where $n \in \mathbb{N}$. Observe that:

$$|x_n - y_n| = \left|\sqrt{2\pi n} - \sqrt{2\pi n + \pi/2}\right| = \left|\frac{\pi/2}{\sqrt{2\pi n} + \sqrt{2\pi n + \pi/2}}\right| \leqslant \frac{2}{2\sqrt{2\pi n}} \leqslant \frac{1}{\sqrt{n}}$$

So, we can make $|x_n - y_n| < \delta$ whenever $n > 1/\delta^2$. But now, the rectangle has been broken since the very moment we define our sequences x_n and y_n. Why? Because $f(x_n) = 0$ and $f(y_n) = 1$ for any n. This implies:

$$|f(x_n) - f(y_n)| = 1$$

In particular, we cannot make the difference any smaller than 1 i.e. $|f(x_n) - f(y_n)| < 1$ is impossible. And we are done. ■

Note that we could have choose ε to be even smaller, e.g. $1/2$ or 0.00001 or the more exotic $\pi^{-20202019}$. It does not matter as long as we've shown the epsilon-delta rectangle is broken.

Example 3.1.8 (Another example of breaking the Rectangle)

Claim: e^x is not uniformly continuous on $(0, \infty)$

Let $x_n = \log(n+1)$ and $y_n = \log(n)$. Then, we have:

$$|x_n - y_n| = |\log(n+1) - \log(n)| = \left|\log\left(\frac{n+1}{n}\right)\right| \to 0 \text{ as } n \to \infty$$

But $|f(x_n) - f(y_n)| = 1$ and we cannot make it any smaller.

There is a powerful shortcut to check whether a function is uniformly continuous. We will, however, prove it later as we shall need the Fundamental Theorem of Calculus.

Theorem 3.1.9 (Lipschitz Continuity Shortcut)

Every differentiable function whose derivative is bounded on I is Lipschitz continuous

In symbols, $|f'| \leqslant C$ for some $C \in \mathbb{R} \implies f$ Lipschitz continuous

Note that we have not define differentiability rigorously yet up to this point. But it is pretty handy to mention the above theorem now as we will able to tame more uniform continuity questions now.

Example 3.1.10

Claim: $f = e^x$ is uniformly continuous on $(-\infty, 0)$

Since $f'(x) = e^x$ and $|e^x| \leqslant 1$ on $(-\infty, 0)$. It follows that f is Lipschitz continuous and hence, uniformly continuous on $(-\infty, 0)$.

Why do we study uniform continuity at all? Well, it turns out that if a function is uniformly continuous on an open interval (a, b), we can easily extend it to its endpoints as well i.e. it is uniformly continuous on $[a, b]$. This is another shortcut to check a function is uniformly continuous but it the domain has to be a compact interval first.

Theorem 3.1.11 (Cantor's Theorem)

Let f be a continuous function on a compact (closed and bounded) interval $[a, b]$. Then, f is uniformly continuous on $[a, b]$.

Proof. Negate the result, go to $\varepsilon - \delta$-s, choose your favorite sequence of δ that dies (obviously choose $\delta = 1/n$) as the sequence progresses. Then, apply Bolzano-Weierstrass and attack using Sequential Criterion for Continuity. Seek for a contradiction. ■

Example 3.1.12

$$\text{Claim: } f(x) = \frac{\sin x}{x} \text{ is uniformly continuous on } (0, 1)$$

As f is continuous on $[0, 1]$, then by Cantor's Theorem, f is uniformly continuous on $[0, 1]$. Now restrict the interval to $(0, 1)$ to get our intended result.

4 Continuous Functions III: Sequences and Series

We introduce a new notion and discuss convergence of sequences and series. Our central object of discussion will be the sequences of the form $\{f_n(x)\}_{n=1}^{\infty}$ where $x \in I$. We will see the interplay between the integer parameter n and the continuous parameter x. It might seem odd at first to see why are we even discussing this in the first place. However, such discussions will lead us to know more about series of functions of the form $\sum_{n=1}^{\infty} f_n(x)$ with $x \in I$ which have substantial use.

4.1 Pointwise and Uniform Convergence

Let $f_n(x)$ be continuous for all n. Roughly speaking, $f_n(x) \to f(x)$ as $n \to \infty$ **pointwise**, if this convergence takes place for every **point** $x \in I$.

Definition 4.1.1 (What is Pointwise?)

Pointwise is an adjective that means: occuring at every point on the set.

Definition 4.1.2 (Pointwise Convergence)

Let $\{f_n(x)\}_{n=1}^{\infty}$ be a sequence of functions defined on an I. Then, $f_n(x) \to f$ pointwise if:

$$\forall x \in I,\ \forall \varepsilon > 0,\ \exists\, N = N(x, \varepsilon) \text{ such that } \forall n \geqslant N,\ |f(x) - f_n(x)| < \varepsilon$$

Main concern: N depends both on x and ε

From the above definition, it is natural to discuss the following question:

> "We have a sequence of continuous functions which converge somewhere. **Is the limit going to be continuous as well?**"

The answer is sometimes yes, and sometimes no. There is no good answer to this. **Pointwise convergence does not guarantee the continuity of the limit.** We shall look at probably the most famous example for pointwise convergence but which its limit is not continuous.

Example 4.1.3

Let $f_n(x) = x^n$ on the interval $[0, 1]$. Then:

$$\lim_{n\to\infty} f_n = f(x) = \begin{cases} 0, & 0 \leqslant x < 1 \\ 1, & x = 1 \end{cases}$$

So it converges pointwise. But clearly, the limit $f(x)$ is not continuous, it has a gap of length 1.

This leads us to the next question. *Is there any way that we can guarantee the continuity of the limit then*? The answer is yes! But we have to develop some new definitions first.

Definition 4.1.4 (Uniform Convergence)

Let $\{f_n(x)\}_{n=1}^{\infty}$ be a sequence of functions defined on I. Then, $f_n(x) \to f$ uniformly if:

$$\forall \varepsilon > 0,\ \exists\, N = N(\varepsilon) \text{ such that } \forall n \geqslant N,\ |f(x) - f_n(x)| < \varepsilon$$

Main concern: N depends on ε only.

Note that this definitions is equivalent to the following:

Definition 4.1.5 (Uniform Convergence II)

Let $\{f_n(x)\}_{n=1}^{\infty}$ be a sequence of functions defined on an I. Then, $f_n(x) \to f$ uniformly if:

$$\forall \varepsilon > 0,\ \exists N = N(\varepsilon) \text{ such that } \forall n \geqslant N,\ \sup_{x \in I} |f(x) - f_n(x)| < \varepsilon$$

Why are they equivalent? Well our main concern of pointwise convergence is that the choice of N so that we can make the difference small is dependent of our position i.e. our choice of x. Our first definition is just saying, oh well, let's just restrict and say that N now cannot be dependent on x – if this happens, we call it uniform convergence.

But now, if we take a step back and think: "what is usually the problem actually?" or a better question, "where are the points of x such that the difference cannot be made small?" i.e. where are actually the positions that our choice of N (dependent on x) make our difference not small enough. Well they are usually the points where the function are able to blow up, i.e. the maximal amount the function can take. Since we are dealing with bounded functions, we can actually measure this by finding the maximum of the function (or rather, supremum). Since we have a notion of size of a function, we have a notion of size of difference of two functions.

If we are able to make this maximal difference small, then we are sure to make any smaller difference small. Hence, why we choose the supremum as part of the definition. Of course, we now don't have to worry about our choice of N being dependent of x as we are talking about x being fixed when we choose the supremum, i.e. we have turned our problem from being a problem of sequence of functions to a problem of sequence of real numbers which we know how to deal with quite easily.

Theorem 4.1.6 (Uniform Limit Theorem)

Let $\{f_n\}_{n=1}^{\infty}$ be a sequence of continuous functions that converge to $f(x)$ uniformly on I. Then, f is continuous on I.

Proof. Use $\varepsilon/3$ argument. ■

4.2 Revision of Series

Here, we give a brief revision to a special type of sequence which we call series. We assume everything in this section to be prior knowledge. However, we put it here as a good refresh to the memory.

Definition 4.2.1

If we define the partial sums $A_N = \sum_{n=1}^{N} a_n$. Then, we say A_N is convergent to a sum A if the following limit exists:

$$\lim_{N \to \infty} A_N = \sum_{n=1}^{\infty} a_n$$

where we define the above limit to be equal to A.

Definition 4.2.2

The series $\sum_{n=1}^{\infty} a_n$ is said to converge absolutely if the series $\sum_{n=1}^{\infty} |a_n|$ converges. Moreover, if the latter fact is true, we write:

$$\sum_{n=1}^{\infty} |a_n| < \infty$$

Theorem 4.2.3

Every absolutely convergent series is convergent.

Theorem 4.2.4 (Estimating Speed of Convergence)

Let $A_N = \sum_{n=1}^{N} a_n$ be the partial sum absolutely converging to a limit A. Then, we have the estimate:

$$|A_N - A| \leqslant \sum_{n=N+1}^{\infty} |a_n|$$

Analysts usually calls the object $\sum_{n=N+1}^{\infty} |a_n|$ to be the *tail of the series.* This might be a new fact to some so we give a proof. We only require the great triangle inequality.

Proof.

$$|A_N - A| = \left| \sum_{n=1}^{N} a_n - \sum_{n=1}^{\infty} a_n \right| = \left| \sum_{n=N+1}^{\infty} a_n \right| \leqslant \sum_{n=N+1}^{\infty} |a_n|$$

■

Finally, we will also need the Comparison Test.

Theorem 4.2.5 (Comparison Test)

Let $\sum_{n=1}^{\infty} b_n$ be a convergent series with $b_n \geqslant 0$ and suppose that for some constant $M > 0$ we have $|a_n| \leqslant M b_n$ for all $n \in \mathbb{N}$. Then, the series $\sum_{n=1}^{\infty} a_n$ converges absolutely.

4.3 Series of Functions

In this subsection, we will discuss one of the most used tools of all, series of functions. A notable series of functions that you might have seen and used since high school is the geometric series defined on the interval $(-1, 1)$. Here, we will define a more general notion of series of functions defined on a more general interval I. We will be most interested in series of the form:

$$\sum_{n=1}^{\infty} f_n(x)$$

where each function f_n is defined on I. We define $F_N(x)$ to be the partial sums of the above series and this is by its own right a sequence. We now apply all the theory that we have developed in Subsection (4.1) to this particular sequence.

Definition 4.3.1 (Types of Convergence)

Let $F_N(x) = \sum_{n=1}^{N} f_n(x)$ where f_n is defined on the interval I.

(i) **Pointwise Convergence**:

$$F_N(x) \to F(x) \text{ as } N \to \infty \text{ for all } x \in I$$

(ii) **Uniform Convergence**:

$$\|F_N - F\|_{C(I)} \to 0 \text{ as } N \to \infty$$

(iii) **Absolute Convergence**:

$$\sum_{n=1}^{\infty} |f_n(x)| < \infty \text{ for all } x \in I$$

(iv) **Absolute and Uniform Convergence**:

$$\sum_{n=1}^{\infty} \|f_n\|_{C(I)} < \infty$$

The above definition is *better* as it goes down. Absolute and uniform convergence is probably the *best* and *nicest* type of convergence here. Why? It is simply because it implies absolute convergence which thus implies pointwise convergence. Thus, it would be wise if we would have a tool in our arsenal to check for absolute and uniform convergence. Thankfully, we do have!

A very powerful theorem to check uniform convergence of series is the so-called Weierstrass M-Test. It says that if the modulus of the function is bounded by some sequence of positive real numbers, α_n; and such α_n converges as a series, then f_n converges uniformly as a series.

Theorem 4.3.2 (Weierstrass M-Test +Version)

*Let f_n be a sequence of continuous functions defined on $I \subset \mathbb{R}$ for every $n \in \mathbb{N}$. Let $\alpha_n > 0$ be **positive real numbers** such that:*

$$|f_n(x)| \leqslant \alpha_n \text{ for all } x \in I, n \in \mathbb{N}$$

*(i) If $\sum_{n=1}^{\infty} \alpha_n$ converges, then $\sum_{n=1}^{\infty} f_n$ **converges absolutely and uniformly** on I to a continuous function $F(x)$.*

(ii) Moreover, denote $F_N(x)$ to be the partial sum of the series of f_n, then, we have the estimate:

$$\|F_N(x) - F\|_{C(I)} \leqslant \sum_{n=N+1}^{\infty} \alpha_n$$

Note that if $|f_n(x)| \leqslant \alpha_n$ for every n, then taking supremums over I on both sides yield

$$\|f_n\|_{C(I)} \leqslant \sup_{x\in I} \alpha_n = \alpha_n$$

The last equality is true because α_n is independent of x. Also, if $\|f_n\|_{C(I)} \leqslant \alpha_n$, then automatically everything else will be less than α_n. In other words, we could replace the condition of $|f_n(x)| \leqslant \alpha_n$ with $\|f_n\|_{C(I)} \leqslant \alpha_n$. We haven't give a proof to the M-Test above so let's start.

Proof. First note that $F_N(x) = \sum_{n=1}^{N} f_n(x)$ is continuous as a sum of finitely many continuous functions is continuous. Since $|f_n(x)| \leqslant \alpha_n$, taking supremums we have $\|f_n\|_{C(I)} \leqslant \alpha_n$. Thus, the Comparison Test tells us that if $\sum \alpha_n$ converges, then so does $\sum \|f_n\|_{C(I)}$. Thus, $\sum f_n(x)$ converges absolutely and uniformly. This partially proves (i). We have to show that it converges to a continuous function. We take a detour to prove (ii), we first have:

$$|F(x) - F_N(x)| \leqslant \sum_{n=N+1}^{\infty} |f_n(x)| \leqslant \sum_{n=N+1}^{\infty} \alpha_n$$

This proves (ii). Taking supremum over I we have:

$$\|F(x) - F_N(x)\|_{C(I)} \leqslant \sup_{x\in I} \sum_{n=N+1}^{\infty} \alpha_n = \sum_{n=N+1}^{\infty} \alpha_n \tag{4.1}$$

The last equality follows because the term is independent of x, so supremum does nothing there. Now, since the RHS of (4.1) goes to 0 as $N \to \infty$, we have that indeed $F_N(x) \to F(x)$ uniformly which implies F is continuous on I by the Uniform Limit Theorem. ■

A clarification. The RHS of (4.1) goes to 0 as $N \to \infty$ does not mean $\sum_{n=1}^{\infty} \alpha_n = 0$. Instead the RHS here is the difference between $\sum_{n=1}^{N} \alpha_n$ and $\sum_{n=1}^{\infty} \alpha_n$. And since we assumed $\sum_{n=1}^{\infty} \alpha_n$

converges, we can make this difference arbitrarily small (i.e. goes to 0 as N gets large enough). This difference is precisely the RHS.

