MATH2017 Real Analysis

Cheatsheet

2023/24

This document collects together the important definitions and results presented throughout the lecture notes. The numbering used throughout will be consistent with that in the lecture notes.

Contents

1 Limits and Continuity

1.1 The Limit of a Convergent Sequence

Reminder: A sequence of real numbers is a function $a : \mathbb{Z}^+ \to \mathbb{R}$, where we denote the output by a_n (instead of the usual notation $a(n)$ for functions). A term in the sequence is denoted by a_n, whereas the whole sequence is denoted by $(a_n)_{n\in\mathbb{Z}^+}$, or just (a_n) for short.

Definition 1.1 A real sequence (a_n) converges to a real number $L \in \mathbb{R}$ if, for each $\varepsilon > 0$, there exists $N \in \mathbb{Z}^+$ such that, for all $n \geq N$, we have $|a_n - L| < \varepsilon$. In this case, we call L the limit of (a_n) and we write either $a_n \to L$ or $\lim_{n\to\infty} a_n = L$. Here, we call (a_n) convergent.

Remark Given a sequence (a_n), we can show that it 'approaches' the number L as $n \in \mathbb{Z}^+$ gets large by showing that for **any** positive number ($\varepsilon > 0$), there exists a point in the sequence a_N (there exists $N \in \mathbb{Z}^+$) for which it and every subsequent term in the sequence (for all $n \geq N$) lies within distance that positive number of the number L ($|a_n - L| < \varepsilon$). Because this needs to work for **any** ε, the idea is that the distance can be as large or as small as you like and we should still be able to find $N \in \mathbb{Z}^+$ to make this work. Geometrically, if we plot n against a_n, every point for $n \geq N$ will live inside a rectangle with width 2ε centred on the line $a_n = L$.

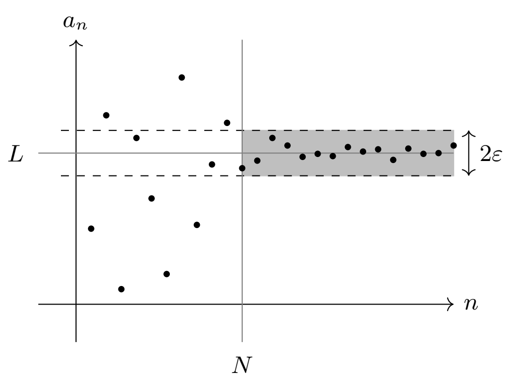


Figure 1: The geometric interpretation of the convergence of the sequence (a_n).

Lemma (Triangle Inequality) *For all $x, y \in \mathbb{R}$, we have $|x + y| \leq |x| + |y|$.*

Proof: By definition, it is clear that $x \leq |x|$ and $y \leq |y|$. Therefore, adding these inequalities tells us that $x+y \leq |x|+|y|$. Furthermore, we also see that $-x \leq |x|$ and $-y \leq |y|$. Adding these inequalities produces $-(x + y) \leq |x| + |y|$. Combining these statements gives what we want:

$$|x + y| \leq |x| + |y|. \qquad \square$$

Note: Another useful trick in analysis is to add zero to something in a non-trivial way.

Proposition 1.4 (Uniqueness of Limits) *The limit of a convergent sequence is unique.*

Proof: Let (a_n) be convergent and suppose that $a_n \to L$ and $a_n \to K$. It is our task to prove $L = K$. To this end, let $\varepsilon > 0$ be given. We can apply Definition 1.1 in the following situations:

- There exists $N_1 \in \mathbb{Z}^+$ such that, for all $n \geq N_1$, $|a_n - L| < \varepsilon/2$.
- There exists $N_2 \in \mathbb{Z}^+$ such that, for all $n \geq N_2$, $|a_n - K| < \varepsilon/2$.

Define $N := \max\{N_1, N_2\}$. Then, for all $n \geq N$, we see that

$$\begin{aligned}
|L - K| &= |L - a_n + a_n - K| \\
&\leq |L - a_n| + |a_n - K|, && \text{by the Triangle Inequality,} \\
&= |a_n - L| + |a_n - K|, && \text{by properties of the absolute value,} \\
&< \varepsilon/2 + \varepsilon/2, && \text{by the inequalities above,} \\
&= \varepsilon.
\end{aligned}$$

This shows that the 'distance' between the real numbers L and K is less than the positive number ε, but this works for **any** ε, so we must have $|L - K| = 0$. In other words, $L = K$ as required. □

Definition 1.5 Let (a_n) be a sequence.

(i) It is bounded above if there exists $M \in \mathbb{R}$ where $a_n \leq M$ for all $n \in \mathbb{Z}^+$. In this case, we call the number M an upper bound on the sequence.

(ii) It is bounded below if there exists $K \in \mathbb{R}$ where $a_n \geq K$ for all $n \in \mathbb{Z}^+$. In this case, we call the number K an lower bound on the sequence.

(iii) It is bounded if it is bounded above and below.

Proposition 1.6 *If a sequence (a_n) is convergent, then it is bounded.*

Proof: Let $a_n \to L$ for some $L \in \mathbb{R}$. Then, there exists $N \in \mathbb{Z}^+$ such that, for all $n \geq N$, we have $|a_n - L| < 1$ (remember this works for **all** $\varepsilon > 0$, so it works for $\varepsilon = 1$ in particular). By the Triangle Inequality, this means that $|a_n| \leq |L| + 1$ for all $n \geq \mathbb{Z}^+$. If we define the number

$$K := \max\{|a_1|, |a_2|, ..., |a_{N-1}|, |L| + 1\},$$

then we see that $|a_n| \leq K$ for **every** $n \in \mathbb{Z}^+$, as required. □

Note: The idea of the proof is this: even though $|a_n| \leq |L| + 1$ for all $n \geq N$, we aren't done yet. It might be that an earlier term of the sequence is larger than this number. Therefore, we must consider the sizes of these earlier terms, namely $|a_n|$ for $n < N$. Hence, K is the largest number amongst the earlier terms and the number $|L| + 1$.

Remark The converse of Proposition 1.6 is false: $a_n = (-1)^n$ is bounded but **not** convergent.

Proposition 1.7 *If $a_n \to L$ and $K \leq a_n \leq M$ for all $n \in \mathbb{Z}^+$, then $K \leq L \leq M$.*

Proof: Since $a_n \to L$, there exists $N \in \mathbb{Z}^+$ such that, for all $n \geq N$, we have $|a_n - L| < \varepsilon$. In particular, it is satisfied by the term a_N, meaning $|a_N - L| < \varepsilon$. We now show each inequality.

- Assume to the contrary that $L > M$. Then, we can choose $\varepsilon = L - M > 0$. But this means that $a_N > L - \varepsilon = M$, a contradiction to the fact that $a_n \leq M$ for **all** $n \in \mathbb{Z}^+$.
- Assume to the contrary that $L < K$. Then, we can choose $\varepsilon = K - L > 0$. But this means that $a_N < L + \varepsilon = K$, a contradiction to the fact that $a_n \geq K$ for **all** $n \in \mathbb{Z}^+$. □

Remark The bounds in Proposition 1.7 must be non-strict, that is if we impose $K < a_n < M$ for all $n \in \mathbb{Z}^+$, the statement is false: $a_n = 1/n$ satisfies $0 < a_n < 2$ for all n, but its limit $L = 0$ does **not** satisfy the same inequality $0 < L < 2$.

Proposition 1.8 (Squeeze Rule) *If $a_n \leq b_n \leq c_n$ with $a_n \to L$ and $c_n \to L$, then $b_n \to L$.*

Proof: For any $\varepsilon > 0$, there exists $N \in \mathbb{Z}^+$ such that, for all $n \geq N$, we have both $|a_n - L| < \varepsilon$ and $|c_n - L| < \varepsilon$. By definition of the absolute value, this means that $a_n > L - \varepsilon$ and $c_n < L + \varepsilon$. Hence, for all $n \geq N$, we have

$$b_n \geq a_n > L - \varepsilon \qquad \text{and} \qquad b_n \leq c_n < L + \varepsilon.$$

Combining these gives us $L - \varepsilon < b_n < L + \varepsilon$, which is equivalent to $-\varepsilon < b_n - L < \varepsilon$ by subtracting L. But this is precisely $|b_n - L| < \varepsilon$, and so $b_n \to L$ as required. □

Remark We can visualise Proposition 1.8 with a geometric interpretation given below in Figure 2. Notice that in the picture, the sequence (b_n) is confined to the rectangle of width 2ε much earlier than either of N_1 and N_2. But the point of Proposition 1.8 is that it gives us a cheap-and-easy way to show that (b_n) does indeed end up staying inside this rectangle.

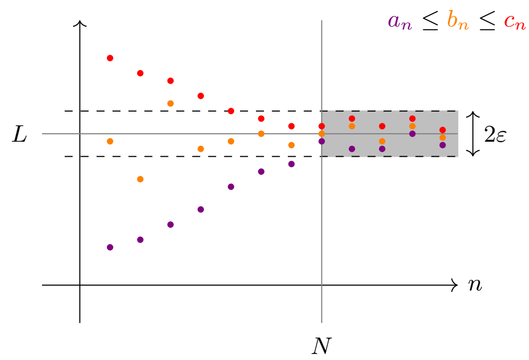


Figure 2: The geometric interpretation of the Squeeze Rule.

Proposition 1.9 (Algebra of Limits) *For $a_n \to A$ and $b_n \to B$, the following are true:*

(i) $a_n + b_n \to A + B$.

(ii) $a_n b_n \to AB$.

(iii) $a_n/b_n \to A/B$ **if** $b_n \neq 0$ *for all* $n \in \mathbb{Z}^+$ *and* $B \neq 0$.

Sketch of Proof: (i) Since (a_n) and (b_n) are convergent, we can assume $|a_n - A| < \varepsilon/2$ and $|b_n - B| < \varepsilon/2$ for large enough n. Then, we apply the Triangle Inequality to $\left|a_n + b_n - (A + B)\right|$.

(ii) Since (b_n) is convergent, it is bounded by Proposition 1.6, that is $|b_n| \leq K$ for some $K \in \mathbb{R}$. But since (a_n) is also convergent, we can assume $|a_n - A| < \varepsilon'$ and $|b_n - B| < \varepsilon'$ for large enough n, where for any $\varepsilon > 0$ we define

$$\varepsilon' := \frac{\varepsilon}{K + |A|}.$$

Then, we apply the Triangle Inequality to $|a_n b_n - AB| = |a_n b_n - Ab_n + Ab_n - AB|$.

(iii) We prove that $1/b_n \to 1/B$ and apply (ii). To this end, since $B \neq 0$, we know that $|B|/2 > 0$. Since (b_n) is convergent, we can assume $|b_n - B| < |B|/2$ for $n \geq N_1$ and $|b_n - B| < \varepsilon'$, where for any $\varepsilon > 0$ we define $\varepsilon' := \varepsilon|B|^2/2$. Then, we write $\left|1/b_n - 1/B\right|$ using a common denominator. $\square$

Definition 1.10 A subsequence of (a_n) is a sequence (b_k) such that there exists a strictly increasing sequence of positive integers (n_k) such that $b_k = a_{n_k}$.

In other words, the terms in (b_k) must occur in (a_n) in the same order. An alternate take is that we can obtain (b_k) from (a_n) by simply omitting the terms we don't want.