4.4 Some Special Series

This subsection is put here to give a motivation of what upcoming Analysis can we see beyond this module. We will have a short discussion on Power Series, Dirichlet Series and Fourier Series.

4.4.1 Power Series

Let $f_n = a_n x^n$ here where $a_n \in \mathbb{R}$ and $x \in \mathbb{R}$. Thus, power series are series of the form:

$$\sum_{n=0}^{\infty} a_n x^n \tag{4.2}$$

Theorem 4.4.1

Suppose the series (4.2) converges absolutely for $x = R > 0$. Then, it converges absolutely and uniformly for every $x \in [-R, R]$ and the sum is continuous on $[-R, R]$.

Proof. Let $I = [-R, R]$ and $f_n(x) = a_n x^n$. Then, we have:

$$|f_n(x)| = |a_n x^n| = |a_n||x_n| \leqslant |a_n|R^n \text{ for } x \in I$$

Absolute convergence of (4.2) at $x = R$ implies that $\sum_{n=0}^{\infty} |a_n|R^n < \infty$. By the Weierstrass M-Test, $\sum_{n=1}^{\infty} f_n$ converges absolutely and uniformly for every $x \in I$ to a continuous function also on I. ■

Note that the *maximum of such* R is called the radius of convergence. Moreover, the theorem above extends verbatim to $\mathbb{C}$. Just redefine $x \in \mathbb{C}$. We would have to make a minor tweak because in the complex setting.

Theorem 4.4.2 (Same Theorem, Complex Version)

Suppose the series $\sum_{n=0}^{\infty} a_n z^n$ converges absolutely for some $z \in \mathbb{C}$ with $z \neq 0$. Denote $R = |z|$. Then, $\sum_{n=0}^{\infty} a_n z^n$ converges absolutely and uniformly for $|z| \leqslant R$ and the sum is a continuous function of z in $|z| \leqslant R$.

4.4.2 Dirichlet Series

4.4.3 Fourier Series

5 Differentiability

5.1 The definitions and their properties

Definition 5.1.1 (Notion of a Derivative)

Let $f : I \to \mathbb{R}$ where $I \subset \mathbb{R}$ is an open interval (could be union of open intervals). Then, we say f is differentiable at $x_0 \in I$ if the following limit exists:

$$\lim_{h\to 0}\left(\frac{f(x_0+h)-f(x_0)}{h}\right) \tag{5.1}$$

If it exists, we write it to be equal to $f'(x_0)$.

There are other ways to define the integral, we could rewrite (5.1) in a different way, we could resort to epsilons and deltas or we could rewrite it in Scott's favorite definition - using the small o's.

Definition 5.1.2 (TFAE Differentiation)

Let $f : I \to \mathbb{R}$ where $I \subset \mathbb{R}$ is an open interval. Then, the following are equivalent:

(i) f is differentiable at x_0.

(ii) $f'(x_0)$

(iii) The limit of expression (5.1) exists

(iv) Let $x = x_0 + h$ in (5.1) to obtain:

$$\lim_{x\to x_0}\left(\frac{f(x)-f(x_0)}{x-x_0}\right) \tag{5.2}$$

and this limit exists.

(v) Scott's favorite definition. $f(x_0 + h)$ can be written in the following way.

$$f(x_0+h) = f(x_0) + h\xi + o(h) \tag{5.3}$$

for some $\xi \in \mathbb{R}$ and if this is possible, we write $\xi = f'(x_0)$.

(vi) For every $\varepsilon > 0$, there exists a $\delta > 0$ such that whenever $0 < |h| < \delta$, then:

$$\left|\frac{f(x_0+h)-f(x_0)}{h} - \xi\right| < \varepsilon \tag{5.4}$$

for some $\xi \in \mathbb{R}$. If we can find a δ without fail, we write $\xi = f'(x_0)$.

For all this definition to make sense, we assume f is defined in a small neighbourhood of x_0 i.e. $f : (x_0 - \mu, x_0 + \mu) \to \mathbb{R}$ is defined for some $\mu > 0$ (with μ small).

The other definitions are standard except for (v). We will therefore prove it by showing that it is equivalent to (iii).

Proof. Suppose (v) holds true. Then, $f(x_0 + h) = f(x_0) = h\xi + o(h)$ for some $\xi \in \mathbb{R}$. We can divide both sides by h to arrive at:

$$\frac{f(x_0+h)-f(x_0)}{h} = \xi + \frac{o(h)}{h}$$

Take limits on both sides. On the LHS, we have exactly (5.1). On the RHS, the limit does nothing to ξ as it is simply a constant here. And by definition of $o(h)$, we have that $o(h)/h$ dies as $h \to 0$. So the RHS is finite. Hence, the limit of (5.1) exists as required. ■

Thus, we are free to use any of the definitions that we like as they are all equivalent. Some quick remarks before we get into examples.

Remark 5.1.3 (A remark on differentiation)

1. It is possible that the limit (5.1) does not exists. In this case, we say f is not differentiable and all of the definitions in TFAE Differentiation (5.1.2) failed.
2. We say f is differentiable on X if $f'(x)$ exists for every $x \in X$. In this case, we can consider f' as a function of variable x defined on X. This function is the so-called *derivative* of f at x. A well-known notation is the Leibniz notation.
3. The right and left limits:

$$\lim_{x \to x_0^+} \left(\frac{f(x) - f(x_0)}{x - x_0} \right), \quad \lim_{x \to x_0^-} \left(\frac{f(x) - f(x_0)}{x - x_0} \right)$$

   if they exist, are said to be the right and left derivatives of f at the point x_0. The function f is differentiable if both derivatives exist and are equal.
4. If x_0 is an end point of the interval X, then it does not make sense to talk about left and right derivatives simultaneously. Only one will be defined and the other one just does not make sense.

Now, we jump into an example. Let us try to differentiate something that we would have differentiated a hundred times at least by now so that we can be surprised by the result.

Example 5.1.4

$f(x) = x^n$ is differentiable everywhere on $\mathbb{R}$. We expand $f(x+h) = (x+h)^n$ using binomial expansion:

$$f(x+h) = (x+h)^n = x^n + nx^{n-1}h + \frac{n(n-1)}{2}x^{n-2}h^2 + \ldots + h^n \tag{5.5}$$

Notice that x^n is just $f(x)$. Here, nx^{n-1} is our candidate derivative. This is true if everything in green is little $o(h)$ by Scott's favorite notation. Now, notice there is a lot of h^2 in the green part. We can rewrite (5.5) as:

$$f(x+h) = f(x) + nx^{n-1}h + h^2(\text{some stuff})$$

and the green stuff when divided by h goes to 0. In other words, it is $o(h)$. So, surprise surprise, $f'(x) = nx^{n-1}$.

Example 5.1.5 (The modulus function)

Consider $f(x) = |x|$. This is differentiable for $x \neq 0$ and not differentiable at $x = 0$. At $x_0 = 0$, we have:

$$\frac{f(0+h) - f(0)}{h} = \frac{|h| - 0}{h} = \frac{|h|}{h}$$

which goes to 1 if $h > 0$ and goes to -1 if $h < 0$. Since the limit has to be unique, we conclude that the limit does not exist at this point and hence, the derivative.

From this example, we could see how the derivative tells you how *smooth* the function is. If it is not differentiable at a point, then the function must not be smooth at that particular point. Now, Let's try a more exotic example.

Example 5.1.6

Consider the function:

$$f(x) = \begin{cases} x^2 & \text{if } x \in \mathbb{Q} \\ -x^2 & \text{if } x \in \mathbb{R} \backslash \mathbb{Q} \end{cases} \tag{5.6}$$

This function is not differentiable anywhere except at $x = 0$. Let's look at the function more closely when $x_0 = 0$:

$$\frac{f(x_0 + h) - f(x_0)}{h} = \frac{f(h) - f(0)}{h} = \frac{f(h)}{h} = \frac{\pm h^2}{h} = \pm h \to 0 \text{ as } h \to 0$$

Elsewhere besides 0, we can easily see no such limit exists by considering the sequence of $a_n \to x_0$ with $a_n \in \mathbb{Q}$ and $a_n \in \mathbb{R} \backslash \mathbb{Q}$. We could see the limit of:

$$\frac{f(a_n + h) - f(a_n)}{h}$$

exists in both cases, but they are not the same limit.

After feeling confident that our definition of the derivative makes sense and is coherent to what we know about it all this while, we now proceed to develop the theory.

5.2 Differentiability and Continuity

This subsection is very short, but it highlights the important one-way street relation between differentiation and continuity. This serves as an important bridge in the next section where we talk about the famous Taylor's theorem and analytic functions.

Theorem 5.2.1

If the derivative $f'(x)$ exists, then f is continuous at x.

If we are being pedantic, we write: Let f be defined on I. If f is differentiable at a point $x_0 \in I$, then f is continuous at x_0.

Proof. The proof just uses $^{\times}(x - x_0)$-manipulation.

$$\lim_{x \to x_0} (f(x) - f(x_0)) = \lim_{x \to x_0} \left(\frac{f(x) - f(x_0)}{x - x_0} \right) (x - x_0) = f'(x_0) \cdot 0 = 0$$

■

A very important comment here is that the **converse is not true!** A continuous function is not necessarily differentiable.

Non-Example 5.2.2 (Modulus function II)

Let $f(x) = |x|$. This function is continuous at $x = 0$. But as we have seen in Example (5.1.5), it is not differentiable there. Here is an immediate counter-example.

5.3 Algebraic properties of differentiation

Here, we recover the fact that the derivative is linear, the Leibniz product rule and respectively its consequence (differentiating quotient). Basically, this subsection is dedicated to the *algebra* that we can do with differentiation.

Theorem 5.3.1 (Differentiation is Awesome)

Let f, g be functions defined on $I \subset \mathbb{R}$ and suppose both functions are differentiable at some point $c \in I$. Then,

(i) $(f+g)'(c) = f'(c) + g'(c)$

(ii) $(\alpha f)'(c) = \alpha f'(c)$ *for any* $\alpha \in \mathbb{R}$.

(iii) $(fg)'(c) = f'(c)g(c) + f(c)g'(c)$

(iv) $(f/g)'(c) = \frac{g(c)f'(c) - f(c)g'(c)}{[g(c)]^2}$ *given that* $g(c) \neq 0$.

Proof. Statement (i) and (ii) are left as exercises. Let's prove (iii). This proof just uses $^{+}(f(x)g(c))$-manipulation in its first line.

$$\begin{aligned}\frac{(fg)(x) - (fg)(c)}{x-c} &= \frac{f(x)g(x) - f(x)g(c) + f(x)g(c) - f(c)g(c)}{x-c} \\ &= f(x)\left(\frac{g(x)-g(c)}{x-c}\right) + g(c)\left(\frac{f(x)-f(c)}{x-c}\right)\end{aligned}$$

Take limits on both sides and note that f, g are both differentiable at c to get the result. Proving (iv) feels like plagiarising ourselves here as it is very similar so we omit it. The only difference is that we use a $^{+}(f(x)g(x))$-manipulation instead. ■

Note that we can prove everything above using Scott's favorite definition with ease. It's just that things will get messier than it should, and the reason why we intend to stick with the standard definition.

Quick remark: The derivative of a constant function is identically zero everywhere. If $f(x)$ is constant for every x everywhere f is defined. Then choose any x, c to have that $f(x) - f(c) = 0$. Divide through by $x - c$ and take the limit of $x \to c$. We can use this fact together with (iii) in the previous theorem to deduce (ii) as well. How? Well let $g = \alpha$ in (iii). Let us move on to the chain rule.

Theorem 5.3.2 (Chain Rule)

Let g be differentiable at c and f be differentiable at $g(c)$. Then $(f \circ g)(x)$ is differentiable at c and:

$$(f \circ g)'(c) = g'(c)f'(g(c)) \tag{5.7}$$

I have to admit that I am indeed being mathematically immoral here. The scrupulous analyst which begs for rigour everywhere would definitely denounce me instantly if I speak of this theorem this way in front of his/her face. Yes, we can write:

> "Let $f : A \to \mathbb{R}$ and $g : B \to \mathbb{R}$ such that $f(A) \subset B$ so that the composition (of functions) we make will be well-defined ... and so on ..."

and we are capable in doing so. However, in these notes, we assume that the readers are experts; that they understand when something is left missing on purpose and they automatically know how to fill those missing bits. Now, let us prove this theorem. We will leave the proof of the chain rule using standard derivative definitions to you as we will proceed using Scott's favorite definition.