Proposition 1.11 *If $a_n \to L$ and (b_k) is a subsequence of (a_n), then $b_k \to L$.*

Proof: Let $\varepsilon > 0$ be given. Since $a_n \to L$, there exists $N \in \mathbb{Z}^+$ such that, for all $n \geq N$, we have $|a_n - L| < \varepsilon$. By Definition 1.10, we can write $b_k = a_{n_k}$ for a strictly increasing sequence (n_k) of positive integers. Note that $n_1 \geq 1$ and, if $n_k \geq k$, then $n_{k+1} \geq n_k + 1 \geq k + 1$. By induction, we conclude that $n_k \geq k$ for all $k \in \mathbb{Z}^+$. Therefore, for all $k \geq N$, we have $n_k \geq n_N \geq N$ which means $|b_k - L| = \left|a_{n_k} - L\right| < \varepsilon$. $\square$

Method – Showing Divergence of a Sequence: To show that a sequence (a_n) does **not** converge, one can show that it has a subsequence which diverges. Alternatively, it suffices to find subsequences that converge to different limits (combine Propositions 1.4 and 1.11).

Definition 1.12 Let (a_n) be a sequence.

(i) It is increasing if $a_{n+1} \geq a_n$ for all $n \in \mathbb{Z}^+$.
(ii) It is decreasing if $a_{n+1} \leq a_n$ for all $n \in \mathbb{Z}^+$.
(iii) It is strictly increasing if $a_{n+1} > a_n$ for all $n \in \mathbb{Z}^+$.
(iv) It is strictly decreasing if $a_{n+1} < a_n$ for all $n \in \mathbb{Z}^+$.
(v) It is monotonic if it is either increasing or decreasing.

We now state but do not prove the following important results (they were proved in MATH1026).

Theorem 1.13 (Monotone Convergence) *Every bounded monotonic sequence converges.*

Theorem 1.14 (Bolzano-Weierstrass) *Any bounded sequence has a convergent subsequence.*

1.2 Convergence of Sequences and the Cauchy Property

Definition 1.15 A real sequence (a_n) is Cauchy (or has the Cauchy property) if, for each $\varepsilon > 0$, there exists $N \in \mathbb{Z}^+$ such that, for all $n, m \geq N$, we have $|a_n - a_m| < \varepsilon$.

Remark The definition of Cauchy is very similar to that of convergent, with a key difference; no mention of a real number L. Instead, we look at the difference between two terms a_n and a_m. In words, where convergence is about having all terms after a certain point being within distance ε of the limit L, the Cauchy property is about having all terms after a certain point being within distance ε of **each other**.

Lemma 1.16 *If (a_n) converges, then it is Cauchy.*

Proof: Let $a_n \to L$, meaning for any $\varepsilon > 0$, there exists $N \in \mathbb{Z}^+$ such that, for all $n \geq N$, we have $|a_n - L| < \varepsilon/2$. But for all $n, m \geq N$, the Triangle Inequality can be applied to produce

$$|a_n - a_m| = |a_n - L + L - a_m| \leq |a_n - L| + |a_m - L| < \varepsilon. \qquad \square$$

Lemma 1.17 *If (a_n) is Cauchy, then it is bounded.*

Proof: By assumption, there exists $N \in \mathbb{Z}^+$ such that, for all $n, m \geq N$, we have $|a_n - a_m| < 1$. In particular, for all $n \geq N$, it is true that $|a_n - a_N| < 1$. This can be rearranged to say that $|a_n| < |a_N| + 1$. Proceeding similarly to the proof of Proposition 1.6, we define the number

$$K := \max\{|a_1|, |a_2|, ..., |a_{N-1}|, |a_N| + 1\}.$$

Then, we see that $|a_n| \leq K$ for **every** $n \in \mathbb{Z}^+$, as required. $\square$

Lemma 1.18 *If (a_n) is Cauchy and has a convergent subsequence, then it itself converges.*

Proof: Let $a_{n_k} \to L$ be the convergent subsequence and suppose $\varepsilon > 0$ is given. Then, there exists $N_1 \in \mathbb{Z}^+$ such that, for all $k \geq N_1$, ew have $\left|a_{n_k} - L\right| < \varepsilon/2$. But the Cauchy property means there exists $N_2 \in \mathbb{Z}^+$ such that, for all $n, m \geq N_2$, we have $|a_n - a_m| < \varepsilon/2$. Define $N = \max\{N_1, N_2\}$. Then, for all $n \geq N$, we see that

$$\begin{aligned} |a_n - L| &= \left|a_n - a_{n_N} + a_{n_N} - L\right| \\ &\leq \left|a_n - a_{n_N}\right| + \left|a_{n_N} - L\right|, && \text{by the Triangle Inequality,} \\ &< \varepsilon/2 + \varepsilon/2, && \text{by the inequalities above,} \\ &= \varepsilon. \end{aligned}$$

□

Theorem 1.19 *A sequence (a_n) converges if and only if it is Cauchy.*

Proof: The "only if" is Lemma 1.16. Conversely, if (a_n) is Cauchy, it is bounded (Lemma 1.17), so has a convergent subsequence (Bolzano-Weierstrass) and thus converges (Lemma 1.18). □

Note: We could define the Cauchy property for a sequence of elements in any so-called *metric space*. However, it is **not** true in general that convergence and the Cauchy property are equivalent. The metric spaces for which this **is** the case care called complete. Hence, Theorem 1.19 can be re-stated in the following succinct way: $\mathbb{R}$ is complete.

1.3 Limits of Functions

1.3.1 Limits at Infinity

Definition 1.22 Let $D \subseteq \mathbb{R}$ be unbounded above. We say $f : D \to \mathbb{R}$ has a limit $L \in \mathbb{R}$ at infinity if, for each $\varepsilon > 0$, there exists $K \in \mathbb{R}$ such that, for all $x \in D$ with $x > K$, we have $\left|f(x) - L\right| < \varepsilon$. In this case, we write $\lim\limits_{x \to \infty} f(x) = L$.

We can make a very similar definition this time for the limit of a function at minus infinity.

Definition Let $D \subseteq \mathbb{R}$ be unbounded below. We say $f : D \to \mathbb{R}$ has a limit $L \in \mathbb{R}$ at minus infinity if, for each $\varepsilon > 0$, there exists $K \in \mathbb{R}$ such that, for all $x \in D$ with $x < K$, we have $\left|f(x) - L\right| < \varepsilon$. In this case, we write $\lim\limits_{x \to -\infty} f(x) = L$.

1.3.2 Cluster Points

Definition 1.26 Let $D \subseteq \mathbb{R}$. We say $a \in \mathbb{R}$ is a cluster point of D if, for each $\varepsilon > 0$, there exists $x \in D$ with $0 < |x - a| < \varepsilon$. Equivalently, $(D \setminus \{a\}) \cap (a - \varepsilon, a + \varepsilon) \neq \emptyset$ for all $\varepsilon > 0$.

A characterisation of cluster points by sequences may help illuminate what they are.

Proposition 1.28 *Let $D \subseteq \mathbb{R}$. Then, $a \in \mathbb{R}$ is a cluster point of D if and only if there exists a sequence in $D \setminus \{a\}$ that converges to a.*

Proof: ($\Rightarrow$) Let $a \in \mathbb{R}$ be a cluster point. By definition, for each $\varepsilon > 0$, there exists $x \in D$ with $0 < |x - a| < \varepsilon$. Hence, $x \in D \setminus \{a\}$. But this works in particular for $\varepsilon = 1/n$ where $n \in \mathbb{Z}^+$, i.e. for each $n \in \mathbb{Z}^+$, there exists $x_n \in D \setminus \{a\}$ such that $0 < |x_n - a| < 1/n$. The sequence (x_n) lies in $D \setminus \{a\}$ and the Squeeze Rule implies that $x_n \to a$.

($\Leftarrow$) Let (x_n) be a sequence in $D \setminus \{a\}$ such that $x_n \to a$. Then, for any $\varepsilon > 0$, there exists $N \in \mathbb{Z}^+$ such that, for all $n \geq N$, we have $|x_n - a| < \varepsilon$. In particular, this is satisfied by the N^{th} term: $|x_N - a| < \varepsilon$. But since $x_N \in D$ and $x_N \neq a$, it is true that $0 < |x_N - a| < \varepsilon$. Hence, we have found $x = x_N$ as in Definition 1.26, that is a is a cluster point of D. $\square$

1.3.3 Limits of Functions: The Main Definition

Definition 1.29 Let $D \subseteq \mathbb{R}$ and $a \in \mathbb{R}$ be a cluster point of D. We say that $f : D \to \mathbb{R}$ has a limit $L \in \mathbb{R}$ at a if, for each $\varepsilon > 0$, there exists $\delta > 0$ such that, for all $x \in D$ with $0 < |x - a| < \delta$, we have $\big|f(x) - L\big| < \varepsilon$. In this case, we write $\lim\limits_{x \to a} f(x) = L$.

Remark In other words, given any positive number ($\varepsilon > 0$), we can find another positive number ($\delta > 0$) such that whenever x is within distance δ of a ($0 < |x - a| < \delta$), it follows that $f(x)$ is within distance ε of L ($\big|f(x) - L\big| < \varepsilon$). Geometrically, if we plot the graph of f and delete from it the point $(a, f(a))$, having a limit at $a \in \mathbb{R}$ tells us that the part of the graph between the vertical lines $x = a \pm \delta$ lies entirely between the horizontal lines $y = L \pm \varepsilon$.

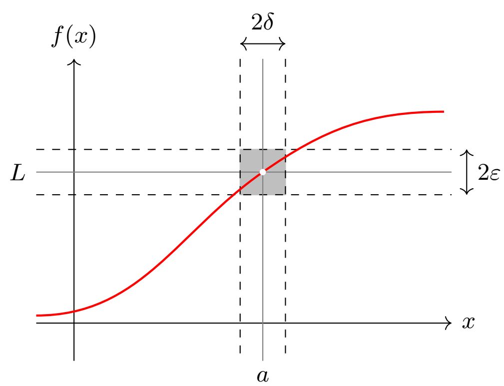


Figure 3: The geometric interpretation of the limit of $f : D \to \mathbb{R}$ at a.

Note: The cluster point inequality $0 < |x - a| < \delta$ is equivalent to $x \in (a - \delta, a + \delta) \setminus \{a\}$.

Method – Finding δ: Suppose we wish to do an ε-δ proof of a limit.

(i) Manipulate the absolute value $\left|f(x) - L\right| = g(x)|x-a|$, where g is some function.
(ii) Pick $n \in \mathbb{R}$ such that $0 < \delta \leq n$ keeps the interval $(a-\delta, a+\delta)$ strictly positive.
(iii) Use $x \leq a + \delta \leq a + n$ to determine an inequality that g must satisfy, i.e. $g(x) < m$.
(iv) Choose $\delta = \min\{n, \frac{\varepsilon}{m}\}$ to ensure that $g(x)|x-a| < \varepsilon$.

1.3.4 Some Basic Properties of Limits

Theorem 1.33 (Uniqueness of Limits) *If a function f has a limit at a, then it is unique.*

Proof: Let $f : D \to \mathbb{R}$ and $a \in \mathbb{R}$ be a cluster point of D. Assume to the contrary that **distinct** $L_1, L_2 \in \mathbb{R}$ satisfy Definition 1.29. For $\varepsilon = |L_2 - L_1|/2 > 0$, apply said definition in each case:

- There exists $\delta_1 > 0$ such that, for all $x \in D$ with $0 < |x - a| < \delta_1$, we have $\left|f(x) - L_1\right| < \varepsilon$.
- There exists $\delta_2 > 0$ such that, for all $x \in D$ with $0 < |x - a| < \delta_2$, we have $\left|f(x) - L_2\right| < \varepsilon$.

Define $\delta \coloneqq \min\{\delta_1, \delta_2\}$ and let $y \in D$ be such that $0 < |y - a| < \delta$; this exists since a is a cluster point of D. Then, for all $x \in D$ with $0 < |x - a| < \delta$, we obtain the following contradiction:

$$\begin{aligned}
2\varepsilon &= |L_2 - L_1| \\
&= \left|L_2 - f(y) + f(y) - L_1\right| \\
&\leq \left|f(y) - L_2\right| + \left|f(y) - L_1\right|, && \text{by the Triangle Inequality,} \\
&< \varepsilon + \varepsilon, && \text{by the inequalities above,} \\
&= 2\varepsilon.
\end{aligned}$$

□

Note: The proof makes essential use of the fact a is a cluster point of D. If we **omit** this assumption, we can't be sure that the point $y \in D \setminus \{a\}$ satisfying $0 < |y - a| < \delta$ exists!

Reminder: The negation of a statement is true if and only if the original statement is false. Under negation, the universal quantifier ($\forall$) and existential ($\exists$) quantifier swap roles.

Theorem 1.35 *Let $f : D \to \mathbb{R}$ and $a \in \mathbb{R}$ be a cluster point of D. Then, $\lim_{x\to a} f(x) = L$ if and only if for* **all** *sequences (x_n) in $D \setminus \{a\}$ such that $x_n \to a$, it follows that $f(x_n) \to L$.*

Proof: ($\Rightarrow$) Let (x_n) be a sequence in $D \setminus \{a\}$ such that $x_n \to a$, and suppose $\varepsilon > 0$ is given. By assumption of the existence of the limit of f at a, there exists $\delta > 0$ such that, for all $x \in D$ with $0 < |x - a| < \delta$, we have $\left|f(x) - L\right| < \varepsilon$. But since $x_n \to a$ and $\delta > 0$, there exists $N \in \mathbb{Z}^+$ such that, for all $n \geq N$, we have $|x_n - a| < \delta$ (notice this is just Definition 1.1 with δ playing the role usually held by ε). But since $x_n \neq a$, we know $|x_n - a| > 0$ for all n. Hence, for all $n \geq N$, we have $0 < |x_n - a| < \delta$. But by the definition of δ, this means $\left|f(x_n) - L\right| < \varepsilon$ for all $n \geq N$. In other words, this shows that $f(x_n) \to L$.

($\Leftarrow$) We will prove the contrapositive: if f does **not** have limit L at a, then there exists a sequence (x_n) in $D \setminus \{a\}$ such that $x_n \to a$ but $f(x_n) \not\to L$. To this end, suppose $\lim_{x \to a} f(x) = L$ is **false**: there *exists* $\varepsilon > 0$ such that, for *all* $\delta > 0$, there *exists* $x \in D$ with $0 < |x - a| < \delta$ such that $\left|f(x) - L\right| \geq \varepsilon$. Because this holds for all $\delta > 0$, it holds in particular for $\delta = 1/n$ where $n \in \mathbb{Z}^+$. In other words, for each n, there exists a point $x_n \in D$ with $0 < |x_n - a| < 1/n$ such that $\left|f(x_n) - L\right| \geq \varepsilon$. But by the Squeeze Rule, it is clear that $x_n \to a$. Because $0 < |x_n - a|$, we know that $x_n \neq a$ (i.e. $x_n \in D \setminus \{a\}$). The fact $\left|f(x_n) - L\right| \geq \varepsilon$ tells us that $f(x_n) \not\to L$. $\square$

Theorem 1.34 (Algebra of Limits) *For* $\lim_{x \to a} f(x) = L$ *and* $\lim_{x \to a} g(x) = K$*, these are true:*

(i) $\lim_{x \to a} \big(f(x) + g(x)\big) = L + K$.

(ii) $\lim_{x \to a} \big(f(x)g(x)\big) = LK$.

(iii) $\lim_{x \to a} \big(1/f(x)\big) = 1/L$ **if** $f(x) \neq 0$ *for all* $x \in D$ *and* $L \neq 0$.

Proof: We can do an ε-δ argument in the style of the proof of the Algebra of Limits for sequences (Proposition 1.9). However, a much more efficient argument follows from Theorem 1.35, alongside a direct application of Proposition 1.9. Indeed, let (x_n) be a sequence in $D \setminus \{a\}$ with $x_n \to a$. By assumption, Theorem 1.35 tells us $f(x_n) \to L$ and $g(x_n) \to K$. Let's now work case-by-case.

(i) By Proposition 1.9(i), we have $f(x_n) + g(x_n) \to L + K$.

(ii) By Proposition 1.9(ii), we have $f(x_n)g(x_n) \to LK$.

(iii) By Proposition 1.9(iii), we have $1/f(x_n) \to 1/L$.

As these hold for any such sequence (x_n), the statement follows directly from Theorem 1.35. $\square$

1.3.5 Continuity and Limits

Definition 1.37 Let $f : D \to \mathbb{R}$ and $a \in D$. We say that f is continuous at a if, for **all** sequences (x_n) in D such that $x_n \to a$, it follows that $f(x_n) \to f(a)$. If f is **not** continuous at a, it is discontinuous at a. Moreover, f is continuous if it is continuous at **every** $a \in D$.

Lemma *Every polynomial* $f : \mathbb{R} \to \mathbb{R}$ *given by* $f(x) = a_0 + a_1 x + \cdots + a_k x^k$ *is continuous.*

Proof: This is continuous by Proposition 1.9, and the easy fact that $\mathrm{id}(x) = x$ is continuous: if $x_n \to a$, then $\mathrm{id}(x_n) = x_n \to a = \mathrm{id}(a)$; this is exactly what Definition 1.37 asks us to show. $\square$

Note: A function which is continuous everywhere **except one** point is any *step function*. On the other hand, a function which is discontinuous everywhere **except one** point is

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

Lemma (Stars over Babylon) *The Stars over Babylon function $f : \mathbb{R} \to \mathbb{R}$ which is given by*

$$f(x) = \begin{cases} 0 & \text{if } x \notin \mathbb{Q} \\ \dfrac{1}{q} & \text{if } x = \dfrac{p}{q} \in \mathbb{Q} \text{ written in lowest terms with } q > 0 \end{cases}$$

is continuous at every irrational number, and discontinuous at every rational number.

Proof: (i) We will show that for all $a \in \mathbb{R} \setminus \mathbb{Q}$, f is continuous at a. Assume to the contrary that there is a sequence (x_n) such that $x_n \to a$ but $f(x_n) \not\to f(a)$. Note first that $a \notin \mathbb{Q}$, so we know $f(a) = 0$. Saying that $f(x_n)$ doesn't converge to zero means there *exists* $\varepsilon > 0$ such that, for *all* $k \in \mathbb{Z}^+$, there *exists* a term $n_k \in \mathbb{Z}^+$ such that $n_k \geq k$ and $\left|f(x_{n_k}) - 0\right| \geq \varepsilon$ (or, in other words, $f(x_{n_k}) \geq \varepsilon$ since f is non-negative). Hence, there is a subsequence $y_k = x_{n_k}$ such that $f(y_k) \geq \varepsilon$. By Proposition 1.11, we know that $y_k \to a$. Moreover, since $f(y_k) \geq \varepsilon > 0$, we know that each $y_k \in \mathbb{Q}$ (since the only non-zero outputs of this function come from rational inputs). This means we can write $y_k = p_k/q_k$ for $p_k, q_k \in \mathbb{Z}$, $q_k > 0$ and $\gcd(p_k, q_k) = 1$. By definition, we know

$$f(y_k) = \frac{1}{q_k} \geq \varepsilon.$$

Hence, this tells us that $0 < q_k \leq 1/\varepsilon$. In other words, (q_k) is a bounded sequence in $\mathbb{Z}^+$. By the Bolzano-Weierstrass Theorem, there exists a convergent subsequence (q_{k_m}), that is $q_{k_m} \to q$ for some $q \in \mathbb{Z}^+$. By Theorem 1.19, we know that (q_{k_m}) is Cauchy. Consequently, there is a point in the sequence after which every positive integer lies within distance $1/2$ of each other (taking $\varepsilon = 1/2$ in Definition 1.15). This is a sequence of integers, so it means all terms after this point are constant. Explicitly, there exists $K \in \mathbb{Z}^+$ such that for all $m \geq K$, $q_{k_m} = q$. For $m \geq K$,

$$y_{k_m} = \frac{p_{k_m}}{q_{k_m}} = \frac{p_{k_m}}{q}.$$

But $p_{k_m}/q \to a$ again by Proposition 1.11. Thus, (p_{k_m}) is bounded and has a convergent subsequence $(p_{k_{m_\ell}})$. by the Bolzano-Weierstrass Theorem. This subsequence is therefore also Cauchy, so every term lies within distance $1/2$ of each other beyond a certain term. But they are all integers, so there exists $L \in \mathbb{Z}^+$ such that, for all $\ell \geq L$, $p_{k_{m_\ell}} = p \in \mathbb{Z}^+$. For $\ell \geq L$,

$$y_{k_{m_\ell}} = \frac{p_{k_{m_\ell}}}{q_{k_{m_\ell}}} = \frac{p}{q}.$$

Hence, $y_{k_{m_\ell}} \to p/q$. But applying Proposition 1.11 (a lot), it follows that $y_{k_{m_\ell}} \to a$. By the Uniqueness of Limits for sequences, we know $a = p/q$, but $a \notin \mathbb{Q}$ and we have a contradiction.

(ii) We will show that for all $a \in \mathbb{Q}$, f is discontinuous at a. Assume to the contrary that $a \in \mathbb{Q}$ and f is continuous at a. For each $n \in \mathbb{Z}^+$, there exists an irrational $i_n \in \mathbb{R} \setminus \mathbb{Q}$ such that $a < i_n < a + 1/n$ (by the density of the irrationals in the rationals; see MATH1025). By the Squeeze Rule, $i_n \to a$. Because we assume f is continuous at a, we know that $f(i_n) \to f(a)$. By the assumption that $q > 0$, we know that $f(a) > 0$ for any rational $a = p/q$. But $f(i_n) = 0 \to 0$ since each i_n is irrational, a contradiction to the Uniqueness of Limits for sequences. $\square$

Note: Strangely, there is **no** function that is continuous on $\mathbb{Q}$ and discontinuous on $\mathbb{R} \setminus \mathbb{Q}$!

Proposition 1.41 *Let $f : D \to \mathbb{R}$ and $a \in D$ such that a* **not** *a cluster point of D. Then, f is continuous at a.*

Proof: Let (x_n) be any sequence in D such that $x_n \to a$. Since a is **not** a cluster point of D, there exists $\varepsilon > 0$ such that there is no element $x \in D$ with $0 < |x - a| < \varepsilon$. Consequently, if $x \in D$ is such that $|x - a| < \varepsilon$, to ensure the previous inequality fails, it must be that $|x - a| = 0$, i.e. $x = a$. Because $x_n \to a$, there exists $N \in \mathbb{Z}^+$ such that, for all $n \geq N$, we have $|x_n - a| < \varepsilon$. Combining this with what we just explained tells us that $x_n = a$. Hence, for all $n \geq N$, the sequence is constant and $f(x_n) = f(a)$. Therefore, $f(x_n) \to f(a)$ as required. $\square$

Theorem 1.42 *Let $f : D \to \mathbb{R}$ and $a \in D$ be a cluster point of D. These are equivalent:*

(i) *f is continuous at a.*

(ii) $\lim\limits_{x \to a} f(x) = f(a)$.

(iii) *For each $\varepsilon > 0$, there exists $\delta > 0$ such that, for all $x \in D$ with $|x - a| < \delta$, we have $\big|f(x) - f(a)\big| < \varepsilon$.*

Proof: $\big((\text{i}) \Rightarrow (\text{ii})\big)$ This is immediate from Theorem 1.35 with $L = f(a)$.