Proof. We know that g is differentiable at c so by Scott's favorite definition:

$$g(c+h) = g(c) + hg'(c) + o(h) \tag{5.8}$$

Also, f is differentiable at $g(c)$ so:

$$f(g(c)+k) = f(g(c)) + kf'(g(c)) + o(k) \tag{5.9}$$

Note that k here can (almost) be any number, depending on the domain of g of course. The rest of the proof is just combining (5.8) into (5.9). We will show this in detail. Then plugging in (5.8) into the composition $(f \circ g)(c+h)$:

$$(f \circ g)(c+h) = f(g(c+h)) = f(g(c) + hg'(c) + o(h)) \tag{5.10}$$

Now, let $k = hg'(c)+o(h)$. We thus get to write (5.10) in the form $(f \circ g)(c+h) = f(g(c)+k)$. Now apply (5.9) to get:

$$(f \circ g)(c+h) = f(g(c)) + (hg'(c) + o(h))f'(g(c)) + o(k) \tag{5.11}$$

Distribute the one in blue to get:

$$(f \circ g)(c+h) = f(g(c)) + h\{g'(c)f'(g(c))\} + o(h)f'(g(c)) + o(k)$$

But $o(k) = o(hg'(c) + o(h))$ by our previous substitution,

$$(f \circ g)(c+h) = f(g(c)) + h\{g'(c)f'(g(c))\} + o(h)f'(g(c)) + o(hg'(c) + o(h)) \tag{5.12}$$

Recall that the function $\phi(x)$ is differentiable at c if $\phi(c+h) = \phi(c) + h\phi'(c) + o(h)$ and we say that $\phi'(c)$ is the derivative at that point. Now if we compare to (5.12), we have that the derivative of $(f \circ g)(x)$ at c is $g'(c)f'(g(c))$ if and only if the term after it are all $o(h)$ - the ones in gold colour. $o(h)f'(g(c))$ is automatically $o(h)$ as $f'(g(c))$ is constant here. The one that is left to check is $o(hg'(c) + o(h))$. But this is also $o(h)$ (exercise) so we are done. ■

We initially intended to leave the last part of checking $o(hg'(c) + o(h))$ is $o(h)$ as an exercise but we feel that we are being too merciless here. The proof is obvious (after spending an hour on it that is). It utilizes $^{\times}(hg'(c) + o(h))$-manipulation and a passing-the-limit argument. Here's the proof:

Proof.

$$\frac{o(hg'(c)+o(h))}{h} = \frac{1}{h}\frac{o(hg'(c)+o(h))}{hg'(c)+o(h)}(hg'(c)+o(h)) = \frac{o(hg'(c)+o(h))}{hg'(c)+o(h)}\left(g'(c) + \frac{o(h)}{h}\right)$$

Let $k = hg'(c) + o(h)$. We have by definition of $o(k)$:

$$\lim_{k\to 0}\frac{o(k)}{k} = 0$$

which is good as we now know that $o(k)/k$ is well-behaved as $k \to 0$. It is tempting to just rewrite $k = h$ and be done with it but that seems to be cheating. What we want here is for $\lim_{h\to 0} o(k)/k = 0$. But, it turns out, we get this for free because k is dependent on h and $k \to 0$ as $h \to 0$. So, we can make h as small as we want it to be and upon reaching some threshold, k will become small as well. And by definition of $o(k)$, we know how $o(k)/k$ behaves when k is really small. In other words, we have:

$$\lim_{h\to 0}\frac{o(k)}{k} = \lim_{k\to 0}\frac{o(k)}{k} = 0$$

And since $g'(c) + o(h)/h$ is convergent as $h \to 0$ (we don't really care what it's value is, we just care it is not divergent), we have that the whole thing goes to 0 as $h \to 0$ (due to multiplicative property of limits). In other words, it is $o(h)$ as required. ■

We also give an alternate proof which is way simpler but requires a stronger grasp over the algebra of little-oh. The fact that we use here is that if the function is $o(h)$, then it is $o(1)$. We will highlight in red where this is (ab)used in the proof below. Moreover, $o(h^a) + o(h^a) = o(h^a)$ and $o(o(h)) = o(h)$.

Proof. We will use all the facts mentioned above. Also, we will be $^+(o(h))$-manipulating.

$$\begin{aligned} o(hg'(c) + o(h)) &= o(o(1) + o(h)) = o(o(1) + o(1)) = o(o(1)) = o(o(1) - o(h) + o(h)) \\ &= o(o(1) - o(1) + o(h)) = o(o(h)) = o(h) \end{aligned}$$

■

I am pretty sure by now you know why we were being merciful. The proof is a bit involved. However, if you write this down on a piece of paper, it shouldn't be this long (albeit the sketch of reaching the conclusion will fill up the whole sheet - both sides).

I am rather confident that at some point of the proof (either the previous theorem or just proving the little bit being $o(h)$), the thought of "why not prove it from the other definitions" would probably came across your mind. Well, we are not surprised by such thoughts because of how the proof was written. If the proof of Theorem (Chain Rule) was written on paper, it will be very short. The proof written here was very lengthy because we wanted to give a careful treatment so that it is accessible to any reader. By using Scott's favorite definition, the proof was kind of designed to be straightforward (another word for "don't have to use the brain that much") unlike when you use standard definitions which requires some thinking. Try to do it using standard definitions, we deeply encourage you.

Example 5.3.3

Let $f(x) = \ln(x)$ where $x > 0$ and $g(x) = e^x$. Then, by definition $f(g(x)) = x$ for all x. If we differentiate this and plug in c, we have $f(g(x))' = 1$ for all x. We know $g'(x) = g(x) = e^x$. So, we have $g'(f(c)) = e^{\ln c} = c$. Now, by Theorem (5.3.2), we have that $f(g(c))' = f'(c)g'(f(c))$. Hence, we plug in what we know already to have $1 = f'(c)c$ and thus $f'(c) = 1/c$ as expected.

5.4 Mean Value Theorems

This subsection is dedicated to proving the Mean Value Theorems, or as I'd like to call it, the *obviously true* Theorems. Before we go into the business of proofs, let's gain some insight and intuition of what we are trying to prove.

Some intuition – Choose your favorite notice board at school or office, fix two thumb tacks a and b anywhere on the board such that b is to the right of a and tie one end of the string to a and another end to b (it need not be tight).

Now, this string is connected in the sense that if you touch the string at point a, and journey your finger through the string to the point b, you are able to do it without lifting at any instance no matter how slow or fast you do it. In other words, the string is continuous everywhere between a and b. Furthermore, however tight or loose you make the string (when you were tying), it wouldn't have sharp points (except at the end points) and it will always be smooth.

If you were to take a ruler and align it so that both thumb tacks are hit, and fix this orientation. You'll discover that there will be somewhere on the string that touches the ruler exactly once when you traverse this ruler throughout the path of the string - provided that you used the same orientation. This is exactly what the Mean Value Theorem says. In fact, the theorems that will be used to prove MVT utilizes the exact same idea as we will see below.

Theorem 5.4.1 (Fermat's Theorem)

Let f be differentiable on (a, b). If f has a local maxima or minima at $c \in (a, b)$. Then, $f'(c) = 0$.

Proof. Suppose $f(c)$ is the local max. Perturb $f(c)$ a bit to see that $f(c+h) \leqslant f(c)$ for some $|h| < \delta$ where δ is small. Now consider when h is positive and negative and use the fact that the limit (when it exists) is unique. Repeat for local min. ■

This statement is obviously true when one see it pictorially and our senses served us really well here. Let's prove another obvious fact which was first realized by Michel Rolle.

Theorem 5.4.2 (Rolle's Theorem)

Let (i) f continuous on $[a, b]$, (ii) f differentiable on (a, b) and (iii) $f(a) = f(b)$. Then, there exists $c \in (a, b)$ with $f'(c) = 0$.

Proof. f is continuous on a compact interval so f attains a maxima and a minima somewhere. If **both** max and min occurs at endpoints then f is constant, you're done.

If either max or min occurs at some $c \in (a, b)$, then Fermat's tell us that $f'(c) = 0$. ■

We now have enough tools to prove the so-called Mean Value Theorem. We state the theorem below and give a proof immediately.

Theorem 5.4.3 (Mean Value Theorem [Cauchy 1823])

Let (i) f be continuous on $[a, b]$ and (ii) f differentiable on (a, b). Then, there exists $\xi \in (a, b)$ with the property:

$$f(b) - f(a) = (b - a)f'(\xi) \tag{5.13}$$

Proof. This is the proof of MVT. Let:

$$g(x) = f(x) - f(a) - (x - a)\left(\frac{f(b) - f(a)}{b - a}\right)$$

Then, $g(a) = g(b) = 0$. Rolle's Theorem implies the existence of some $\xi \in (a, b)$ such that $g'(\xi) = 0$. Compute $g'(x)$ and you know what to do. ■

Note that (5.13) can be rewritten in other different forms. Most notably, it is written as:

$$\frac{f(b) - f(a)}{b - a} = f'(\xi) \tag{5.14}$$

by dividing both sides by $b - a$ provided that $b - a \neq 0$; or in this form:

$$f(b) = f(a) + (b - a)f'(\xi) \tag{5.15}$$

by adding $f(a)$ to both sides. These are quite obvious manifestations of the MVT as the manipulations needed to reach the changed form are just simple algebraic tweaks.

Remark 5.4.4

One often think of b as a variable x and $a = x_0$ fixed. Then, (5.15) can be rewritten as:

$$f(x) = f(x_0) + (x - x_0)f'(\xi) \tag{5.16}$$

or writing $x = x_0 + h$ in (5.16):

$$f(x_0 + h) = f(x_0) + hf'(\xi) \tag{5.17}$$

Let's test MVT's power to prove a very obvious but useful estimate for the sine function.

Example 5.4.5

Claim: $|\sin x| \leqslant |x|$ for all $x \in \mathbb{R}$

Suppose $x \neq 0$ as if $x = 0$, the inequality turns into a true equality. We know that $\sin x$ is differentiable everywhere on $\mathbb{R}$. Fix an arbitrary $a \in \mathbb{R}$. By the MVT, there exists a $c \in (0, a) \subset \mathbb{R}$ such that:

$$\sin(a) - \sin(0) = a\cos(c) \implies \sin(a) = a\cos(c)$$

Thus,

$$|\sin(a)| = |a||\cos(c)| \leqslant |a|$$

But since a was arbitrary, we conclude that $|\sin(x)| \leqslant |x|$ for every $x \neq 0$.

Theorem 5.4.6 (Cauchy's MVT)

Let f, g be continuous on $[a, b]$ and differentiable on (a, b) and $g' \neq 0$ on (a, b). Then, there exists $c \in (a, b)$ with:

$$\frac{f'(c)}{g'(c)} = \frac{f(b) - f(a)}{g(b) - g(a)} \tag{5.18}$$

Proof. Note that $g(b) - g(a) \neq 0$ (exercise, hint: what will Rolle's say). Let:

$$\phi(x) = (g(b) - g(a))f(x) - (f(b) - f(a))g(x)$$

Then, $\phi(a) = \phi(b) = 0$. Rolle's Theorem implies the existence of some $\xi \in (a, b)$ such that $\phi'(\xi) = 0$. Compute $\phi'(x)$ and you know what to do. ■

Observe that the real superstar in proving MVT (and Cauchy's MVT) is Rolle's Theorem. Moreover, the proof for MVT and Cauchy's MVT is isomorphic. Let us highlight some easy consequences to the MVT.

Corollary 5.4.7

Let f be differentiable on (a, b). If $f'(x) = 0$ for all $x \in [a, b]$, then f is constant.

Proof. MVT tell us that for every $x \in [a, b]$, there exists a $c \in (a, x)$ such that:

$$f(x) - f(a) = (x - a)f'(c)$$

But $f'(x) = 0$ everywhere. ■

Corollary 5.4.8

Let f be continuous on $[a, b]$ and differentiable on (a, b).

(1) If $f'(x) > 0$ for all $x \in (a, b)$, then f is (strictly) increasing.

(2) If $f'(x) < 0$ for all $x \in (a, b)$, then f is (strictly) decreasing.

(3) If $f'(x) \geqslant 0$ for all $x \in (a, b)$, then f is non-decreasing.

(4) If $f'(x) \leqslant 0$ for all $x \in (a, b)$, then f is non-increasing.

Proof. Choose $x, y \in [a, b]$ such that $a \leqslant x < y \leqslant b$ so that $y - x > 0$. By MVT, there exists a $c \in (x, y)$ such that:

$$f(y) - f(x) = (y - x)f'(c)$$

(1) If $f'(c) > 0$, then $f(y) - f(x) > 0$.

(2) If $f'(c) < 0$, then $f(y) - f(x) < 0$.

(3) If $f'(c) \geqslant 0$, then $f(y) - f(x) \geqslant 0$.

(4) If $f'(c) \leqslant 0$, then $f(y) - f(x) \leqslant 0$.

■

Corollary 5.4.9

Let f be continuous on $[a, b]$ and twice differentiable on (a, b). Let $c \in (a, b)$.

(1) If $f'(c) = 0$ and $f''(c) > 0$, then f has a local minimum at c.

(2) If $f'(c) = 0$ and $f''(c) < 0$, then f has a local maximum at c.

Proof. We will only proof (1). The proof of (2) is step-isomorphic. Since $c \in (a, b)$ and f is twice differentiable on (a, b), f' is differentiable at c. We move to $\varepsilon - \delta$. We have then, for every $\varepsilon > 0$, there exists a $\delta > 0$ such that whenever $0 < |h| < \delta$, we have:

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

But $f'(c) = 0$. So $f'(c)$ vanishes in the above expression. Moreover, we can expand the modulus and write:

$$f''(c) - \varepsilon < \frac{f'(c+h)}{h} < f''(c) + \varepsilon$$

Remember that ε can be anything positive. Here's the magic, let $\varepsilon = f''(c)/2$. Then, on one side, we have:

$$\frac{f'(c+h)}{h} > f''(c) - \frac{f''(c)}{2} > \frac{f''(c)}{2} > 0$$

- For $h > 0$, we have that $f'(c+h) > 0$ which by the previous corollary, implies that f is strictly increasing.
- For $h < 0$, we have $f'(c+h) < 0$ which implies f is strictly decreasing.

Draw the graph and convince yourself that these two facts imply that f has a local minimum at c. Repeat to see local max. ■

5.5 Differentiating convergent sequences

We haven't mention anything about it, but this has been one of our bigger goals of this section. We were mesmerised by a theorem of the last section that we seek to find a differentiable analogue for it. However, it is not as easy this time.

Theorem 5.5.1 (Term-by-Term Differentiation)

Let f_n be a sequence of differentiable functions on (a, b) such that:

(i) There exists $x_0 \in [a, b]$ such that $f_n(x_0)$ is a convergent sequence of real numbers.

(ii) The sequence of derivative functions f_n' converges uniformly to ϕ for some $\phi : [a, b] \to \mathbb{R}$.

Then, $f_n \to g$ uniformly to a function $g : [a, b] \to \mathbb{R}$ which is differentiable on (a, b) with $g' = \phi$. In other words:

$$\frac{d}{dx} \lim f_n = \lim \frac{d}{dx} f_n$$

It is not enough to hope $f_n \to g$ uniformly to get this result (i.e. $g' = \phi$) no matter how intuitive it may seem. We will need $f_n'(x)$ to be uniformly convergent as well. This can be seen through the counter-example below.

Non-Example 5.5.2

Take $f_n = \sin(nx)/\sqrt{n}$. Then, $|f_n| < 1/\sqrt{n} \to 0$ so f_n converges uniformly. However, $f_n'(x) = \sqrt{n}\cos(nx)$ is **not** convergent.

The proof of Theorem (5.5.1) will be postponed although we have enough machinery to prove it. We have the Mean Value Theorem which is sufficient. The only thing left to prove is the notion of f_n being uniformly Cauchy iff f_n is uniformly convergent. Being uniformly Cauchy is just being Cauchy with the threshold N being independent of x:

$\forall \varepsilon > 0, \exists N \in \mathbb{N}$ (independent of x) such that $\forall m, n \geqslant N$, we have the estimate:

$$|f_n(x) - f_m(x)| < \varepsilon$$

In this way, if we can prove f_n with the the assumptions as stated in (5.5.1) to be uniformly Cauchy, then we are done. But the proof is very tedious. As we will see and prove later, with the help of more advanced machinery - that is the Fundamental Theorem of Calculus, we can prove the statement with fewer lines. If your curiosity supersedes you, see page 42, Corollary (8.4.5) for this proof.

Remark 5.5.3 (On Infinitely Differentiable Functions)

Let f, g **be infinitely differentiable functions on** $X \subset \mathbb{R}$ (in other words $f, g \in C^\infty(X)$). Then, their **derivatives**, **linear combination**, their **product** (and hence, quotient provided the denominator is non-zero), their **reciprocal** (multiplicative inverse) and their **composition** are all **infinitely differentiable on** X.

6 Taylor's Theorem and Analytic Functions

6.1 Taylor's Theorem

Theorem 6.1.1 (Taylor's Theorem)

Let $f : (a-\varepsilon, a+\varepsilon) \to \mathbb{R}$ for some $\varepsilon > 0$, be n-times continuously differentiable. Then, for each $x \in (a-\varepsilon, a+\varepsilon)$, there exists a c between a and x such that:

$$f(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(a)}{k!}(x-a)^k + R_n(x) \tag{6.1}$$

where

$$R_n(x) = \frac{(x-a)^n}{n!} f^{(n)}(c) \text{ is called the remainder term} \tag{6.2}$$

Remark 6.1.2 (Some comments on Taylor's)

1. If $f(x)$ is a polynomial, then $R_n(x) = 0$.
2. When $n = 1$, we get the MVT back which led us to calling Taylor's Theorem as the n-th Mean Value Theorem
3. We may rewrite (6.1) to be simpler by letting $h = x - a$:

$$f(a+h) = f(a) + hf'(a) + \frac{h^2}{2!} f''(a) + \ldots + \frac{h^{n-1}}{(n-1)!} f^{(n-1)}(a) + R_n(h) \tag{6.3}$$

where $R_n(h) = h^n/n!\ f^{(n)}(a)$. Rewriting it in this way can bring some insight of thinking Taylor's theorem as perturbing the function by a small value h (hence, $f(a+h)$).

We will prove the theorem below. A first word of advice: do not be discouraged by how it looked overall. The idea is really simple - using the MVT again and again and again.

Proof. The proof is really easy but a bit more involved. Suppose WLOG that $x < a$ (otherwise we could just replace x with a and vice-versa and everything is still the same). The idea is to apply **Cauchy's MVT** many times on (x, a) by setting $b = x$ (in the MVT). Here's the clever trick, set:

$$g(x) = (x-a)^n \tag{6.4}$$

$$\phi(x) = f(x) - f(a) - (x-a)f'(a) - \ldots - \frac{(x-a)^{n-1}}{(n-1)!} f^{(n-1)}(a) \tag{6.5}$$

What's clever here is that:

$$g^{(j)}(a) = \phi^{(j)}(a) = 0 \text{ for } 1 \leqslant j \leqslant n-1 \text{ and } g^{(n)}(x) = n! \text{ for any } x. \tag{$\dagger$}$$

Then, take the quotient $\phi(x)/g(x)$ to see:

$$\frac{\phi(x)}{g(x)} = \frac{\phi(x) - \phi(a)}{g(x) - g(a)} = \frac{\phi'(c_1)}{g'(c_1)} \tag{6.6}$$

for some $c_1 \in (a, x)$. By ($\dagger$), $\phi'(a) = g'(a) = 0$ so we can do like (6.6) again. Eventually, we will reach:

$$\frac{\phi(x)}{g(x)} = \frac{\phi'(c_1)}{g'(c_1)} = \frac{\phi''(c_2)}{g''(c_2)} = \ldots = \frac{\phi^{(n-1)}(c_{n-1}) - \phi^{(n-1)}(a)}{g^{(n-1)}(c_{n-1}) - g^{(n-1)}(a)} \tag{6.7}$$

for some $c_{n-1} \in (a, c_{n-2})$. This is the maximum we can go as:

$$g^{(n)}(a) = n! \neq 0$$

We can still apply Cauchy's MVT one last time to see that there exists a $c_n \in (a, c_{n-1})$ such that:

$$\frac{\phi^{(n)}(c_n)}{g^{(n)}(c_n)} = \frac{\phi^{(n-1)}(c_{n-1}) - \phi^{(n-1)}(a)}{g^{(n-1)}(c_{n-1}) - g^{(n-1)}(a)} = \frac{\phi(x)}{g(x)} \tag{6.8}$$

But now note $\phi^{(n)}(c) = f^{(n)}(c)$ and $g^{(n)}(x) = n!$ for all x so we can rewrite (6.8) as:

$$\frac{\phi(x)}{g(x)} = \frac{f^{(n)}(c)}{n!} \implies \phi(x) = \frac{f^{(n)}(c)}{n!} g(x) = \frac{f^{(n)}(c)}{n!}(x-a)^n \tag{6.9}$$

By recalling what is $\phi(x)$, do some rearranging and letting $R_n(x)$ be as in the theorem, we are done. ■

6.2 Analytic Functions

So we got Taylor's. What's next? Well, we are particularly interested in the case when we can let $n \to \infty$ and obtain an exact equality i.e. when:

$$f(a+h) = \sum_{k=0}^{\infty} \frac{h^k}{k!} f^{(k)}(a) \tag{6.10}$$

Thus, we get the polynomial approximation (6.1) evolved into a **Power series identity** (6.10). Note that this need not happen in general. If this holds, f is said to be analytic on $(a-\varepsilon, a+\varepsilon)$.

Definition 6.2.1 (Local Analyticity)

Let f be infinitely differentiable on the interval $(a-\varepsilon, a+\varepsilon)$. If there exists $\varepsilon > 0$ such that for all $x \in (a-\varepsilon, a+\varepsilon)$, one has

$$f(x) = \sum_{k=0}^{\infty} f^{(k)}(a) \frac{(x-a)^k}{k!} \tag{6.11}$$

Then, the function is said to be **analytic** on the interval $(a-\varepsilon, a+\varepsilon)$. Moreover, the equation (6.11) is called the **Taylor Series** of f at the point a.

Definition 6.2.2 (Global Analyticity)

A function $f : I \to \mathbb{R}$ where $I \subset \mathbb{R}$ is **analytic on I** if for every $a \in I$, there exists a neighbourhood $(a-\varepsilon, a+\varepsilon) \subset I$ of a on which f is analytic.

Being analytic is a local property. Therefore, we would need a global definition just like we needed one for continuity and differentiability. However, the difference for analyticity between local and global is not as straightforward as those of, say, continuity. If we could compare the two with a single sentence difference, it would be this:

$$\text{Local Analyticity} = \text{Analytic on a particular interval } (a-\varepsilon, a+\varepsilon) \text{ where } a \text{ is fixed}$$
$$\text{Global Analyticity} = \text{Analytic on every interval } (a-\varepsilon, a+\varepsilon) \text{ for every possible } a$$

By the definition above, checking for analyticity is quite straightforward. Being analytic just looks like we are running the remainder term to infinity. In fact, it is exactly just that.

Lemma 6.2.3 (Analyticity Lemma)

A function f is analytic on $(a-\varepsilon, a+\varepsilon)$ if and only if $R_n(x) \to 0$ as $n \to \infty$ for every $x \in (a-\varepsilon, a+\varepsilon)$.

Proof. Do we even require one here? ■

Lemma 6.2.4 (Sum and Products of Analytic Functions are Analytic)

If f, g are analytic on $(a-\varepsilon, a+\varepsilon)$, then:

$$f+g,\ cf \text{ (for any constant } c\text{) and } fg \text{ are all analytic on } (a-\varepsilon, a+\varepsilon)$$

Lemma 6.2.5 (Taylor's Inequality)

Let f be an infinitely differentiable function on $(a-\varepsilon, a+\varepsilon)$. If there exists a sequence C_n such that:

(i) $\left|f^{(n)}(x)\right| \leqslant C_n$ *for all* $x \in (a-\varepsilon, a+\varepsilon)$

(ii) $C_n \dfrac{\varepsilon^n}{n!} \to 0$ *as* $n \to \infty$

Then the function f is analytic on $(a-\varepsilon, a+\varepsilon)$.

For *(i)*; in particular, if this is so for some fixed positive constant $B > 0$, such that for all $n \in \mathbb{N}$, one has:

$$|f^{(n)}(x)| \leqslant B$$

Proof.

$$|R_n(x)| = \left|f^{(n)}(c)\frac{(x-a)^n}{n!}\right| \leqslant C_n \frac{|x-a|^n}{n!} \leqslant C_n \frac{\varepsilon^n}{n!} \to 0 \text{ as } n \to \infty$$

which implies that $R_n(x) \to 0$ as $n \to \infty$. By Analyticity Lemma, this implies analyticity of f on $(a-\varepsilon, a+\varepsilon)$. ■

6.3 Power Series

Let x_k be a sequence of real (or complex) numbers. We make sense of the infinite series $\sum_{k=0}^{\infty} x_k$ formally by defining it to be the limit of the sequence of partial sums $\sum_{k=0}^{n} x_k$ as $n \to \infty$ if the limit exists. We then call $\sum_{k=0}^{\infty} x_k$ to be the sum of the series.

For **notation**, whenever it is clear, we might sometimes denote the series as a shorthand by just writing $\sum x_k$ to mean summing from $k = 0$ to ∞. By $\sum x_k < \infty$, we will mean that the series is convergent. By $\sum x_k = \infty$, we mean that the series is divergent.

6.3.1 Convergence of Power Series

Lemma 6.3.1

Let $\sum_{n=0}^{\infty} x_n$ and $\sum_{n=0}^{\infty} y_n$ be convergent series. Then,

$$\forall a, b \in \mathbb{R}, \text{ the series } \sum_{n=0}^{\infty}(ax_n + by_n) \text{ converges and}$$
$$\sum_{n=0}^{\infty}(ax_n + by_n) = a\sum_{n=0}^{\infty} x_n + b\sum_{n=0}^{\infty} y_n$$

Proof. Direct from definition. ■

Lemma 6.3.2

$$\sum x_k < \infty \iff \forall \varepsilon > 0, \exists N_\varepsilon \text{ such that } \forall n > m > N_\varepsilon, \text{ one has } \left|\sum_{k=m}^{n} x_k\right| < \varepsilon$$

Proof. Suppose $\sum x_k < \infty$. Then, the sequence of partial sums $A_m = \sum_{k=0}^{m} x_k$ is Cauchy. This implies that if we fix $\varepsilon > 0$, we can find an N_ε such that for all $n > m > N_\varepsilon$ we have:

$$|A_n - A_{m-1}| = \left|\sum_{k=0}^{n} x_k - \sum_{k=0}^{m-1} x_k\right| = \left|\sum_{k=m}^{n} x_k\right| < \varepsilon$$

This logic works backwards as well. ■

Here we state an almost similar result to the Comparison Test as a direct corollary.

Corollary 6.3.3

Suppose there exists $c_k > 0$ with $|x_k| < c_k$ for $k > M$ where $M \in \mathbb{N}$. If $\sum c_k < \infty$, then $\sum x_k < \infty$.

Proof. Suppose n, m is sufficiently large so that $n, m > M$ for some $M \in \mathbb{N}$. Now, by the triangle inequality we have:

$$\left|\sum_{k=0}^{n} x_k - \sum_{k=0}^{m-1} x_k\right| = \left|\sum_{k=m}^{n} x_k\right| \leqslant \sum_{k=m}^{n} |x_k| < \sum_{k=m}^{n} c_k$$

for $n, m > M$. Since $\sum c_k < \infty$ by assumption, the previous lemma tells us that you can make the RHS in the above inequality as small as we want whenever we make n, m big. ■

Proposition 6.3.4 (n-th Root Test)

Let $\limsup |w_n|^{1/n} = L$. Then, the series $\sum_{n=0}^{\infty} w_n$ converges absolutely if $L < 1$ and diverges if $L > 1$.

Proposition 6.3.5 (Ratio Test)

Suppose $w_n \neq 0$ for all $n \in \mathbb{N} \cup \{0\}$ and that the sequence $|w_{n+1}|/|w_n|$ converges. Let $L = \lim |w_{n+1}|/|w_n|$. Then, the series $\sum_{n=0}^{\infty}$ converges absolutely if $L < 1$ and diverges if $L > 1$.

6.3.2 Functions defined by Power Series

We make our definition precise first. A real power series is a function $f(x)$ of the form $\sum_{n=0}^{\infty} c_n(x-a)^n$ that spits out real values where $c_n, x, a \in \mathbb{R}$.

> **Question:** When does the power series converges absolutely?

That's not too hard to answer. We go back to definitions to see that the the power series converges absolutely if the series $\sum_{n=0}^{\infty} |c_n||x-a|^n$ converges. In particular, this converges absolutely for $x = a$. Here we see that indeed the power series is the function whose domain is the set of points (in $\mathbb{R}$) such that the limit of the partial sums $\sum_{n=0}^{M} c_n(x-a)^n$ exists pointwise.

A quick notation remark. By $\hat{R}$ we will mean the radius of convergence of a power series. We will give a precise meaning to what does this mean.

Proposition 6.3.6 (When does it converges?)

For any power series $\sum_{n=0}^{\infty} c_n(x-a)^n$, there exists a value $0 \leqslant \hat{R} \leqslant \infty$ such that:

(i) If $0 < \hat{R} < \infty$, the series converges absolutely for x such that $|x-a| < \hat{R}$ and diverges for x such that $|x-a| > \hat{R}$

(ii) If $\hat{R} = \infty$, the series converges absolutely for all $x \in \mathbb{R}$

(iii) If $\hat{R} = 0$, the series is divergent for all $x \neq a$.