$\big((\text{ii}) \Rightarrow (\text{iii})\big)$ For any $\varepsilon > 0$, there exists $\delta > 0$ such that, for all $x \in D$ with $0 < |x - a| < \delta$, we have $\big|f(x) - f(a)\big| < \varepsilon$ by Definition 1.29. But if $x = a$, we see that $\big|f(x) - f(a)\big| = 0 < \varepsilon$. Hence, we can restrict merely to $x \in D$ with $|x - a| < \delta$ and we are done.

$\big((\text{iii}) \Rightarrow (\text{i})\big)$ Let (x_n) be any sequence in D such that $x_n \to a$ and $\varepsilon > 0$ is given. By assumption, there exists $\delta > 0$ such that, for all $x \in D$ with $|x - a| < \delta$, we have $\big|f(x) - f(a)\big| < \varepsilon$. But since $x_n \to a$, there exists $N \in \mathbb{Z}^+$ such that, for all $n \geq N$, we have $|x_n - a| < \delta$. By the definition of δ, it follows that $\big|f(x_n) - f(a)\big| < \varepsilon$ for all $n \geq N$, which is to say that $f(x_n) \to f(a)$. $\square$

Remark In other words, a function $f : D \to \mathbb{R}$ is continuous at $a \in D$ if a small change of the input (for all $x \in D$ with $|x - a| < \delta$) results in a small change of the output ($\big|f(x) - f(a)\big| < \varepsilon$). Geometrically, the graph of $f(x)$ should stay inside the rectangle as in Figure 4.

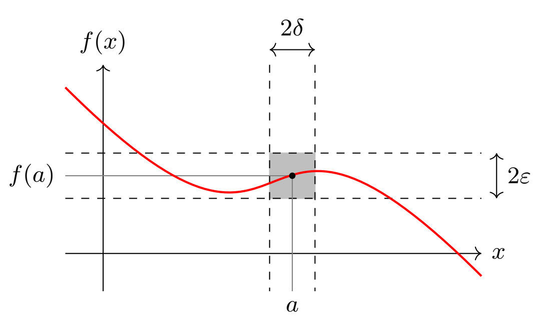


Figure 4: The geometric interpretation of continuity of $f : D \to \mathbb{R}$.

Theorem 1.45 *Let $D, E \subseteq \mathbb{R}$ with $f : D \to E$ continuous at $a \in D$ and $g : E \to \mathbb{R}$ continuous at $f(a) \in E$. Then, the composition $g \circ f : D \to \mathbb{R}$ is continuous at a.*

Proof: Let (x_n) be any sequence in D such that $x_n \to a$. Because f is continuous at a, we know that the sequence $y_n := f(x_n) \to f(a)$. But since g is continuous at $f(a)$, we know that $g(y_n) \to g(f(a))$. In other words, this shows $(g \circ f)(x_n) \to (g \circ f)(a)$ as required. □

Reminder: A closed bounded interval is $[a, b] := \{x \in \mathbb{R} : a \leq x \leq b\}$ for any $a, b \in \mathbb{R}$. It is non-empty if and only if we have $a \leq b$. Note that if $a = b$, then the interval is $[a, a] = \{a\}$.

Theorem 1.47 (Intermediate Value Theorem) *Let $f : [a, b] \to \mathbb{R}$ be continuous and y be a number between $f(a)$ and $f(b)$. Then, there exists $c \in [a, b]$ such that $f(c) = y$.*

Proof: Omitted, but proved in MATH1026 (we prove it differently later). □

Remark The classic (but highly non-rigorous) explanation of a continuous function is this: "f is continuous if it's graph can be drawn without taking your pen off the paper". This isn't accurate (since $f(x) = 1/x$ is continuous on $\mathbb{R} \setminus \{0\}$), but the Intermediate Value Theorem asserts that this slogan **is** true for continuous functions on an interval. Geometrically, any horizontal line between $f(a)$ and $f(b)$ must intersect the graph of f at least once.

Theorem 1.48 (Extreme Value Theorem) *Let $f : [a, b] \to \mathbb{R}$ be continuous. Then, f is bounded and attains both a minimum and maximum value.*

Proof: Omitted, but also proved in MATH1026. □

Remark Geometrically, the Extreme Value Theorem asserts that the graph of f as we vary its input from $x = a$ to $x = b$ can be covered by a rectangle, that is there exist two horizontal lines $y = m$ and $y = M$ which define the lower and upper edges, respectively, of said rectangle. These are the minimum and maximum values of the function.

2 Differentiable Functions

2.1 The Main Definition

Definition 2.1 Let $D \subseteq \mathbb{R}$, $f : D \to \mathbb{R}$ and $a \in D$ a cluster point of D. We say that f is differentiable at a if the following limit, called called the derivative of f at a, exists:

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a}.$$

We denote this by $f'(a)$. Moreover, f is differentiable if it is differentiable at **every** $a \in D$.

Remark Look back at Definition 1.29 and notice that to define a limit, we need only that a is a cluster point of the domain, **not** that it is also an element of the domain. However, we do impose this extra condition here. Therefore, we can very easily argue the following by definition:

- $f : D \to \mathbb{R}$ is **not** differentiable at any $a \notin D$.
- $f : \mathbb{Z} \to \mathbb{R}$ is **not** differentiable anywhere, since $\mathbb{Z}$ has no cluster points.

Note: We give an explicit ε-δ criteria of differentiability by referring to Definition 1.29. Namely, $f : D \to \mathbb{R}$ is differentiable at a if there exists a real number $f'(a) \in \mathbb{R}$ such that, for each $\varepsilon > 0$, there exists $\delta > 0$ where, for all $x \in D$ with $0 < |x - a| < \delta$, we have

$$\left| \frac{f(x) - f(a)}{x - a} - f'(a) \right| < \varepsilon.$$

Remark The slope of the straight line through the points $(a, f(a))$ and $(x, f(x))$ is given by

$$\frac{f(x) - f(a)}{x - a}.$$

If we take the limit as $x \to a$, these points approach one another, and the line segment between them (called a *chord*) approaches a straight line through $(a, f(a))$ with gradient $f'(a)$. Hence, we interpret the derivative as being the slope of the tangent line to the graph of f at a.

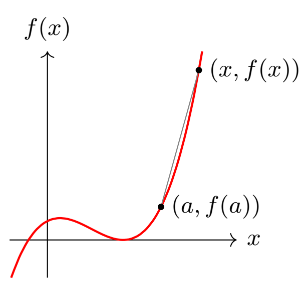


Figure 5: The geometric interpretation of differentiability of $f : D \to \mathbb{R}$.

Lemma *The following (constant) function $f : \mathbb{R} \to \mathbb{R}$ is differentiable at every $a \in \mathbb{R}$:*

$$f(x) = c, \qquad \text{with } f'(a) = 0.$$

Proof: Let $a \in \mathbb{R}$ be arbitrary, $\varepsilon > 0$ be given and take $\delta = 1$. For all $x \in \mathbb{R}$ with $0 < |x - a| < 1$,

$$\left| \frac{f(x) - f(a)}{x - a} - 0 \right| = \left| \frac{c - c}{x - a} \right| = 0 < \varepsilon. \qquad \square$$

Lemma *The following function $f : \mathbb{R} \to \mathbb{R}$ is differentiable at every $a \in \mathbb{R}$:*

$$f(x) = x, \qquad \text{with } f'(a) = 1.$$

Proof: Let $a \in \mathbb{R}$ be arbitrary, $\varepsilon > 0$ be given and take $\delta = 1$. For all $x \in \mathbb{R}$ with $0 < |x - a| < 1$,

$$\left| \frac{f(x) - f(a)}{x - a} - 1 \right| = \left| \frac{x - a}{x - a} - 1 \right| = 0 < \varepsilon. \qquad \square$$

Note: In the above proofs, **any** $\delta > 0$ work. We just decided on $\delta = 1$ because "why not?".

Lemma *The following function $f : \mathbb{R} \to \mathbb{R}$ is differentiable at every $a \in \mathbb{R}$:*

$$f(x) = x^2, \qquad \text{with } f'(a) = 2a.$$

Proof: Let $a \in \mathbb{R}$ be arbitrary, $\varepsilon > 0$ be given and take $\delta = \varepsilon$. For all $x \in \mathbb{R}$ with $0 < |x - a| < \delta$,

$$\left| \frac{f(x) - f(a)}{x - a} - 2a \right| = \left| \frac{x^2 - a^2}{x - a} - 2a \right| = |x + a - 2a| = |x - a| < \delta = \varepsilon. \qquad \square$$

Lemma *The following function $f : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ is differentiable at every $a \in \mathbb{R} \setminus \{0\}$:*

$$f(x) = \frac{1}{x}, \qquad \text{with } f'(a) = -\frac{1}{a^2}.$$

Proof: Let $a \in \mathbb{R} \setminus \{0\}$ be arbitrary, $\varepsilon > 0$ be given and take $\delta = \min\{|a|/2, |a|^3 \varepsilon/2\}$. For all $x \in \mathbb{R} \setminus \{0\}$ with $0 < |x - a| < \delta$,

$$\begin{aligned} \left| \frac{f(x) - f(a)}{x - a} + \frac{1}{a^2} \right| &= \left| \frac{(1/x) - (1/a)}{x - a} + \frac{1}{a^2} \right| \\ &= \left| \frac{a - x}{ax(x - a)} + \frac{1}{a^2} \right| \end{aligned}$$

$$
\begin{aligned}
&= \left| -\frac{1}{ax} + \frac{1}{a^2} \right| \\
&= \frac{1}{|a|} \left| \frac{1}{a} - \frac{1}{x} \right| \\
&= \frac{1}{|a|^2 |x|} |x - a| \\
&\leq \frac{2}{|a|^3} |x - a|, && \text{as } |x-a| < |a|/2 \text{ and so } |x| > |a|/2, \\
&< \varepsilon, && \text{as } |x-a| < |a|^3 \varepsilon/2. \qquad \square
\end{aligned}
$$

Remark You may think "where on Earth did you decide on that value of δ?", but secretly we did the absolute value calculation **first**. Indeed, we get to the fifth line (the one with $|x-a|/|a|^2|x|$) and, from there, we consider the following: if $|x - a| < \delta$, how do I make this whole expression less than ε? Well, we choose δ so that $1/|x|$ doesn't become unbounded (i.e. is away from zero, so half the distance from a will do), and then we introduce the second inequality to make the result less than ε. Hence, δ should be the smaller of these numbers to ensure **both** cases hold.

Lemma *The following function $f : [0, \infty) \to \mathbb{R}$ is differentiable at every $a \in (0, \infty)$:*

$$f(x) = \sqrt{x}, \qquad \text{with } f'(a) = \frac{1}{2\sqrt{a}}.$$

Proof: Let $a \in (0, \infty)$, $\varepsilon > 0$ be given and $\delta = 2a^{3/2}\varepsilon$. For all $x \in [0, \infty)$ with $0 < |x - a| < \delta$,

$$
\begin{aligned}
\left| \frac{f(x) - f(a)}{x - a} - \frac{1}{2\sqrt{a}} \right| &= \left| \frac{\sqrt{x} - \sqrt{a}}{x - a} - \frac{1}{2\sqrt{a}} \right| \\
&= \left| \frac{\sqrt{x} - \sqrt{a}}{x - a} \cdot \frac{\sqrt{x} + \sqrt{a}}{\sqrt{x} + \sqrt{a}} - \frac{1}{2\sqrt{a}} \right| \\
&= \left| \frac{x - a}{(x - a)(\sqrt{x} + \sqrt{a})} - \frac{1}{2\sqrt{a}} \right| \\
&= \left| \frac{1}{\sqrt{x} + \sqrt{a}} - \frac{1}{2\sqrt{a}} \right| \\
&= \left| \frac{\sqrt{x} - \sqrt{a}}{2\sqrt{a}(\sqrt{x} + \sqrt{a})} \right| \\
&= \left| \frac{\sqrt{x} - \sqrt{a}}{2\sqrt{a}(\sqrt{x} + \sqrt{a})} \cdot \frac{\sqrt{x} + \sqrt{a}}{\sqrt{x} + \sqrt{a}} \right| \\
&= \frac{|x - a|}{2\sqrt{a}(\sqrt{x} + \sqrt{a})^2} \\
&\leq \frac{|x - a|}{2a^{3/2}}, && \text{as } \sqrt{x} \geq 0, \\
&< \varepsilon. \qquad \square
\end{aligned}
$$

Note: This proof uses the standard trick, when working with square roots, of multiplying by one in a slightly obscure way. This ensures that we can get $|x-a|$ somewhere, which then allows us to do an estimate involving δ and ε.

To show that a function is **not** differentiable at a point, one option is to use the negation of Definition 2.1. However, it is often more convenient to use Theorem 1.35: it suffices to find **one** sequence (x_n) in $D \setminus \{a\}$ such that $x_n \to a$ but where this sequence doesn't converge:

$$\left(\frac{f(x_n)-f(a)}{x_n-a}\right).$$

Lemma *The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = |x|$ is* **not** *differentiable at* 0.

Proof: Assume to the contrary that f is differentiable at zero, with derivative $f'(0)$. Then,

$$\lim_{x \to 0} \frac{f(x)-f(0)}{x-0} = f'(0).$$

Consider the sequence $x_n = (-1)^n/n \to 0$ which lies in $\mathbb{R} \setminus \{0\}$. But Theorem 1.35 ensures that

$$\frac{f(x_n)-f(0)}{x_n-0} \to f'(0).$$

However, the left-hand side is precisely

$$\frac{f(x_n)-f(0)}{x_n-0} = \frac{\left|(-1)^n/n\right|}{(-1)^n/n} = \frac{1}{(-1)^n} = (-1)^n$$

which does not converge, a contradiction. □

Lemma *The function $f : [0,\infty) \to \mathbb{R}$ given by $f(x) = \sqrt{x}$ is* **not** *differentiable at* 0.

Proof: Assume to the contrary that f is differentiable at zero, with derivative $f'(0)$. Then,

$$\lim_{x \to 0} \frac{f(x)-f(0)}{x-0} = f'(0).$$

Consider the sequence $x_n = 1/n \to 0$ which lies in $(0,\infty)$. But Theorem 1.35 ensures that

$$\frac{f(x_n)-f(0)}{x_n-0} \to f'(0).$$

However, the left-hand side is precisely

$$\frac{f(x_n)-f(0)}{x_n-0} = \frac{\sqrt{1/n}}{1/n} = \frac{n}{\sqrt{n}} = \sqrt{n}$$

which does not converge, a contradiction. □

Proposition 2.8 *Let $f : D \to \mathbb{R}$ be differentiable at $a \in D$. Then, it is continuous at a.*

Proof: By Theorem 1.42, it suffices to show that $\lim_{x\to a} f(x) = f(a)$. Define $s : D \setminus \{a\} \to \mathbb{R}$ by

$$s(x) = \frac{f(x) - f(a)}{x - a}.$$

By assumption, we know that s as a limit at a, namely $f'(a)$. But for all $x \in D \setminus \{a\}$,

$$f(x) = f(a) + (x - a)s(x).$$

Appealing to the Algebra of Limits, it is clear that

$$\lim_{x\to a} f(x) = f(a) + 0f'(a) = f(a).$$ □

Note: The converse is **false**, that is continuity does not imply differentiability. For example, the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = |x|$ is continuous, but not differentiable, at zero.

Lemma *The following function $f : \mathbb{R} \to \mathbb{R}$ is differentiable at 0 but* **nowhere** *else:*

$$f(x) = \begin{cases} x & \text{if } x \in \mathbb{Q} \\ x - x^2 & \text{if } x \notin \mathbb{Q} \end{cases}, \qquad \text{with } f'(0) = 1.$$

Proof: Let $a \in \mathbb{R}$ and assume that f is differentiable at a. We aim to prove that $a = 0$; this does **not** prove that f is differentiable at zero (because we are assuming differentiability). The logic here is that we are proving the following by contrapositive: f is **not** differentiable at $a \neq 0$. But f is continuous by Proposition 2.8. For each $n \in \mathbb{Z}^+$, there exist $r_n \in \mathbb{Q}$ and $i_n \in \mathbb{R} \setminus \mathbb{Q}$ with

$$a < r_n < a + \frac{1}{n} \qquad \text{and} \quad a < i_n < a + \frac{1}{n}.$$

By the Squeeze Rule, we see that $r_n \to a$ and $i_n \to a$. By continuity, it follows that $f(r_n) \to f(a)$ and $f(i_n) \to f(a)$. But because r_n is rational and i_n is irrational, the definition of f tells us

$$f(r_n) = r_n \to a \qquad \text{and} \qquad f(i_n) = i_n - i_n^2 \to a - a^2.$$

But both limits are equal to $f(a)$, which tells us that $a = a - a^2$. The only solution is $a = 0$. □

Corollary *The function f from above is* **not** *increasing on any neighbourhood of zero.*

Proof: We must show that there does **not** exist $\varepsilon > 0$ such that $f : (-\varepsilon, \varepsilon) \to \mathbb{R}$ is increasing. Indeed, for any $\varepsilon > 0$ given, we can find an irrational number $x \in \mathbb{R} \setminus \mathbb{Q}$ such that $0 < x < \varepsilon$ (since the irrationals are dense in the reals). By the definition of the function f, it is clear that

$$f(x) = x - x^2 < x.$$

On the other hand, we can find a rational number $y \in \mathbb{Q}$ such that $x - x^2 < y < x$. Again by definition, $f(y) = y > x - x^2 = f(x)$, so $y < x$ does **not** imply that $f(y) < f(x)$ as required. □

2.2 The Rules of Differentiation

Proposition 2.10 (Linearity) *Let $f : D \to \mathbb{R}$ and $g : D \to \mathbb{R}$ be differentiable at $a \in D$ with derivatives $f'(a)$ and $g'(a)$ respectively, and $c \in \mathbb{R}$ some constant. Then, these are true:*

(i) *The function $cf : D \to \mathbb{R}$ is differentiable at a with derivative $cf'(a)$.*

(ii) *The function $f + g : D \to \mathbb{R}$ is differentiable at a with derivative $f'(a) + g'(a)$.*

Proof: This is immediate from the Algebra of Limits (Theorem 1.34), since we have

$$\lim_{x\to a} \frac{cf(x) - cf(a)}{x - a} = \lim_{x\to a} c\left(\frac{f(x) - f(a)}{x - a}\right) = cf'(a)$$

and

$$\lim_{x\to a} \frac{f(x) + g(x) - \big(f(a) + g(a)\big)}{x - a} = \lim_{x\to a} \left(\frac{f(x) - f(a)}{x - a} + \frac{g(x) - g(a)}{x - a}\right) = f'(a) + g'(a). \qquad \square$$

Note: The above proof uses $\lim_{x\to a} c = c$, which is easy to prove directly (with an ε-δ proof).

Reminder: Let $f : A \to B$ and $g : B \to C$. Their composition $g \circ f : A \to C$ is given by

$$(g \circ f)(x) = g(f(x)).$$

Of course, we can only compose functions if the range of f is a subset of the domain of g. Recall we know that the composition of continuous functions is itself continuous (Theorem 1.45). We can now look at the differentiability on the composition of two differentiable functions.

Proposition 2.12 (Carathéodory's Criterion) *Let $f : D \to \mathbb{R}$ and $a \in D$ be a cluster point of D. Then, f is differentiable at a if and only if there exists a function $\phi : D \to \mathbb{R}$ which is continuous at a and satisfies the following equation for all $x \in D$:*

$$f(x) - f(a) = \phi(x)(x - a). \qquad (*)$$

In this case, the derivative $f'(a) = \phi(a)$.

Proof: ($\Rightarrow$) Let f be differentiable at a with derivative $f'(a)$. We define $\phi : D \to \mathbb{R}$ by

$$\phi(x) = \begin{cases} \dfrac{f(x) - f(a)}{x - a} & \text{if } x \neq a \\ f'(a) & \text{if } x = a \end{cases}.$$

Clearly, ϕ satisfies $(*)$ for all $x \in D \setminus \{a\}$. But when $x = a$, both sides of $(*)$ are zero and thus it holds automatically. Moreover, we see that ϕ is continuous at a by appealing to Theorem 1.42:

$$\lim_{x\to a} \phi(x) = \lim_{x\to a} \frac{f(x) - f(a)}{x - a} = f'(a) = \phi(a).$$

($\Leftarrow$) Let such a function ϕ exist. Then, dividing $(*)$ by $(x-a)$ and taking a limit produces

$$\lim_{x\to a}\frac{f(x)-f(a)}{x-a}=\lim_{x\to a}\phi(x)=\phi(a),$$

by continuity. But this says precisely that f is differentiable at a with derivative $f'(a)=\phi(a)$. $\square$

Remark What is the motivation behind Carathéodory's Criterion? If a function $g:\mathbb{R}\setminus\{a\}\to\mathbb{R}$ has a limit L at a, we can extend the function to the domain $\mathbb{R}$ by **defining** $g(a):=L$. The resulting extended function is continuous by Theorem 1.42. On the other hand, if such a function $g:\mathbb{R}\setminus\{a\}\to\mathbb{R}$ admits a continuous extension to all of $\mathbb{R}$, then its limit at a is precisely $g(a)$. We have applied this idea to the difference quotient $s(x)$ defined in the proof of Proposition 2.8.

Theorem 2.11 (Chain Rule) *Let $f:D\to E$ be differentiable at $a\in D$ and $g:E\to\mathbb{R}$ be differentiable at $f(a)\in E$. Then, $g\circ f:D\to\mathbb{R}$ is differentiable at a, with derivative*

$$(g\circ f)'(a)=g'(f(a))f'(a).$$

Proof: By Carathéodory's Criterion, there exist functions $\phi:D\to\mathbb{R}$ and $\psi:E\to\mathbb{R}$ such that ϕ is continuous at a, ψ is continuous at $f(a)$ and, for all $x\in D$ and $y\in E$, we have the following:

$$f(x)-f(a)=\phi(x)(x-a)\qquad\text{and}\qquad g(y)-g(f(a))=\psi(y)(y-f(a)).$$

We define $\Phi:D\to\mathbb{R}$ by $\Phi(x)=\psi(f(x))\phi(x)$. By the Algebra of Limits and Theorem 1.45, it is clear that Φ is continuous at a. Moreover, for all $x\in D$, we conclude that

$$\begin{aligned}\Phi(x)(x-a)&=\psi(f(x))\phi(x)(x-a)\\&=\psi(f(x))\left(f(x)-f(a)\right), &&\text{by the first equality above,}\\&=g(f(x))-g(f(a)), &&\text{by the second equality above with } y=f(x).\end{aligned}$$

Hence, Carathéodory's Criterion applies again: $g\circ f$ is differentiable at a with derivative

$$\Phi(a)=\psi(f(a))\phi(a)=g'(f(a))f'(a).$$

$\square$

Proposition 2.15 (Product Rule) *Let $f:D\to\mathbb{R}$ and $g:D\to\mathbb{R}$ be differentiable at $a\in D$ with derivatives $f'(a)$ and $g'(a)$, respectively. Then, $fg:D\to\mathbb{R}$ is differentiable at a with derivative $f'(a)g(a)+f(a)g'(a)$.*

Proof: Let $\sigma:\mathbb{R}\to\mathbb{R}$ be the function $\sigma(x)=x^2$; recall this is differentiable (proven in an earlier lemma). Hence, the Chain Rule and linearity imply that the following are differentiable at a:

- $\sigma\circ f$, with derivative $(\sigma\circ f)'(a)=2f(a)f'(a)$.
- $\sigma\circ g$, with derivative $(\sigma\circ g)'(a)=2g(a)g'(a)$.
- $\sigma\circ(f+g)$, with derivative $\left(\sigma\circ(f+g)\right)'(a)=2\left(f(a)+g(a)\right)\left(f'(a)+g'(a)\right)$.