Moreover, $\hat{R} = (\limsup |c_n|^{1/n})^{-1}$

Proof of (i). Suppose $0 < \hat{R} < \infty$. Let $w_n = c_n(x-a)^n$. Then, define $L = \limsup |w_n|^{1/n}$. We immediately have $L = |x-a| \limsup |c_n|^{1/n} = C|x-a|$ where we define $C = \limsup |c_n|^{1/n}$. Now we make the observation that:

$$L < 1 \iff |x-a| < C^{-1}, \quad L > 1 \iff |x-a| > C^{-1}$$

Hence, by the n-th Root Test, the series converges absolutely for x such that $|x-a| < C^{-1}$ and diverges for x such that $|x-a| > C^{-1}$. Rename $\hat{R} = C^{-1} = (\limsup |c_n|^{1/n})^{-1}$ to get an identical result to what was stated above. ■

Proof of (ii). Suppose $\hat{R} = \infty$. ■

Proof of (iii). Suppose $\hat{R} = 0$. ■

Definition 6.3.7 (Radius of Convergence)

Suppose the power series $\sum_{n=0}^{\infty} c_n(x-a)^n$ converges for $|x-a| < \hat{R}$ and diverges for $|x-a| > \hat{R}$. Then, we call $0 \leqslant \hat{R} \leqslant \infty$ the radius of convergence of the power series.

It is the largest interval for which the power series can exist.

Theorem 6.3.8 (Root Formula)

Let $f(x) = \sum_{n=0}^{\infty} c_n x^n$ be a power series. We define:

$$\hat{R} = \frac{1}{\limsup |c_n|^{1/n}}$$

Then, $f(x)$ converges absolutely if $|x| < \hat{R}$ and divergent if $|x| > \hat{R}$.

Now, in view of the Ratio Test which we have seen earlier, it is quite natural to ask whether we have another characterization of the radius of convergence. We indeed have one.

Theorem 6.3.9 (Ratio Formula)

Let $f(x) = \sum_{n=0}^{\infty} c_n x^n$ be a power series. Assume that $c_n \neq 0$. We define:

$$\hat{R} = \frac{1}{\limsup |c_{n+1}/c_n|}$$

Then, $f(x)$ converges absolutely if $|x| < \hat{R}$ and divergent if $|x| > \hat{R}$.

Important remark. Our theorems never tell us what is happening at the boundary, i.e. what is happening when $|x - a| = \hat{R}$. Unfortunately, we have to check it by hand as this depends strictly on the coefficients c_n. In general, the power series may converge or diverge there.

Theorem 6.3.10 (Main Theorem of Subsection)

Let $\hat{R} > 0$ be the radius of convergence of the power series

$$f(x) = \sum_{n=0}^{\infty} c_n (x-a)^n$$

Then, $f(x)$ is differentiable on the interval $(a - \hat{R}, a + \hat{R})$ with derivative:

$$f'(x) = \sum_{n=1}^{\infty} n c_n (x-a)^{n-1} = \sum_{n=0}^{\infty} (n+1) c_{n+1} (x-a)^n$$

Corollary 6.3.11 (Power Series is Infinitely Differentiable)

Let $f(x) = \sum_{n=0}^{\infty} c_n (x-a)^n$. Then:

(i) $f(x)$ is infinitely differentiable on the interval $(a - \hat{R}, a + \hat{R})$ given by

$$f^{(k)}(x) = \sum_{n=k}^{\infty} c_n n(n-1)\dots(n-k+1)(x-a)^{n-k} \tag{6.12}$$

(ii) $c_n(a) = \dfrac{f^{(n)}(a)}{n!}$

Together (i) and (ii) implies the Taylor Series:

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!} (x-a)^n$$

Proof of (i). Apply previous theorem to the power series defining $f(x)$ many times. ■

Proof of (ii). Observe that (6.12) can be rewritten as:

$$
\begin{aligned}
f^{(k)}(x) &= \sum_{n=k}^{\infty} c_n n(n-1)\dots(n-k+1)(x-a)^{n-k} \\
&= c_k k! + \sum_{n=k+1}^{\infty} c_n n(n-1)\dots(n-k+1)(x-a)^{n-k}
\end{aligned}
$$

Now set $x = a$ and after a small rearrangement, the result follows. ■

Theorem 6.3.12 (Cauchy Product of Power Series)

Let $f(x) = \sum_{i=0}^{\infty} a_i(x-x_0)^i$ *and* $g(x) = \sum_{j=0}^{\infty} b_j(x-x_0)^j$. *Assume that both* $f(x), g(x)$ *converges on the open interval* $(x_0 - \hat{R}, x_0 + \hat{R})$. *Then,*

$$
f(x)g(x) = \sum_{k=0}^{\infty} c_k(x-x_0)^k \quad \text{where} \quad c_k = \sum_{l=0}^{k} a_l b_{k-l}
$$

and the series converges on the same interval $(x_0 - \hat{R}, x_0 + \hat{R})$.

The above Cauchy Product is sometimes complicated to use and apply. As the indices can be quite arbitrary. A more direct computation using the Cauchy Product is shown in the remark below.

Remark on using the Cauchy Product. Let $f(x) = \sum_{n=0}^{\infty} a_n(x-c)^n$ and $g(x) = \sum_{n=0}^{\infty} b_n(x-c)^n$ be two power series. The power series of the product is then:

$$
\begin{aligned}
f(x)g(x) &= \left(\sum_{n=0}^{\infty} a_n(x-c)^n\right)\left(\sum_{n=0}^{\infty} b_n(x-c)^n\right) \\
&= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} a_i b_j (x-c)^{i+j} \\
&= \sum_{n=0}^{\infty}\left(\sum_{i=0}^{n} a_i b_{n-i}\right)(x-c)^n
\end{aligned}
$$

Theorem 6.3.13 (Sums and Products of Power Series)

Let $\sum a_n x^n$ *and* $\sum b_n x^n$ *be two power series with radius of convergence* R *and* S *respectively. Then:*

(i) The radius of convergence of $\sum (a_n + b_n)x^n$ *is at least* $\min\{R, S\}$

(ii) The radius of convergence of $\sum a_n b_n x^n$ *is at least* RS.

Proof of (i). All the sums below are running from $n = 0$ to ∞. By the triangle inequality, we have:

$$
\sum |a_n + b_n||x|^n \leqslant \sum |a_n||x|^n + \sum |b_n||x|^n
$$

If $\sum |a_n||x|^n < \infty$ and $\sum |b_n||x|^n < \infty$. The Comparison Test implies that the LHS above is $< \infty$. Now observe that:

$$
\begin{aligned}
&\sum |a_n||x|^n < \infty \text{ when } |x| < R \\
&\sum |b_n||x|^n < \infty \text{ when } |x| < S
\end{aligned}
$$

So the LHS converges if $|x| < \min\{R, S\}$. In other words, denote $\hat{R}$ to be the radius of convergence of LHS, then $\hat{R} \geqslant \min\{R, S\}$. ■

Proof of (ii). For $\sum a_n b_n x^n$, we use the n-th root test. Let $\hat{R}$ be the radius of convergence of the sum. Then,

$$\hat{R}^{-1} = \limsup \sqrt[n]{|a_n||b_n|} = \limsup \sqrt[n]{|a_n|}\sqrt[n]{|b_n|} \leqslant \limsup \sqrt[n]{|a_n|} \limsup \sqrt[n]{|b_n|} = R^{-1}S^{-1}$$

This implies that $1/\hat{R} \leqslant 1/(RS)$ which further implies $\hat{R} \geqslant RS$ as required. ■

Important remark for (i). Note that the radius of convergence **must be at least** $\min\{R, S\}$. Being **equal is not sufficient!**

For example, if $a_n = -b_n$, then $\hat{R} = \infty$.

6.3.3 Power Series Expansions for Common Functions

7 Integration I

7.1 A single step closer

We drop our motivation here first. If f is a function that feeds and spits values from $\mathbb{R}$, we intend the integral to be a *measure* of the area under the graph of f. We know how it looks like from prior knowledge of Calculus I. We also know how we'd hope the integral to be like - linear, breakable into pieces, the integral being the anti-differentiating process and further more. Unfortunately, in mathematics, we have to do everything from the very foundational axioms and our bias can only be put aside for now.

Differentiation has been kind, it follows exactly as what we hoped for it to be like. As we will see soon, integration will be the same.

> **Question**: Are all functions integrable?

Along the ways, we will see that it is simply not possible to integrate every function there ever was. There are cases for which the integral is just - infinite. That we cannot assign any real-number value to it. There would also be cases where the area under the graph is infinite. When we start to poke even deeper questions such as asking to integrate over intervals without boundaries or perhaps integrating within a compact interval but which the function blows up to infinity, we will come to new problems. However, with giving the right and proper definition, we will also be able to tame these beasts as well.

We start off with some **notation**. In these last two sections, we will always work on the bounded interval $[a, b]$. By $B[a, b]$, we will mean the space of all bounded functions on $[a, b]$. A bounded function is a function of which we can sandwich it by a positive real constant. By $\mathcal{S}[a, b]$ we will mean the space of step functions. By $R[a, b]$ we will mean the space of Riemann-integrable functions. Both $\mathcal{S}[a, b]$ and $R[a, b]$ will be defined properly in the upcoming subsections. By the norm notation $\|f\|_{C[a,b]}$, we mean the supremum of x over $[a, b]$ of $|f(x)|$ for some function $f \in B[a, b]$. This norm, like all norms, have the triangle inequality property which we will use (and abuse).

7.2 Integrable Steps

We will make use of something called the step functions. They are constant functions. It lives within the world (or space if you are being pedantic) of piecewise continuous functions. Our motivation to use this is simply because they are easy to work with. For example, the difference of two step functions is just the difference of two constants and a 4 year old could understand that.

We will develop its theory with the clear goal at the back of our head that it will be used for defining the so-called Riemann-integrable functions. As a matter of fact, we will define Riemann-integrable functions to be functions that can be approximated very nicely by step functions. In other words, they are the uniform limits of a sequence of step functions.

Definition 7.2.1 (Step Functions)

Let $\psi : [a, b] \to \mathbb{R}$ be a real-valued function. Then, ψ is a step function if:

there exists a finite collection of intervals $I_k \subset [a, b]$ with $k \in \{1, 2, \ldots, N\}$

with the property that:

(1) $\displaystyle\bigcup_{k=1}^{N} I_k = [a, b]$

(2) $j \neq k \implies I_j \cap I_k = \emptyset$

(3) $\forall\, I_k$, ψ is constant

Some important remarks that needs to be brought to attention regarding the previous definition is the following:

Remark 7.2.2 (Remarks on previous definition)

1. The intervals I_k are not necessarily open or closed, may also be degenerate (a singleton).
2. We write $\psi \sim \{I_k, c_k\}$ if ψ is constant on the intervals I_k and takes the value c_k on I_k. Note here that ψ is not a single constant function. Instead, it is a collection of constant functions
3. Hence, every step function ψ is **dependent** on the collection of intervals I_k and constants c_k.
4. If $\psi \sim \{I_k, c_k\}$, we may write c_k as $c_k = \psi|_{I_k}$.

Now, it is not too difficult to see that the step functions form a linear vector space. This is a useful fact that is coherent to our *intuition* which we will abuse later on. Let us denote the linear vector space of step functions as $\mathcal{S}[a, b]$.

Lemma 7.2.3 (Steps Are Linear)

$\mathcal{S}[a, b]$ is a linear space.

Proof. We need to check for linearity. Our first goal is thus to prove:

$$\text{If } \lambda \in \mathbb{R} \text{ and } \psi \in \mathcal{S}[a, b], \text{ then } \lambda\psi \in \mathcal{S}[a, b].$$

Take $\lambda \in \mathbb{R}$ and $\psi \sim \{I_k, c_k\}$. What is $\lambda\psi$? We know it has no effect on I_k as this is just the collection of intervals that partition $[a, b]$ and that remains constant. So only c_k is hit, hence, $\lambda\psi \sim \{I_k, \lambda c_k\} \implies \lambda\psi \in \mathcal{S}[a, b]$. Our second goal is to prove:

$$\psi_1, \psi_2 \in \mathcal{S}[a, b] \implies \psi_1 + \psi_2 \in \mathcal{S}[a, b]$$

This is a bit trickier (but not harder). Take $\psi_1, \psi_2 \in \mathcal{S}[a, b]$:

$$\begin{aligned} \psi_1 &\sim \{I_k, c_k\}, \quad k = 1, 2, \ldots, N \\ \psi_2 &\sim \{\tilde{I}_j, \tilde{c_j}\}, \quad j = 1, 2, \ldots, M \end{aligned}$$

We make our first observation that $I_k \cap \tilde{I}_j$ are disjoint intervals for all k, j. Secondly, we notice that $I_k \cap \tilde{I}_j \subset [a, b]$. So we are still in check, these two observations gives us a sigh of relief that we still have hope to prove this lemma. Furthermore, we have that:

$$\bigcup_{k=1}^{N} \bigcup_{j=1}^{M} (I_k \cap \tilde{I}_j) = [a, b]$$

because $\bigcup\limits_{k=1}^{N} I_k = [a, b]$ and $\bigcup\limits_{j=1}^{M} \tilde{I}_j = [a, b]$. Finally, we have that $\psi_1 + \psi_2$ is constant on each interval $I_k \cap \tilde{I}_j$ and takes the value $c_k + \tilde{c_j}$. We have therefore satisfy all properties of being a Step Function which implies that $\psi_1 + \psi_2 \sim \{I_k \cap \tilde{I}_j, c_k + \tilde{c_j}\}$ is a step function. ■

Our intuitive idea of the end result (the integral as how we know it) being the area under the graph of the function leads us to try doing it for just step functions for now. We might not know how to precisely calculate areas under the graph of x^2 for now (although we might resort to other elementary methods hailing from geometry), but we for sure know how to calculate the area of a rectangle. It is just the rectangle's height multiplied by its width.

We try to extend this idea by doing it for many rectangles that *cover* the interval $[a, b]$ and summing them up. We begin our first attempt in defining the integral.