The trick here is to write the product fg that we are working with as

$$\begin{aligned} fg &= \frac{1}{2}\left((f+g)^2 - f^2 - g^2\right) \\ &= \frac{1}{2}\left((\sigma \circ (f+g)) - (\sigma \circ f) - (\sigma \circ g)\right). \end{aligned}$$

Applying the Chain Rule and simplifying, we see that fg is differentiable at a with derivative

$$\begin{aligned} (fg)'(a) &= \frac{1}{2}\Big(2\left(f(a)+g(a)\right)\left(f'(a)+g'(a)\right) - 2f(a)f'(a) - 2g(a)g'(a)\Big) \\ &= f'(a)g(a) + f(a)g'(a). \end{aligned}$$

□

Proposition 2.17 *Every polynomial $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = a_0 + a_1 x + \cdots a_k x^k$ is differentiable, and its derivative is another polynomial $f' : \mathbb{R} \to \mathbb{R}$ which is given by*

$$f'(x) = a_1 + 2a_2 x + \cdots + k a_k x^{k-1}.$$

Sketch of Proof: Use linearity and proceed by induction on the degree k of the polynomial. □

Corollary *Let $m \in \mathbb{Z}^+$ and $f : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ be the function defined as $f(x) = 1/x^m$. Then, f is differentiable with derivative given by $f'(x) = -m/x^{m+1}$.*

Proof: We can view $f = h \circ g$, where $g : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ is defined by $g(x) = 1/x$ and $h : \mathbb{R} \to \mathbb{R}$ is defined by $h(x) = x^m$. We know that g is differentiable (a previous lemma) and h is differentiable (Proposition 2.17). Thus, we can apply the Chain Rule to conclude that

$$\begin{aligned} f'(x) &= h'(g(x))g'(x) \\ &= mg(x)^{m-1}\left(-\frac{1}{x^2}\right) \\ &= -m\left(\frac{1}{x}\right)^{m-1}\left(\frac{1}{x^2}\right) \\ &= -\frac{m}{x^{m+1}}. \end{aligned}$$

□

Proposition 2.19 (Quotient Rule) *Let $f : D \to \mathbb{R}$ and $g : D \to \mathbb{R} \setminus \{0\}$ be differentiable at $a \in D$. Then, f/g is differentiable at a with derivative*

$$(f/g)'(a) = \frac{f'(a)g(a) - f(a)g'(a)}{g(a)^2}.$$

Sketch of Proof: Use the same trick as in the above corollary: view $f/g = f \cdot (1/g)$ and use the Product Rule. Note that in order to differentiate $1/g$, we view this as the composition of g with the function $1/x$ and apply the Chain Rule. □

Note: For all $k \in \mathbb{Z}$, we know now that $f(x) = x^k$ is differentiable everywhere it is well-defined (i.e. if $k < 0$, it is not defined for $x = 0$ of course), and that $f'(x) = kx^{k-1}$.

Proposition 2.20 *Let $r \in \mathbb{Q}$. Then, the function $f : (0, \infty) \to \mathbb{R}$ given by $f(x) = x^r$ is differentiable with $f'(x) = rx^{r-1}$.*

Proof: This is postponed; something more general is proven using the natural logarithm. □

2.3 Open Sets and the Localisation Lemma

Definition 2.21 A subset $U \subseteq \mathbb{R}$ is open if for every element $a \in U$, there exists $\delta > 0$ such that the interval $(a - \delta, a + \delta) \subseteq U$.

Remark The geometric intuition is that a subset of $\mathbb{R}$ is open if around any point of our subset, we can squeeze in an open interval that doesn't spill out of our subset. For instance, the subset $[0, 1)$ is **not** open because for $a = 0$, we want $(-\delta, \delta) \subseteq [0, 1)$ where $\delta > 0$; this is clearly not possible since $-\delta/2 \in (-\delta, \delta)$ but $-\delta/2 \notin [0, 1)$.

Note: Not only is $\mathbb{R}$ an open subset of itself, but the empty set $\emptyset$ **is** indeed an open subset.

Lemma 2.24 (Localisation Lemma) *Let $f : D \to \mathbb{R}$ coincide with a differentiable function $g : U \to \mathbb{R}$ on some open set $U \subseteq D$. Then, on U, f is differentiable with derivative g'.*

Proof: Let $a \in U$ be arbitrary. Per Definition 2.1, we must show that the limit

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a} = g'(a).$$

To that end, let $\varepsilon > 0$ be given. Since U is open, there exists $\delta_1 > 0$ where $(a - \delta_1, a + \delta_1) \subseteq U \subseteq D$. But since g is differentiable, there exists $\delta_2 > 0$ such that, for all $x \in U$ with $0 < |x - a| < \delta_2$,

$$\left| \frac{g(x) - g(a)}{x - a} - g'(a) \right| < \varepsilon.$$

Let $\delta := \min\{\delta_1, \delta_2\}$ so that both situations above happen concurrently. Then, for all $x \in D$ with $0 < |x - a| < \delta$, we know that $x \in U$. Because f and g coincide on U, this means that $f(x) = g(x)$ for all such $x \in U$, in particular $f(a) = g(a)$. Therefore, since $0 < |x - a| < \delta \leq \delta_2$,

$$\left| \frac{f(x) - f(a)}{x - a} - g'(a) \right| = \left| \frac{g(x) - g(a)}{x - a} - g'(a) \right| < \varepsilon. \qquad \square$$

Note: Having U be open in the statement of the Localisation Lemma is crucial. Indeed, $f(x) = |x|$ coincides with the differentiable function $g(x) = x$ on the **non-open** set $[0, \infty)$, but we know that f is **not** differentiable at zero.

3 Functions Differentiable on an Interval

3.1 The Interior Extremum Theorem

Reminder: Let $f : D \to \mathbb{R}$ be a function.
- It attains a maximum at $a \in D$ if $f(x) \leq f(a)$ for all $x \in D$.
- It attains a minimum at $a \in D$ if $f(x) \geq f(a)$ for all $x \in D$.
- It attains an extremum at $a \in D$ if it has either a maximum or minimum at a.

Theorem 3.1 (Interior Extremum Theorem) *Let $f : (a, b) \to \mathbb{R}$ be differentiable and attain an extremum at $c \in (a, b)$. Then, $f'(c) = 0$.*

Proof: Suppose first that f attains a maximum at $c \in (a, b)$. By assumption, $f'(c)$ exists. Hence, by Theorem 1.35, for any sequence (x_n) in $(a, b) \setminus \{c\}$ where $x_n \to c$, we know the sequence

$$s(x_n) := \frac{f(x_n) - f(c)}{x_n - c} \to f'(c).$$

In particular, this is true for the sequence $x_n = c + (b - c)(n + 1) \in (c, b)$. Note that $x_n > c$ for all c, so the denominator $x_n - c > 0$. Because $f(c)$ is the maximum, we know that $f(x_n) \leq f(c)$ and thus the numerator satisfies $f(x_n) - f(c) \leq 0$. Therefore, $s(x_n) \leq 0$ for every n, so it follows from Proposition 1.7 that $f'(c) \leq 0$. But on the other hand, this is also true for the sequence $x_n = c - (c - a)/(n + 1) \in (a, c)$. Note that $x_n < c$ for all c, so the denominator $x_n - c < 0$. But still we have that $f(c)$ is the maximum, so the numerator satisfies $f(x_n) - f(c) \leq 0$. Therefore, $s(x_n) \geq 0$ for every n, so it follows from Proposition 1.7 that $f'(c) \geq 0$. Combining these inequalities shows that $f'(c) = 0$.

Suppose next that f attains a minimum at $c \in (a, b)$. We can define the function $g(x) = -f(x)$. Then, g is differentiable on (a, b) by the linearity property, and it attains a maximum at $c \in (a, b)$. By the above argument, we know that $g'(c) = 0$. Hence, $f'(c) = -g'(c) = 0$. $\square$

Definition 3.2 Let $f : D \to \mathbb{R}$ be differentiable. Then, $c \in D$ is a critical point if $f'(c) = 0$.

The Interior Extremum Theorem says that any extremum is automatically a critical point.

Method – Finding Extrema: Suppose we want to find the extrema of $f : [a, b] \to \mathbb{R}$.
(i) Evaluate f at the endpoints $f(a)$ and $f(b)$, and at the critical points $c \in (a, b)$.
(ii) State which of the values from Step (i) are largest and smallest.

Proposition 3.4 *Let $f : [a, b] \to \mathbb{R}$ be differentiable.*
(i) *If $f'(a) > 0$, then $f(a)$ is* **not** *the maximum value attained by f.*
(ii) *If $f'(a) < 0$, then $f(a)$ is* **not** *the minimum value attained by f.*
(iii) *If $f'(b) > 0$, then $f(b)$ is* **not** *the minimum value attained by f.*
(iv) *If $f'(b) < 0$, then $f(b)$ is* **not** *the maximum value attained by f.*

Proof: (i) Assume to the contrary that f does attain a maximum at a. Since f is differentiable at a, for any sequence (x_n) in $[a, b] \setminus \{a\} = (a, b]$ where $x_n \to a$, we know that the sequence

$$s(x_n) := \frac{f(x_n) - f(a)}{x_n - a} \to f'(a) > 0.$$

But $f(x_n) \leq f(a)$ since we assume that $f(a)$ is a maximum. However, $x_n > a$ since our sequence (i.e. each term) lies in the interval $(a, b]$. Combining these inequalities tells us that $s(x_n) \leq 0$ for all n. But Proposition 1.7 tells us that $f'(a) \leq 0$, a contradiction. The proofs of the other statements (ii), (iii) and (iv) are near-identical, so we omit them. □

3.2 The Mean Value Theorem

Theorem 3.6 (Rolle's Theorem) *Let f be continuous on $[a, b]$ and differentiable on (a, b). If $f(a) = f(b)$, then there exists $c \in (a, b)$ such that $f'(c) = 0$.*

Proof: By the Extreme Value Theorem (Theorem 1.48), we know that f attains both a maximum and a minimum on $[a, b]$. If both of these occur at the endpoints, the fact that $f(a) = f(b)$ tells us that the maximum and minimum are equal; the function $f : [a, b] \to \mathbb{R}$ is therefore constant. Consequently, $f'(x) = 0$ for **all** $x \in (a, b)$, so it is true in particular for $c = (a + b)/2$.

On the other hand, if either the maximum or minimum does **not** occur at an endpoint, it must be that f attains an extremum at some interior point $c \in (a, b)$. But we know immediately that $f'(c) = 0$ by the Interior Extremum Theorem. □

Theorem 3.7 (Mean Value Theorem) *Let f be a real function that is continuous on $[a, b]$ and differentiable on (a, b). Then, there exists $c \in (a, b)$ such that*

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

Proof: Define $\alpha := (f(b) - f(a))/(b - a)$ and let the function $g : [a, b] \to \mathbb{R}$ be given by

$$g(x) := f(x) - \alpha(x - a).$$

It is clear that g is continuous, and differentiable on (a, b) since f is. But notice that $g(a) = f(a)$ and $g(b) = f(b) - \alpha(b - a) = f(b) - (f(b) - f(a)) = f(a)$. In other words, the values at the endpoints of g coincide and so it satisfies the hypotheses of Rolle's Theorem: there exists $c \in (a, b)$ such that $g'(c) = 0$. That being said, the linearity of the derivative and the Product Rule imply

$$0 = g'(c) = f'(c) - \alpha \qquad \Rightarrow \qquad f'(c) = \alpha.$$ □

Remark Consider the graph $y = f(x)$ of a differentiable function $f : \mathbb{R} \to \mathbb{R}$. For each pair of numbers (a, b) where $a < b$, we can construct the chord (line segment) between the points $(a, f(a))$ and $(b, f(b))$. The Mean Value Theorem asserts that at some point $(c, f(c))$ on the graph between the previous two points, the tangent line to the graph is parallel to the chord.

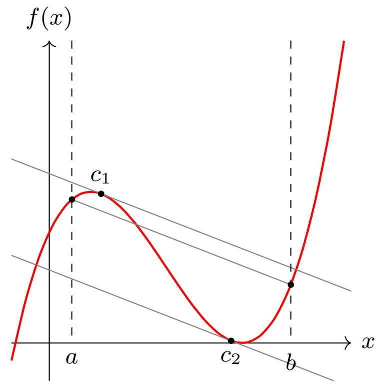


Figure 6: The geometric interpretation of the Mean Value Theorem.

The statement guarantees the existence of a point $c \in (a, b)$ with the required value of $f'(c)$, but it does **not** say it is unique. An example with two such points is presented in Figure 6 above.

Note: We can view Rolle's Theorem as a special case of the Mean Value Theorem, because setting $f(a) = f(b)$ will result in $\alpha = 0$ (geometrically, this means the chord is horizontal).

Corollary *Let $I \subseteq \mathbb{R}$ be an interval and $f : I \to \mathbb{R}$ be differentiable. If $f'(x) = 0$ for all $x \in I$, then f is constant.*

Proof: Assume to the contrary that f is **not** constant, that is there exist $a, b \in I$ with $a < b$ such that $f(a) \neq f(b)$. But we apply the Mean Value Theorem to f on the subinterval $[a, b]$ to determine that there exists a point $c \in (a, b) \subseteq I$ such that

$$f'(c) = \frac{f(b) - f(a)}{b - a} \neq 0.$$ □

Method – Solving Differential Equations: Suppose $f : \mathbb{R} \to \mathbb{R}$ is such that we know $f'(x)$ and an initial condition $f(0) = c$. We want to determine an expression for the function f.

(i) Find a function $F : \mathbb{R} \to \mathbb{R}$ whose derivative is $F'(x) = f'(x)$.
(ii) Define the function $g(x) := f(x) - F(x)$.
(iii) Differentiate to obtain $g'(x) = f'(x) - F'(x) = 0$, using Step (i).
(iv) Apply the corollary above to conclude that $g(x) = g(0) = 0$, i.e. $F(x) = f(x)$.

Definition 3.9 Let $f : D \to \mathbb{R}$ be a function.

- It is increasing if for all $x, y \in D$ with $x < y$, we have $f(x) \leq f(y)$.
- It is decreasing if for all $x, y \in D$ with $x < y$, we have $f(x) \geq f(y)$.
- It is strictly increasing if for all $x, y \in D$ with $x < y$, we have $f(x) < f(y)$.
- It is strictly decreasing if for all $x, y \in D$ with $x < y$, we have $f(x) > f(y)$.

Proposition 3.11 *Let $I \subseteq \mathbb{R}$ be an interval and $f : I \to \mathbb{R}$ be differentiable.*

(i) *For all $x \in I$, $f'(x) \geq 0$ if and only if f is increasing.*
(ii) *For all $x \in I$, $f'(x) \leq 0$ if and only if f is decreasing.*
(iii) *For all $x \in I$, $f'(x) = 0$ if and only if f is constant.*
(iv) *If $f'(x) > 0$ for all $x \in I$, then f is* **strictly** *increasing.*
(v) *If $f'(x) < 0$ for all $x \in I$, then f is* **strictly** *decreasing.*

Proof: (i) Assume to the contrary $f'(x) \geq 0$ for all $x \in I$ but that f is **not** increasing. Then, there exist $a, b \in I$ such that $a < b$ and $f(a) > f(b)$. But the Mean Value Theorem guarantees the following contradiction: there exists a point $c \in (a, b)$ such that

$$f'(c) = \frac{f(b) - f(a)}{b - a} < 0.$$

Conversely, assume f is increasing. If $a \in I$ is the left-hand endpoint, then f attains a minimum at a. It follows from the contrapositive of Proposition 3.4(ii) that $f'(a) \geq 0$. Similarly, if $b \in I$ is the right-hand endpoint, then f attains a maximum at b. It follows from the contrapositive of Proposition 3.4(iv) that $f'(b) \geq 0$. Now, let $x \in (a, b) \subseteq I$ be an interior point and define

$$x_n = x + \frac{b - x}{n + 1} \in (x, b) \subseteq I.$$

It is clear that $x_n \to x$ by the Algebra of Limits. Since f is differentiable (at x), we have

$$s(x_n) := \frac{f(x_n) - f(x)}{x_n - x} \to f'(x).$$

But f is increasing and $x_n > x$, so $s(x_n) \geq 0$. Hence, its limit $f'(x) \geq 0$ by Proposition 1.7.

(ii) This is very similar to (i) with some modifications, and thus is omitted.

(iii) If $f(x) = c$ for all $x \in I$, then $f'(x) = 0$ by an earlier lemma. The converse is proved in the previous corollary. Alternatively, if $f'(x) = 0$ for all $x \in I$, then we know that f is increasing by (i) and decreasing by (ii). The only instance that both can happen is if for all $x, y \in I$, we have $f(x) = f(y)$.

(iv) Assume to the contrary $f'(x) > 0$ for all $x \in I$ but that f is **not strictly** increasing. Then, there exist $a, b \in I$ such that $a < b$ and $f(a) \geq f(b)$. But the Mean Value Theorem guarantees the following contradiction: there exists a point $c \in (a, b)$ such that

$$f'(c) = \frac{f(b) - f(a)}{b - a} \leq 0.$$

(v) This is very similar to (iv) with some modifications, and thus is omitted. $\square$

Note: The converse of Proposition 3.11(iv) and (v) are false, e.g. $f(x) = x^3$ has $f'(0) = 0$.

Remark The condition that $I \subseteq \mathbb{R}$ is an interval in Proposition 3.11 is actually very important. Indeed, note that $f : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ given by $f(x) = 1/x$ has $f'(x) = -1/x^2 < 0$ everywhere, but $f(-1) = f(1)$. Thus, f is **not** decreasing. But this doesn't contradict Proposition 3.11 precisely because the domain of this function $\mathbb{R} \setminus \{0\}$ is **not** an interval.

Method – Showing Injectivity: Suppose we wish to show that a differentiable function $f : I \to \mathbb{R}$ is injective. It suffices to compute its derivative and prove $f'(x) > 0$ for all $x \in I$. Indeed, this means f is strictly increasing by Proposition 3.11(iv). So, $f(x) = f(y)$ implies $x = y$. If not, then $x < y$ without loss of generality and thus f is **not** increasing.

Proposition *The following function $f : \mathbb{R} \to \mathbb{R}$ is differentiable everywhere and is such that $f'(0) = 1 > 0$, but it is* **not** *increasing on* **any** *neighbourhood of zero*:

$$f(x) = \begin{cases} x + 2x^2 \sin\bigl(1/x\bigr) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}.$$

Proof: On $U = \mathbb{R} \setminus \{0\}$, f coincides with $g : U \to \mathbb{R}$ given by $g(x) = x + 2x^2 \sin\bigl(1/x\bigr)$. Because g is differentiable by the Product and Chain Rules, we know that f is differentiable on U by the Localisation Lemma. Consequently, for all $a \in U$, we have

$$f'(a) = g'(a) = 1 + 4a \sin\bigl(1/a\bigr) - 2\cos\bigl(1/a\bigr).$$

We consider the derivative at zero. So let (x_n) be any sequence in $\mathbb{R} \setminus \{0\}$ with $x_n \to 0$. Then,

$$s(x_n) := \frac{f(x_n) - f(0)}{x_n - 0} = 1 + 2x_n \sin\bigl(1/x_n\bigr).$$

Notice $-|x_n| \leq x_n \sin\bigl(1/x_n\bigr) \leq |x_n|$, so the Squeeze Rule implies that $x_n \sin\bigl(1/x_n\bigr) \to 0$. Hence, the Algebra of Limits tells us $s(x_n) \to 1$. By definition (and Theorem 1.35), f is differentiable at zero with $f'(0) = 1$. It remains to show that f is not increasing on a neighbourhood of zero. To that end, assume to the contrary that there exists $\varepsilon > 0$ such that $f : (-\varepsilon, \varepsilon) \to \mathbb{R}$ **is** increasing. Then, $f'(x) \geq 0$ for all $x \in (-\varepsilon, \varepsilon)$ by Proposition 3.11(i). But (by the Archimedean Property) we can find $m \in \mathbb{Z}^+$ such that $m > 1/(2\pi\varepsilon)$. Note that $a = 1/(2\pi m) \in (0, \varepsilon) \subseteq U$. Substituting it into the above derivative formula, we obtain this contradiction:

$$f'(a) = -1 < 0. \qquad \square$$

3.3 Darboux's Theorem

Note: We know that any differentiable function $f : D \to \mathbb{R}$ is automatically continuous by Proposition 2.8. But it is **not** true in general that $f' : D \to \mathbb{R}$ is itself continuous (e.g. the function in the above proposition is differentiable but not continuous at zero).

Theorem 3.15 (Darboux's Theorem) *Let f be differentiable on $[a, b]$ and k be a number between $f'(a)$ and $f'(b)$. Then, there exists $c \in [a, b]$ such that $f'(c) = k$.*

Proof: If $k = f'(a)$ or $k = f'(b)$, then we can take $c = a$ and $c = b$, respectively. Otherwise, there are two cases to consider: $f'(a) < k < f'(b)$ and $f'(b) < k < f'(a)$. We will prove the first case directly, and then use a trick to transform the second case into the first case. Indeed, assume $f'(a) < k < f'(b)$ and consider the function

$$g(x) \coloneqq kx - f(x).$$

This is differentiable with derivative $g'(x) = k - f'(x)$, and therefore continuous, on the interval $[a, b]$. Thus, the Extreme Value Theorem implies g attains a maximum somewhere on $[a, b]$. We see from the inequality we assume at the start that

$$g'(a) = k - f'(a) > 0 \qquad \text{and} \qquad g'(b) = k - f'(b) < 0.$$

The first tells us that the maximum can't be at a, and the second tells us that the maximum can't be at b (Proposition 3.4). Therefore, the maximum is obtained at some interior point $c \in (a, b)$. But $g'(c) = 0$ by the Interior Extremum Theorem. Substituting this into the expression for g' tells us that $0 = k - f'(c)$, which rearranges to $k = f'(c)$.

For the second case, assume $f'(b) < k < f'(a)$ and define the function $h(x) \coloneqq -f(x)$. Then, we see that h is differentiable on $[a, b]$ and satisfies $h'(a) < -k < h'(b)$. But by the above argument, there exists $c \in (a, b)$ such that $h'(c) = -k$. This is equivalent to $f'(c) = k$, so we are done. $\square$

> **Note:** The slogan for Darboux's Theorem: f' has the *intermediate value property* on $[a, b]$.