Definition 7.2.4 (Step Function Integral)

If $\psi \sim \{I_k, c_k\}_{k=1}^N$ is a step function on $[a, b]$, then we define:

$$\int_a^b \psi(x)\ \mathsf{d}x = \sum_{k=1}^{N} c_k \mu(I_k) \tag{7.1}$$

where $\mu(I_k)$ is the length of the interval I_k.

Our intuition served us very well here. Roughly speaking, the definition of integral we have up to this point is:

"Calculate the area of rectangles and summing them all up"

If we want to be more precise but still loose, taking the integral of $\psi \sim \{I_k, c_k\}_{k=1}^N$ over $[a, b]$ is:

"Calculating the area of all N rectangles and sum them all up"

This is not too bad for a start but we can do better. Let $\psi \sim \{I_k, c_k\}$. It is pictorially clear that scaling ψ by a constant (i.e. multiplying c_k by some constant for each k) and then take the integral (multiply scaled c_k by I_k, do this for all k - get rectangles; sum all of them) is equivalent to taking the integral and then scaling the integral itself by the same constant. It is also pictorially clear that if we have two different step functions and we add them and take the integral is equivalent to taking the integral separately and adding them afterwards. We hope this is true if we do this mathematically because this is how we imagined our final integral would be. If it is not, we might have to find a better definition. Spoiler alert: fortunately, this is true.

Lemma 7.2.5 (Linearity of the Step Function Integral)

(1) If $\psi \in \mathcal{S}[a, b], \lambda \in \mathbb{R}$, then:

$$\int_a^b \lambda\psi(x)\ dx = \lambda \int_a^b \psi(x)\ dx$$

(2) If $\psi_1, \psi_2 \in \mathcal{S}[a, b], \lambda \in \mathbb{R}$, then:

$$\int_a^b \psi_1(x) + \psi_2(x)\ dx = \int_a^b \psi_1(x)\ dx + \int_a^b \psi_2(x)\ dx$$

Proof. The proof of (1) is trivial. If $\psi \sim \{I_k, c_k\}$, then $\lambda\psi \sim \{I_k, \lambda c_k\}$. The rest is exercise.

Proof of (2). Let $\psi_1 \sim \{I_k, c_k\}_{k=1}^N$ and $\psi_2 \sim \{\tilde{I}_k, \tilde{c_k}\}_{k=1}^{\tilde{N}}$. Now notice we can partition the interval $[a, b]$ as:

$$\bigcup_{k=1}^{N} \bigcup_{j=1}^{M} (I_k \cap \tilde{I}_j) = [a, b]$$

Then, take the integral of the sum of the two step functions:

$$\begin{aligned}\int_a^b (\psi_1(x)+\psi_2(x))\ \mathrm{d}x &= \sum_{k=1}^{N}\sum_{j=1}^{M}(c_k+\tilde{c_j})\,\mu(I_k\cap\tilde{I}_j)\\ &= \sum_{k=1}^{N}c_k\sum_{j=1}^{M}\mu(I_k\cap\tilde{I}_j)+\sum_{j=1}^{M}\tilde{c_j}\sum_{k=1}^{N}\mu(I_k\cap\tilde{I}_j)\\ &= \sum_{k=1}^{N}c_k\,\mu(I_k)+\sum_{j=1}^{M}\tilde{c_j}\,\mu(\tilde{I}_j)\\ &= \int_a^b \psi_1(x)\ \mathrm{d}x+\int_a^b \psi_2(x)\ \mathrm{d}x\end{aligned}$$

The second last equality is true because $\mu(I_k)=\sum_{j=1}^{M}\mu(I_k\cap\tilde{I}_j)$ and $\mu(\tilde{I}_j)=\sum_{k=1}^{N}\mu(I_k\cap\tilde{I}_j)$. This is non-obvious but can be seen quite clearly through a picture. ■

Linearity of the integral, checked! We end this subsection with something I call the "Obvious Estimate Lemma". This lemma is true because we are dealing with a bounded interval (both endpoints are real numbers). There will be problems if the interval is unbounded, for example $[3,\infty)$. We will see why the implication is false in such cases when we traverse through the proof.

The idea of the lemma is saying that we can estimate the integral of ψ over $[a,b]$ by the *area of the largest rectangle* that contains the integral. This estimate is obvious (hence, the name) if you see it on a picture. Now, the *largest rectangle's area* is rather vague here but we will make it precise.

Since step functions are piecewise continuous, and continuous functions on compact intervals attains a maximum somewhere. We define the *largest rectangle's area* to be the length of the interval $[a,b]$ which is $b-a$ multiplied by the maximum of this step function. But since in a compact interval maximums are just equivalent to supremums, using the beautiful language of norms that we have defined earlier, we define the *largest rectangle's area* to be $(b-a)\,\|\psi\|_{C[a,b]}$.

Lemma 7.2.6 (Obvious Estimate Lemma)

If $\psi\in\mathcal{S}[a,b]$, then:

$$\left|\int_a^b \psi(x)\ dx\right|\leqslant(b-a)\,\|\psi\|_{C[a,b]}$$

Proof. Let $\psi\sim\{I_k,c_k\}$. For any k, we have $|c_k|\leqslant\|\psi\|_{C[a,b]}$. Then,

$$\left|\int_a^b \psi(x)\ \mathrm{d}x\right| = \left|\sum_{k=1}^{N}c_k\,\mu(I_k)\right|\leqslant\sum_{k=1}^{N}|c_k|\mu(I_k)\leqslant\|\psi\|_{C[a,b]}\sum_{k=1}^{N}\mu(I_k)=\|\psi\|_{C[a,b]}\,(b-a)$$

The first inequality we have is just the triangle inequality. ■

As promised, we will discuss why this is false for unbounded intervals. If we revisit the proof, all the inequalities make use of the fact that $\sum\mu(I_k)=b-a$ is finite. If we have that the interval is infinite, then $b-a$ is infinite and we could not have estimated the integral as simply as we had done.

7.3 Riemann-integrable Functions

Definition 7.3.1 (Integrable in Riemann's fashion)

(1) Let $f \in B[a,b]$. f is said to be **Riemann-integrable** if there exists $\{\psi_n\}_{n=1}^{\infty} \subset \mathcal{S}[a,b]$ such that:

$$\|f - \psi_n\|_{C[a,b]} \longrightarrow 0 \quad \text{as} \quad n \to \infty$$

(2) Let f be Riemann-integrable, and let $\{\psi_n\}_{n=1}^{\infty}$ be as in (1) above. Then, we define:

$$\int_a^b f(x) \, \mathrm{d}x = \lim_{n\to\infty} \int_a^b \psi_n(x) \, \mathrm{d}x$$

Notation: If f satisfies Definition (1), we write $f \in R[a,b]$.

Whenever we define something, be it some linear map or a group isomorphism, we have to check that it is always well-defined first. That is, we have to check what we define is correct - it maps to the things that we say it maps to; and that we are not cheating. One obvious question from the definition above is the definition correct at all? Another less obvious question is that are all bounded functions (Riemann-) integrable?

Questions: **(A)** Is Definition (2) correct? **(B)** What kind of f's are Riemann-integrable?

Non-Example 7.3.2 ($\exists f \in B[a,b]$ that cannot be approximated by Step Functions)

We recall our old friend the Dirichlet function (which is bounded) defined by:

$$f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \notin \mathbb{Q} \end{cases}$$

To see this clearly, we have to go to the world of pictures. Observe that however you choose ψ_k,

$$\|f - \psi_k\|_{C[a,b]} \geqslant \frac{1}{2}$$

So there does not exists a sequence of step functions $\{\psi_k\}$ such that it approximates f uniformly.

Lemma 7.3.3 (Riemann-integrable Functions Are Linear)

$R[a,b]$ is a linear space.

Proof. Exercise. ■

Theorem 7.3.4 (Continuous functions are Riemann-integrable)

$$C[a,b] \subset R[a,b]$$

Ingredient: **(1)** Cantor's powerful theorem and **(2)** $\varepsilon - \delta$ definition of uniform continuity

Proof. Let $f \in C[a,b]$ be given and let $\varepsilon > 0$.

Goal: Need to find $\psi \in \mathcal{S}[a,b]$ such that it approximates f uniformly on $[a,b]$

Here come's Cantor's theorem. Since f is continuous on the compact interval $[a,b]$. Then

f is uniformly continuous on $[a, b]$. This means that there is a $\delta > 0$ such that $|x - y| < \delta$ implies $|f(x) - f(y)| < \varepsilon$ (†). The next step is not obvious without a picture.

Let us divide $[a, b]$ into finite intervals $\{I_k\}_{k=1}^N$ such that $\mu(I_k) < \delta$ for any $\delta > 0$ for all k (meaning we can make the *length* of the interval I_k sufficiently small). Now for any k, define $\psi|_{I_k} = f(x_k)$ where $x_k \in I_k$. We are so close to the result as we can just rewrite (†) with $y = x_k$ for some $x_k \in I_k$. Then, whenever $|x - x_k| < \delta$, we have:

$$|f(x) - \phi(x)| = |f(x) - f(x_k)| < \varepsilon$$

Taking supremum on $[a, b]$ on both sides of the inequality complete the proof as:

$$\|f(x) - \psi(x)\|_{C[a,b]} < \sup \varepsilon = \varepsilon$$

■

Definition 7.3.5

$f \in B[a, b]$ is called piecewise continuous if:

$$[a, b] = \bigcup_{k=1}^{N} I_k \text{ such that:}$$

(i) $I_j \cap I_k = \emptyset$ for $j \neq k$;

(ii) f continuous on each I_k and has finite limits at the endpoints of I_k.

Piecewise continuous = "Finitely-many jump discontinuities"

Theorem 7.3.6

If f is piecewise continuous on $[a, b]$, then $f \in R[a, b]$.

Proof. Exercise. ■

7.4 Extending the integral to $R[a, b]$

Lemma 7.4.1

Let $f \in R[a, b], \{\psi_n\}_{n=1}^{\infty} \subset \mathcal{S}[a, b]$ such that $\|f - \psi_n\|_{C[a,b]} \to 0$ as $n \to \infty$. Then:

$$\int_a^b \psi_n(x)\ dx \text{ converges as } n \to \infty$$

This lemma features the interplay between (uniform) convergence of functions and the convergence of numbers (the integral).

Ingredient: **(1)** Cauchy sequence definition applied to $s_n = \int \psi_n$, **(2)** Obvious Estimate Lemma (7.2.6) with ^+f-manipulation, **(3)** Triangle inequality for norms.

Proof.

$$
\begin{aligned}
|s_n - s_m| &= \left| \int_a^b \psi_n(x)\ \mathrm{d}x - \int_a^b \psi_m(x)\ \mathrm{d}x \right| \\
&= \left| \int_a^b (\psi_n(x)\ \mathrm{d}x - \psi_m(x))\ \mathrm{d}x \right| \\
&\leqslant (b-a)\|\psi_n - \psi_m\|_{C[a,b]} \\
&= (b-a)\|(\psi_n - f) + (f - \psi_m)\|_{C[a,b]} \\
&\leqslant (b-a)\left\{ \|\psi_n - f\|_{C[a,b]} + \|f - \psi_m\|_{C[a,b]} \right\}
\end{aligned}
$$

Choose N_ε sufficiently large and we are done. ■

Proposition 7.4.2

Let (i) $f \in R[a,b]$, *(ii)* $\{\psi_n\}_{n=1}^{\infty}, \{\phi_n\}_{n=1}^{\infty} \subset \mathcal{S}[a,b]$ *such that* $\|f - \psi_n\|_{C[a,b]} \to 0$ *and* $\|f - \phi_n\|_{C[a,b]} \to 0$. *Then:*

$$
\lim_{n\to\infty} \int_a^b \psi_n(x)\ dx = \lim_{n\to\infty} \int_a^b \phi_n(x)\ dx
$$

Phrased differently, if we have a Riemann-integrable function f and two (can be same or distinct) sequences of step functions such that both these sequences approximates f uniformly. Then, the limits of the step integral has to be the same - no other choice.

This proof is *isomorphic* to the proof the previous lemma.

Proof. Let:

$$
\alpha_n = \int_a^b \psi_n(x)\ \mathrm{d}x, \quad \beta_n = \int_a^b \phi_n(x)\ \mathrm{d}x
$$

By previous Lemma (7.4.1), we know that $\lim \alpha_n$ and $\lim \beta_n$ exists. We then estimate the difference of the two integrals in question. Some steps below are *proof-isomorphic* to the previous lemma's so they will be omitted.

$$
|\alpha_n - \beta_n| = \left| \int_a^b \psi_n(x)\ \mathrm{d}x - \int_a^b \phi_n(x)\ \mathrm{d}x \right| \leqslant (b-a)\left\{ \|\psi_n - f\|_{C[a,b]} + \|f - \phi_n\|_{C[a,b]} \right\} \to 0
$$

Hence, we get that:

$$
\lim(\alpha_n - \beta_n) = 0 \implies \lim \alpha_n = \lim \beta_n
$$

■

8 Integration II: Further properties

8.1 Identities for the Integral on $R[a, b]$

In this subsection, we develop all the nice properties of integration as we know it. We start with showing that the integral is linear.

Proposition 8.1.1 (Linearity on $R[a, b]$)

The map $f \mapsto \int_a^b f(x)\, dx$ is linear on $R[a, b]$

Proving linearity of the integral is equivalent to showing that for any $f, g \in R[a, b]$ and $\lambda, \mu \in \mathbb{R}$:

$$\int_a^b (\lambda f + \mu g)\ \mathrm{d}x = \lambda \int_a^b f\ \mathrm{d}x + \mu \int_a^b g\ \mathrm{d}x \tag{8.1}$$

Ingredients: (1) Definition of Riemann-integrability (both (1) and (2)) **(2)** Triangle inequality for norms

Proof. Let $\lambda, \mu \in \mathbb{R}$. Take $f, g \in R[a, b]$. By Definition (7.3.1), there exists sequences of step functions ψ_n and ϕ_n that estimates f and g uniformly. Automatically, we have from Definition (7.3.1):

$$\int_a^b \psi_n(x)\ \mathrm{d}x \to \int_a^b f(x)\ \mathrm{d}x, \quad \int_a^b \phi_n(x)\ \mathrm{d}x \to \int_a^b g(x)\ \mathrm{d}x \quad (\dagger)$$

$$\textbf{First Goal: Show that } \int_a^b \lambda\psi_n + \mu\phi_n\ \mathbf{d}x \longrightarrow \int_a^b \lambda f + \mu g\ \mathbf{d}x \quad (*)$$

$$\begin{aligned} \|\lambda f + \mu g - (\lambda\psi_n + \mu\phi_n)\|_{C[a,b]} &= \|\lambda(f - \psi_n) + \mu(g - \phi_n)\|_{C[a,b]} \\ &\leqslant |\lambda|\, \|f - \psi_n\|_{C[a,b]} + |\mu|\, \|g - \phi_n\|_{C[a,b]} \\ &\longrightarrow 0 \quad (\psi_n \text{ approximates } f, \text{ same for } \phi_n \text{ for } g) \end{aligned}$$

Good, **first goal done**. Now, Lemma (7.2.5) tells us that step function integrals are linear:

$$\int_a^b \lambda\psi_n(x) + \mu\phi_n(x)\ \mathrm{d}x = \lambda \int_a^b \psi_n(x) + \mu \int_a^b \phi_n(x)\ \mathrm{d}x$$

The LHS converges to $(*)$. The RHS converges to $(\dagger)$. We can stop now. ■

Next, we prove what seems to be one of the most intuitive idea about the integral, that is - it is breakable into parts. We will prove the fact that it can be broken into two parts, but if we iterate this process, we can break the integral into many-many parts.

Proposition 8.1.2 (Integral is Breakable)

Let $f \in R[a, b]$ and $c \in [a, b]$. Then:

$$\int_a^b f(x)\ dx = \int_a^c f(x)\ dx + \int_c^b f(x)\ dx \tag{8.2}$$

Proof. Here are our steps. **(1)** We prove this for step functions first (as we haven't done this before). Then, **(2)** we take limits.

Step 1: Let $\psi \sim \{I_k, c_k\}_{k=1}^N$ and here, $\psi|_{I_k} = c_k$. Define $I_j = I_j' \cup I_j''$ where $I_j' = I_j \cap (-\infty, c]$

and $I_j'' = I_j \cap (c, \infty)$ (this is clearer from a picture). Notice that by defining it this way, we have that:

$$\mu(I_j) = \mu(I_j') + \mu(I_j'')$$

Now, choose any $c \in I_j$. Then:

$$\begin{aligned}
\int_a^b \psi(x)\ \mathrm{d}x = \sum_{k=1}^N c_k\mu(I_k) &= \sum_{k=1}^{j-1} c_k\mu(I_k) + {\color{red}c_j\mu(I_j)} + \sum_{k=j+1}^{N} c_k\mu(I_k) \\
&= \left(\sum_{k=1}^{j-1} c_k\mu(I_k) + {\color{red}c_j\mu(I_j')}\right) + \left({\color{red}c_j\mu(I_j'')} + \sum_{k=j+1}^{N} c_k\mu(I_k)\right) \\
&= \int_a^c \psi(x)\ \mathrm{d}x + \int_c^b \psi(x)\ \mathrm{d}x
\end{aligned}$$

Step 2: From Step 1, we have:

$$\int_a^b \psi(x)\ \mathrm{d}x = \int_a^c \psi(x)\ \mathrm{d}x + \int_c^b \psi(x)\ \mathrm{d}x$$

Take limits on both sides. We can stop now because $f \in R[a,b]$ and $c \in [a,b]$ so f is Riemann-integrable in smaller intervals $\subset [a,b]$ and hence, both sides are convergent - and they converge to what we want. ■

8.2 Inequalities for the Integral on $R[a,b]$

Lemma 8.2.1

Let $\alpha \in \mathbb{R}$. Define $\alpha_+ = \max\{\alpha, 0\}$. Then, $|\alpha_+ - \beta_+| \leqslant |\alpha - \beta|$.

The proof is an exercise. Note that the definition of $\alpha_+ = \max\{\alpha, 0\}$ can be written equivalently as:

$$\alpha_+ = \begin{cases} \alpha & \text{if } \alpha \geqslant 0 \\ 0 & \text{if } \alpha < 0 \end{cases}$$

From this definition, it should not be too difficult to prove the above lemma. Just consider all the 4 cases plus the case where both $\alpha_+ = \beta_+ = 0$. We will need this lemma to prove the next lemma.

Lemma 8.2.2

If $f \in R[a,b]$ is non-negative for all $x \in [a,b]$. Then:

$$\int_a^b f(x)\ dx \geqslant 0$$

This is quite intuitive, and we will need it to prove more intuitive things ahead.

Proof. In the words of Pushnitski, the idea of the proof in one line is:

"Let me try to not make my approximation worse."

More precisely speaking, we already have that $f \in R[a,b]$, so $\|f - \psi_n\|_{C[a,b]} \to 0$ for some sequence of step function ψ_n. Let's try to not make this approximation worse by trying something better - showing $\|f-(\psi_n)_+\|_{C[a,b]} \to 0$ where the subscript of $+$ is similar to the notation we proved in the previous lemma. Now, for any $x \in [a,b]$, we have:

$$|f(x)-(\psi_n(x))_+| = |(f(x))_+ - (\psi_n(x))_+| \leqslant |f(x)-\psi_n(x)|$$

Take supremum on both sides over $[a,b]$ to have:

$$\|f-(\psi_n)_+\|_{C[a,b]} \leqslant \|f-\psi_n\|_{C[a,b]}$$

We assumed $f \in R[a,b]$, so the RHS goes to 0. Hence, $(\psi_n)_+$ converges to f. Since $(\psi_n)_+ \geqslant 0$,

$$\int_a^b f(x)\ \mathrm{d}x = \lim_{n\to\infty} \int_a^b ((\psi_n(x))_+\ \mathrm{d}x \geqslant 0$$

■

The proof seems to be a bit more involved. The most interesting tool used in it is probably our definition of α_+. It seems to come out of nowhere but the natural intuition kicks in if one sees what we are trying to prove pictorially. The hard work we've put in the early bit of this section up to this point can finally be celebrated for once as we move on to the next proposition and its proof.

Proposition 8.2.3 (Monotonicity of the Integral)

Let $f, g \in R[a,b]$ and $f(x) \leqslant g(x)$ for all $x \in [a,b]$. Then:

$$\int_a^b f(x)\ dx \leqslant \int_a^b g(x)\ dx$$

Proof. Consider the function $g(x) - f(x) \geqslant 0$. Apply the previous lemma and use linearity of integral. ■

You could probably understand what I mean before by now. Well it's a chain reaction from now on. Let's try to form some estimates for our well-defined integral.

Lemma 8.2.4 (Estimating the Integral I)

If $f \in R[a,b]$ and $m \leqslant f(x) \leqslant M$ for all $x \in [a,b]$. Then:

$$m(b-a) \leqslant \int_a^b f(x)\ dx \leqslant M(b-a)$$

Proof. By assumption $f(x) \leqslant M$. Apply previous proposition and done. For the the lower bound, the proof is the same. ■

The following next two lemmas are probably the most important estimate of the integral so far. The first next lemma is the intuitive notion saying that the integral is less than the area of the largest rectangle that *contains* the function. This is the Riemann-integrable analogue of our Obvious Estimate Lemma. Turns out, it is true in $R[a,b]$ as well. The second next lemma is the triangle inequality for integrals - this one is of immense important. We will use it to prove the Fundamental Theorem of Calculus.

Lemma 8.2.5 (Estimating the Integral II - Obvious Estimate Lemma on $R[a,b]$)

If $f \in R[a,b]$, then:

$$\left| \int_a^b f(x)\ dx \right| \leqslant (b-a)\|f\|_{C[a,b]} \tag{8.3}$$

Proof. Let M in the previous lemma be defined by $M = \|f\|_{C[a,b]}$. Apply the lemma's result. Similar for lower bound, let $m = -\|f\|_{C[a,b]}$. Combine the two and done. ■

Lemma 8.2.6 (Estimating the Integral III - Triangle Inequality)

Let $f \in R[a,b]$, then $|f| \in R[a,b]$ and

$$\left| \int_a^b f(x)\ dx \right| \leqslant \int_a^b |f(x)|\ dx \tag{8.4}$$

Proof. For the first part - showing $|f| \in R[a,b]$. This is straightforward from the reversed triangle inequality:

$$\left|\,|f| - |\psi_n|\,\right| \leqslant |f - \psi_n|$$

Take supremum on both sides and by definition of $f \in R[a,b]$, the RHS goes to 0 and **we are done**. Now to the real business, $f(x) \leqslant |f(x)|$ for all $x \in [a,b]$. Integral is monotonous, so apply Proposition (8.2.3) to get:

$$\int_a^b f(x)\ \mathrm{d}x \leqslant \int_a^b |f(x)|\ \mathrm{d}x$$

Similarly, $-f \leqslant |f|$, again apply the same proposition. Combine the result. ■

The above lemma is called the *triangle inequality for integrals* because it is an analogue of the generalized triangle inequality for sums. We have, for sums:

$$\left| \sum_{k=1}^n a_k \right| \leqslant \sum_{k=1}^n |a_k|$$

Observe the similarity with the above inequality (8.4). A useful analyst insight from Prof. Pushnitski is that in checking inequalities for integrals, one should ask whether *is it true for sums* first. If it is not true, it is usually the case that it is not true for integrals as well.

8.3 Fundamental Theorem of Calculus

We now go to prove the so-called Fundamental Theorem of Calculus which says that integration is the process of anti-differentiation and vice-versa. From this theorem, we will develop more facts that we hoped to be true (as they are intuitively true) but with more careful treatment.

Theorem 8.3.1 (FTC I)

Let $F(x) = \int_a^x f(s)\ ds$. If $f \in R[a,b]$ is continuous at $x \in (a,b)$, then:

$F(x)$ is differentiable at x and $F'(x) = f(x)$.

Proof. Let $\delta > 0$. We observe that:

$$\begin{aligned}\frac{F(x+\delta)-F(x)}{\delta} &= \frac{1}{\delta}\left\{\int_a^{x+\delta} f(s)\ \mathrm{d}s - \int_a^x f(s)\ \mathrm{d}s\right\} \\ &= \frac{1}{\delta}\left\{\int_a^x f(s)\ \mathrm{d}s + \int_x^{x+\delta} f(s)\ \mathrm{d}s - \int_a^x f(s)\ \mathrm{d}s\right\} \\ &= \frac{1}{\delta}\int_x^{x+\delta} f(s)\ \mathrm{d}s\end{aligned}$$

Now, take the difference with $f(x)$ (note here that x is fixed, so $f(x)$ is a constant):

$$\begin{aligned}\frac{F(x+\delta)-F(x)}{\delta} - f(x) &= \frac{1}{\delta}\int_x^{x+\delta} f(s)\ \mathrm{d}s - f(x)\cdot \frac{1}{\delta}\int_x^{x+\delta}\ \mathrm{d}s \\ &= \frac{1}{\delta}\int_x^{x+\delta} (f(s)-f(x))\ \mathrm{d}s\end{aligned}$$

The one highlighted in red is just $=1$. If we take the modulus here and use the triangle inequality for integrals:

$$\left|\frac{F(x+\delta)-F(x)}{\delta} - f(x)\right| \leqslant \frac{1}{\delta}\int_x^{x+\delta} |f(s)-f(x)|\ \mathrm{d}s$$

Remember we assumed that f is continuous at x. So there is a δ such that whenever $0<|x-s|<\delta$, then $|f(x)-f(s)|<\varepsilon$ for any $\varepsilon>0$. Then:

$$\frac{1}{\delta}\int_x^{x+\delta} |f(s)-f(x)|\ \mathrm{d}s \leqslant \frac{\varepsilon}{\delta}\int_x^{x+\delta}\ \mathrm{d}s = \varepsilon$$

Notice that we have not yet proved the complete statement as we have only shown this is true for $\delta \to 0^+$. Fortunately, proving the case for $\delta \to 0^-$ is very similar. ■

The proof looks way simpler than how it seems. It is lengthy because we give separate treatments when finding the upper bound of $|(F(x+\delta)-F(x))/\delta - f(x)|$ which can actually be done in one swoop.

Theorem 8.3.2 (FTC II)

Suppose $f:[a,b]\to\mathbb{R}$ is continuous and has continuous derivative f' on $[a,b]$. Then:

$$f(x) = f(a) + \int_a^x f'(s)\, ds, \quad \forall x\in[a,b] \tag{8.5}$$

Proof. The clever thing about this proof is this. Let:

$$G(x) = f(a) + \int_a^x f'(s)\ \mathrm{d}s - f(x) \tag{8.6}$$

FTC I tells us that $G'(x)=0$ everywhere. By **MVT**, $G(x)$ is constant everywhere on $[a,b]$. For example, evaluate at $x=a$ to see that $G(a)=0$. Hence, $G(x)$ is zero everywhere. ■

The proof utilizes a corollary to the Mean Value Theorem, that is if f differentiable on $[a,b]$ and its derivative is 0 everywhere, then f has no choice but to be some constant function in $\mathbb{R}$. FTC II has an interesting corollary.

Corollary 8.3.3 (Definite Integral of Zero is Zero)

The definite integral of zero is zero. In other words, let $[a,b] \subset \mathbb{R}$. Then,

$$\int_a^b 0 \; dx = 0, \quad \forall x \in [a,b]$$

Proof. Let $f(x) = C$ for all $x \in [a,b] \subset \mathbb{R}$ for some constant $C \in \mathbb{R}$. Indeed f is continuous and has a continuous derivative on $[a,b]$. Then, **FTC II** tells us that for all $x \in [a,b]$,

$$C = C + \int_a^x 0 \; \mathrm{d}s$$

and hence, the result. ∎

8.4 Integrating convergent sequences of functions

At this point, we have basically recovered almost all (if not all) of the results that we encountered before. We now move into uncharted waters, where it was inaccessible in Calculus I before.

Theorem 8.4.1 (Passing Limit to Integrals)

Let $F_n, F \in R[a,b]$ for all $n \in \mathbb{N}$ and $||F_n - F||_{C[a,b]} \to 0$ as $n \to \infty$. Then:

$$\int_a^b F_n(x) \; dx \longrightarrow \int_a^b F(x) \; dx$$

In other words,

$$\lim_{n\to\infty} \int_a^b F_n(x) \; dx = \int_a^b \lim_{n\to\infty} F_n(x) \; dx$$

It **is important that F_n converges to F uniformly** i.e. $||F_n - F||_{C[a,b]} \to 0$. It is also important that the **integration interval is finite**. We will see some examples further down for why this is the case but we will first prove the theorem.

Proof. The proof is very simple. It uses the triangle inequality and the Obvious Estimate Lemma.

$$\begin{aligned} \left| \int_a^b F_n(x) \; \mathrm{d}x - \int_a^b F(x) \; \mathrm{d}x \right| &= \left| \int_a^b (F_n(x) - F(x)) \; \mathrm{d}x \right| \\ &\leqslant \int_a^b |F_n(x) - F(x)| \; \mathrm{d}x \\ &\leqslant (b-a)||F_n(x) - F(x)||_{C[a,b]} \end{aligned}$$

∎

What happens if we relax some of the conditions that we have mentioned as important?

Non-Example 8.4.2 (Interval Is Infinite)

Let the sequence of function $F_n(x)$ be defined by:

$$F_n(x) = \begin{cases} 1/n & \text{if } x \in [0,n] \\ 0 & \text{otherwise} \end{cases}$$

If we sketch this graph, we would get a step function between 0 and n which is elsewhere 0. Now, note that $||F_n(x)||_{C(\mathbb{R})} = 1/n \to 0$ as $n \to \infty$. So this is OK. But then:

$$\lim_{n\to\infty} \int_{\mathbb{R}} F_n(x) \, \mathrm{d}x = 1$$

and if we take limits inside:

$$\int_{\mathbb{R}} \lim_{n\to\infty} F_n(x) \, \mathrm{d}x = 0$$

The problem here is that the interval is infinite. If we go back to the proof, **the proof fails exactly at the last inequality** (and elsewhere is OK), **where we estimate the difference by** "$\leqslant (b-a)$(the norm)". If the interval is infinite, $b-a$ has to be infinite there, so we can't do this estimate simply as what we did.

How about the case if we assume it is OK to have, just pointwise convergence?

Non-Example 8.4.3 (Convergence is Not Uniform)

Let the sequence of function $f_n(x)$ be defined by:

$$f_n(x) = nxe^{-nx^2} \text{ for } x \in [0,1]$$

Then, we would have $f_n(x) \to 0$ as $n \to \infty$ for all $x \in [0,1]$ i.e. we have pointwise convergence. However, with the aid of a picture, we can easily see that $||f_n(x)||_{C[0,1]} \to \infty$ i.e. $f_n(x)$ converge pointwise but not uniformly. As a consequence, the theorem does not hold:

$$\lim_{n\to\infty} \int_0^1 f_n(x) \, \mathrm{d}x = \lim_{n\to\infty} \left(\frac{1-e^{-n}}{2}\right) = \frac{1}{2}$$

but then if we take the limits inside:

$$\int_0^1 \lim_{n\to\infty} f_n(x) \, \mathrm{d}x = \int_0^1 0 \, \mathrm{d}x = 0$$

and they are not equal.

Now, we are going to apply the previous theorem to series and get the neat result that we can swap summation with integration.

Theorem 8.4.4 (Term-by-Term Integration)

Let $f_k \in R[a,b]$ for all $k \in \mathbb{N}$. Let $F \in R[a,b]$. Assume that $\sum_{k=1}^{\infty} f_k(x)$ converges uniformly to $F(x)$, i.e.

$$\left\| F - \sum_{k=1}^{n} f_k(x) \right\|_{C[a,b]} \longrightarrow 0 \text{ as } n \to \infty$$

Then,

$$\sum_{k=1}^{\infty} \int_a^b f_k(x) \, dx = \int_a^b F(x) \, dx$$

In other words, we can swap integrals with summations:

$$\sum_{k=1}^{\infty} \int_a^b f_k(x) \, dx = \int_a^b \left(\sum_{k=1}^{\infty} f_k(x) \, dx \right)$$

Proof. Apply Theorem (8.4.1) to $F_n = \sum_{k=1}^{n} f_k(x)$. ∎

We now move to prove a result that we have actually stated before but which the proof was postponed until now.

Corollary 8.4.5 (Term-by-Term Differentiation)

Let f_n be continuously differentiable functions on $[a, b]$ such that:

(1). $\sum_{n=1}^{\infty} f_n(x) = f(x), \quad \forall x \in [a, b]$

(2). $\sum_{n=1}^{\infty} f_n'(x)$ *is* ***uniformly convergent***

Then, f is continuously differentiable and $f'(x) = \sum_{n=1}^{\infty} f_n'(x)$.

In other words, we can swap differentiation with summations:

$$\frac{\mathrm{d}}{\mathrm{d}x}\sum_{n=1}^{\infty} f_n(x) = \sum_{n=1}^{\infty}\frac{\mathrm{d}}{\mathrm{d}x} f_n(x)$$

Proof. Main ingredients of the proof are: FTC I & II and the recently proved Theorem (8.4.4) Term-by-Term Integration.

For simplicity, denote $\tilde{f}(x) = \sum_{n=1}^{\infty} f_n'(x)$. Consider the integral from a to x of $\tilde{f}(x)$:

$$\int_a^x \tilde{f}(s)\ \mathrm{d}s = \int_a^x \left(\sum_{n=1}^{\infty} f_n'(s)\right)\ \mathrm{d}s = \sum_{n=1}^{\infty}\int_a^x f_n'(s)\ \mathrm{d}s = \sum_{n=1}^{\infty}(f_n(x) - f_n(a)) = f(x) - f(a) \tag{8.7}$$

Since each f_n' is continuous and convergence of $\sum f_n'$ is uniform (by assumptions of the theorem), $\tilde{f} = \sum_{n=1}^{\infty} f_n'$ is continuous. This is nice as we can apply FTC I now:

$$\frac{\mathrm{d}}{\mathrm{d}x}\int_a^x \tilde{f}(s)\ \mathrm{d}s = \tilde{f}(x)$$

Also from (8.7), we have that:

$$\tilde{f}(x) = \frac{\mathrm{d}}{\mathrm{d}x}(f(x) - f(a)) = f'(x) - 0 = f'(x)$$

Recalling what $\tilde{f}(x)$ means recover our wanted conclusion. ∎

8.5 Improper Integrals I: Unbounded Intervals

The aim of this subsection is to make sense of objects of the form:

$$\int_0^{\infty} f(x)\ \mathrm{d}x, \quad \int_{-\infty}^{1} f(x)\ \mathrm{d}x, \quad \int_{-\infty}^{\infty} f(x)\ \mathrm{d}x$$

To tame these infinities, we shall make the notion of integration on unbounded intervals precise.

Definition 8.5.1 (Improper Integrals)

Let $f \in R[a, b]$ for all $a < b$. Then we define:

(i)
$$\int_a^\infty f(x)\ \mathrm{d}x = \lim_{b\to\infty}\int_a^b f(x)\ \mathrm{d}x \tag{8.8}$$

(ii)
$$\int_{-\infty}^b f(x)\ \mathrm{d}x = \lim_{a\to-\infty}\int_a^b f(x)\ \mathrm{d}x \tag{8.9}$$

(iii)
$$\int_{-\infty}^\infty f(x)\ \mathrm{d}x = \int_{-\infty}^0 f(x)\ \mathrm{d}x + \int_0^\infty f(x)\ \mathrm{d}x \tag{8.10}$$

Remark 8.5.2

1. The existence of the limits in the definition above is not immediate. If the limit exist, we say that these *improper integrals* converge/well-defined/exists.
2. It is important that in (8.10), we take **separate limits**. It may happen that:
$$\lim_{b\to\infty}\int_{-b}^b f(x)\ \mathrm{d}x \text{ exists, yet } \int_{-\infty}^\infty f(x)\ \mathrm{d}x \text{ diverges.}$$
In other words, for (8.10) to exists, both limits that sum up to (8.10) must exists.
3. Another remark to (8.10). Instead of integrating over $(-\infty, 0)$ and $(0, \infty)$, we could have integrated over $(-\infty, c)$ and (c, ∞) for any $c \in \mathbb{R}$. The integral is independent of the choice of this point.

Example 8.5.3 (Why (2) is important to note in the previous remark)

Let $f(x) = \sin x$. We have that:

$$\lim_{b\to\infty}\int_{-b}^b \sin x\ \mathrm{d}x = 0$$

However, we have (both) the individual sums that form the above integral to be divergent:

$$\lim_{b\to\infty}\int_0^b \sin x\ \mathrm{d}x = 1 - \cos b$$
$$\lim_{b\to\infty}\int_{-b}^0 \sin x\ \mathrm{d}x = -1 + \cos b$$

It turns out that their sum cancels out each other, resulting in 0, which tricked us into thinking the improper integral over $(-\infty, \infty)$ is convergent.

Example 8.5.4 (For which α is the improper integral defined?)

Let $\alpha > 0$. Consider the integral:

$$I(b) := \int_1^b \frac{\mathrm{d}x}{x^\alpha} \tag{8.11}$$

and define $I(\infty) = \lim_{b\to\infty} I(b)$. By this definition, $I(\infty)$ is an improper integral. Then, for which α is the improper integral defined? We have to check separate cases.

(1) If $\alpha = 1$:

$$I(b) = \ln b \to \infty \text{ as } b \to \infty$$

(2) If $\alpha \neq 1$:

$$I(b) = \frac{1}{1-\alpha}\left(b^{1-\alpha} - 1\right)$$

If $\alpha < 1$, then $b^{1-\alpha} \to \infty$ so $I(\infty)$ diverges. If $\alpha > 1$, then $b^{1-\alpha} \to 0$ and so:

$$I(\infty) \to \frac{1}{1-\alpha}$$

Hence, we can conclude that $I(\infty)$ converges for $\alpha > 1$ and diverges for $0 < \alpha \leqslant 1$.

We can extend this result and do some more exotic examples.

Example 8.5.5

Let $\alpha > 0$. Consider the improper integral:

$$\int_1^b \frac{\mathrm{d}x}{x(\ln x)^\alpha}$$

or even more exotic, consider this improper integral:

$$\int_1^b \frac{\mathrm{d}x}{x(\ln x)(\ln(\ln x))^\alpha}$$

For what α is this convergent and divergent. Turns out, the conclusion is the same as the previous example's.

8.6 Improper Integral II: Unbounded Functions

Definition 8.6.1 (More Improper Integrals)

1. Suppose $f \in R[a+\varepsilon, b]$ for all $\varepsilon > 0$ (but possibly, $f \notin R[a,b]$). Then, we define, if the limit exists:

$$\int_a^b f(x)\ \mathrm{d}x = \lim_{\varepsilon \to 0} \int_{a+\varepsilon}^b f(x)\ \mathrm{d}x \tag{8.12}$$

2. Similarly for $f \in R[a, b-\varepsilon]$ for all $\varepsilon > 0$. We define, if the limit exists:

$$\int_a^b f(x)\ \mathrm{d}x = \lim_{\varepsilon \to 0} \int_a^{b-\varepsilon} f(x)\ \mathrm{d}x \tag{8.13}$$

Example 8.6.2

Let $\alpha > 0$. Consider the almost-same integral as in (8.11). But now, we fix the upper bound, vary the lower bound and define it this way:

$$J(\varepsilon) = \int_\varepsilon^1 \frac{\mathrm{d}x}{x^\alpha}, \quad J(0) = \lim_{\varepsilon \to 0} J(\varepsilon)$$

For which α this is convergent and divergent? Turns out, it is a mirror-opposite to the case of unbounded intervals.

(1) If $\alpha \neq 1$:

$$J(\varepsilon) = \frac{1}{1-\alpha}(1 - \varepsilon^{1-\alpha})$$

So as $\varepsilon \to 0$, we have that $\varepsilon^{1-\alpha} \to 0$ if $\alpha < 1$ and $\varepsilon^{1-\alpha} \to \infty$ if $\alpha > 1$. Hence, for $\alpha < 1, J(0) = 1/(1-\alpha)$ and for $\alpha > 1$, $J(0)$ is not defined.

(2) If $\alpha = 1$:

$$J(\varepsilon) = -\ln \varepsilon \to \infty \text{ as } \varepsilon \to 0$$

so $J(0)$ diverges.

Hence, we conclude that $J(0)$ converges for $0 < \alpha < 1$ and diverges for $\alpha \geqslant 1$.

Then, we think about what happens if we combine the condition of unbounded intervals together with unbounded functions.

Example 8.6.3 (A Very Improper Integral)

We use our same notation as the previous example(s). Let $\alpha > 0$. Consider this integral:

$$\int_0^\infty \frac{dx}{x^\alpha}$$

We can break this into two integrals:

$$\int_0^\infty \frac{dx}{x^\alpha} = \int_0^1 \frac{dx}{x^\alpha} + \int_1^\infty \frac{dx}{x^\alpha} = J(0) + I(\infty)$$

where the last inequality is by definition of our choice of notation. We know that:

1. $J(0)$ converges for $0 < \alpha < 1$ and diverges for $\alpha \geqslant 1$
2. $I(\infty)$ converges for $\alpha > 1$ and diverges for $0 < \alpha \leqslant 1$.

So the integral $J(0) + I(\infty)$ never converges as we can't have any simultaneous convergent condition.

As before, we can proceed to tame more exotic examples.

Example 8.6.4

Let $\alpha > 0$. Consider the integral:

$$\int_0^\infty \frac{\mathrm{d}x}{x^\alpha(1+x)^\beta} = \int_0^1 \frac{\mathrm{d}x}{x^\alpha(1+x)^\beta} + \int_1^\infty \frac{\mathrm{d}x}{x^\alpha(1+x)^\beta}$$

It is an exercise to check that:

$$\int_0^1 \frac{\mathrm{d}x}{x^\alpha(1+x)^\beta}$$

converges when $\alpha < 1$ and that:

$$\int_1^\infty \frac{\mathrm{d}x}{x^\alpha(1+x)^\beta}$$

converges when $\alpha + \beta > 1$. A hint for this second condition is that when you evaluate the integral at very large values, the denominator seems to look just like $x^{\alpha+\beta}$. Then, applying what we have proved in the earlier example, this converges when the exponent is strictly greater than 1.