3.4 The Extended Mean Value Theorem and L'Hôpital's Rule

> **Theorem 3.16** (Extended Mean Value Theorem) *Let f and g be real functions that are continuous on $[a, b]$ and differentiable on (a, b), where also $g'(x) \neq 0$ for all $x \in (a, b)$. Then, there exists $c \in (a, b)$ such that*
>
> $$\frac{f'(c)}{g'(c)} = \frac{f(b) - f(a)}{g(b) - g(a)}.$$

Proof: Let's call the right-hand side of the above equality α, and note that it is a constant. This is well-defined since $g(a) \neq g(b)$. Indeed, if $g(a) = g(b)$, then Rolle's Theorem implies there exists $c \in (a, b)$ with $g'(c) = 0$, contradicting the fact $g' \neq 0$ on this open interval. We now define

$$h(x) \coloneqq f(x) - \alpha g(x).$$

This is also continuous on $[a, b]$ and differentiable on (a, b), since it is a linear combination of functions with these properties. By linearity, we see that $h'(x) = f'(x) - \alpha g'(x)$. Furthermore,

$$h(a) = \frac{f(a)g(b) - f(b)g(a)}{g(b) - g(a)}$$

by substituting and cancelling. But if we interchange a and b, we get the same expression (this can be checked explicitly). In other words, $h(a) = h(b)$. This allows us to apply Rolle's Theorem to h, so there exists $c \in (a, b)$ such that $h'(c) = 0$. But this means that $f'(c) - \alpha g'(c) = 0$, equivalent to the expression $f'(c)/g'(c) = \alpha$ we wanted to show. $\square$

Note: If we apply the Extended Mean Value Theorem in the case where $g(x) = x$, we reduce to the usual Mean Value Theorem. Even better, if we apply it to $g(x) = x - a$, this reduction still holds and the **proof** of Theorem 3.16 reduces to the **proof** of Theorem 3.7.

Remark Similar to that of the Mean Value Theorem, we can give a geometric interpretation of the Extended Mean Value Theorem. However, instead of considering a graph, we consider a parametrised curve whose components are g and f. Namely, let us define the parametrised curve

$$\gamma : [a, b] \to \mathbb{R}^2, \qquad \gamma(t) = (g(t), f(t)).$$

The tangent to γ at $\gamma(t)$ is parallel to the vector $\gamma'(t) = (g'(t), f'(t))$. Therefore, its slope is precisely $f'(t)/g'(t)$. The Extended Mean Value Theorem says that there is a point on γ between its start $\gamma(a)$ and end $\gamma(b)$ at which its tangent is parallel to the chord connecting the endpoints.

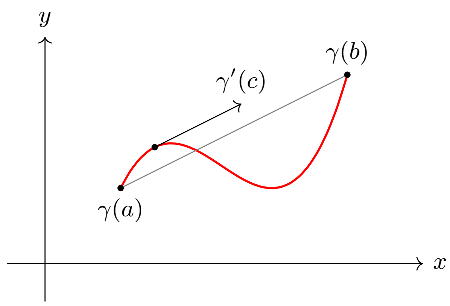


Figure 7: The geometric interpretation of the Extended Mean Value Theorem.

Theorem 3.17 (L'Hôpital's Rule) *Let $I \subseteq \mathbb{R}$ be an open interval and $f, g : I \to \mathbb{R}$ be differentiable functions with $f(a) = g(a) = 0$ for some $a \in I$, where also $g'(x) \neq 0$ for all $x \in I$ and $g(x) \neq 0$ for all $x \in I \setminus \{a\}$. Then, it follows that*

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{x \to a} \frac{f'(x)}{g'(x)}.$$

Proof: Let (x_n) be any sequence in $I \setminus \{a\}$ where $x_n \to a$. For each $n \in \mathbb{Z}^+$, consider f and g restricted to the closed interval with endpoints a and x_n. These functions satisfy the hypotheses of the Extended Mean Value Theorem. As such, there exists c_n between a and x_n such that

$$\frac{f'(c_n)}{g'(c_n)} = \frac{f(x_n) - f(a)}{g(x_n) - g(a)} = \frac{f(x_n)}{g(x_n)}.$$

But $c_n \to a$ by the Squeeze Rule. Since we assume that the limit $L \coloneqq \lim_{x \to a} f'(x)/g'(x)$ exists, $f'(c_n)/g'(c_n) \to L$ by Theorem 1.35, and this is the same as $f(x_n)/g(x_n) \to L$ by the above. $\square$

3.5 Higher Derivatives and Taylor's Theorem

Definition 3.19 A function $f : D \to \mathbb{R}$ is continuously differentiable if it is differentiable and its derivative $f' : D \to \mathbb{R}$ is continuous. We say it is n-times continuously differentiable if all its derivatives up to the n^{th} exist everywhere on the domain D and are continuous. We say that f is smooth if it is n-times continuously differentiable for **every** $n \in \mathbb{Z}^+$.

Notation The set of n-times continuously differentiable functions on the domain D is denoted $C^n(D)$. If $f \in C^n(D)$, we say that "f is C^n". Similarly, the set of smooth functions on D is denoted $C^\infty(D)$. For completeness, the set of continuous functions on D is denoted $C^0(D)$.

Theorem 3.21 (Taylor's Theorem) *Let $I \subseteq \mathbb{R}$ be an interval, $f : I \to \mathbb{R}$ be $(n+1)$-times differentiable and $a, x \in I$. Then, there exists c between a and x such that we can write*

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}.$$

Proof: If $x = a$, the claim is trivial since we can choose $c = a$. Hence, let's assume now that $x \neq a$. We consider a function similar to the one in the statement, namely $F : I \to \mathbb{R}$ given by

$$F(t) = f(x) - f(t) - f'(t)(x-t) - \frac{f''(t)}{2!}(x-t)^2 - \cdots - \frac{f^{(n)}(t)}{n!}(x-t)^n.$$

The claim to be proved is that there exists c between a and x whereby

$$F(a) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}.$$

By linearity and the Product Rule, we see that F is differentiable and compute

$$F'(t) = -\frac{f^{(n+1)}(t)}{n!}(x-t)^n.$$

Define $G(t) \coloneqq (x-t)^{n+1}$, which has derivative $G'(t) = -(n+1)(x-t)^n$. Now, F and G satisfy the hypotheses of the Extended Mean Value Theorem; there exists c between a and x such that

$$\frac{F'(c)}{G'(c)} = \frac{F(x) - F(a)}{G(x) - G(a)} \qquad \Leftrightarrow \qquad \frac{-f^{(n+1)}(c)(x-c)^n}{n!\,\big(-(n+1)(x-c)^n\big)} = \frac{F(a)}{(x-a)^{n+1}}.$$

Multiplying by $(x-a)^{n+1}$ gives us precisely the expression we wanted for $F(a)$. □

Note: The n^{th} Taylor approximant of f about a refers to the polynomial

$$p_n(x) \coloneqq f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n,$$

and the remainder refers to the additional term

$$\frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}.$$

We know from MATH1026 that the following series converges, using the Alternating Series Test, but this does not give us a way to access the actual *limit* of this series. However, Taylor's Theorem provides the missing piece to this puzzle.

Proposition (Alternating Harmonic Series) *The series* $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}$ *converges to* $\ln(2)$.

Proof: Let $f : (0, \infty) \to \mathbb{R}$ be given by $f(x) = \ln(x)$. This is a smooth function with

$$f'(x) = \frac{1}{x}, \qquad f''(x) = -\frac{1}{x^2}, \qquad f'''(x) = \frac{2}{x^3}, \qquad f^{(4)}(x) = -\frac{6}{x^4}, \dots.$$

Indeed, we can prove by induction that $f^{(m)}(x) = (-1)^{m+1}(m-1)!x^{-m}$ for all $m \in \mathbb{Z}^+$. The base case ($m = 1$) says that $f^{(1)}(x) = x^{-1}$ which agrees with the explicit derivative computed above. As for the inductive step, assume it is true for $m = k$, that is

$$f^{(k)}(x) = (-1)^{k+1}(k-1)!x^{-k}.$$

But we can view $f^{(k+1)}(x)$ as the derivative of the above, so differentiating gives us

$$f^{(k+1)}(x) = (-1)^{k+1}(k-1)!(-k)x^{-k-1} = (-1)^{k+2}k!x^{-(k+1)}.$$

It follows by induction that the derivative formula is correct. We use this to see that, for each $m \in \mathbb{Z}^+$, $f^{(m)}(1) = (-1)^{m+1}(m-1)!$. Next, the n^{th} Taylor approximant for f about 1 is

$$\begin{aligned} p_n(x) &= f(1) + f'(1)(x-1) + \frac{f''(1)}{2!}(x-1)^2 + \cdots + \frac{f^{(n)}(1)}{n!}(x-1)^n \\ &= 0 + (x-1) - \frac{1}{2}(x-1)^2 + \cdots + \frac{(-1)^{n+1}(n-1)!}{n!}(x-1)^n \\ &= (x-1) - \frac{1}{2}(x-1)^2 + \frac{1}{3}(x-1)^3 + \cdots + \frac{(-1)^{n+1}}{n}(x-1)^n. \end{aligned}$$

Evaluating this at $x = 2$, we have precisely the n^{th} partial sum of the alternating harmonic series:

$$p_n(2) = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots + \frac{(-1)^{n+1}}{n}.$$

But Taylor's Theorem applies, and gives us the existence of some $c \in (1, 2)$ such that

$$f(2) = p_n(2) + \frac{f^{(n+1)}(c)}{(n+1)!}(2-1)^{n+1} = p_n(2) + \frac{(-1)^{n+2}n!c^{-(n+1)}}{(n+1)!} = p_n(2) + \frac{(-1)^{n+2}}{n+1}\frac{1}{c^{n+1}}.$$

Therefore, the size of the error term when approximating via Taylor approximant is precisely

$$0 < \left|f(2) - p_n(2)\right| = \left|\frac{(-1)^{n+1}}{n+1}\frac{1}{c^{n+1}}\right| = \frac{1}{n+1}\frac{1}{c^{n+1}} < \frac{1}{n+1}.$$

Applying the Squeeze Rule, the sequence of partial sums $p_n(2) \to f(2) = \ln(2)$ as required. $\square$

4 Integration

4.1 Suprema and Infima

Reminder: Let $A \subseteq \mathbb{R}$ be any subset.

- We say A is bounded above if there exists $K \in \mathbb{R}$ such that $x \leq K$ for all $x \in A$.
- We say A is bounded below if there exists $L \in \mathbb{R}$ such that $x \geq L$ for all $x \in A$.
- We say A is bounded if it is bounded both above and below.

Definition 4.1 Let $A \subseteq \mathbb{R}$ be non-empty. If they exist, the supremum $\sup(A)$ of A is the least upper bound on A, and the infimum $\inf(A)$ of A is the greatest lower bound on A.

In other words, the supremum is an upper bound with the property that no number less than it is also an upper bound. Similarly, the infimum is a lower bound with the property that no number greater than it is also a lower bound.

Note: There is a difference between a supremum and a maximum (similarly for an infimum and a minimum). For example, $A = (1, 2)$ has no maximal element, but $\sup(A) = 2 \notin A$.

One of the defining properties of the real numbers is the below axiom (an additional assumption).

Axiom (Completeness) *Every non-empty subset of $\mathbb{R}$ bounded above has a supremum in $\mathbb{R}$.*

Theorem 4.4 *Every non-empty subset of $\mathbb{R}$ bounded below has an infimum in $\mathbb{R}$.*

Proof (Obvious): Let $A \subseteq \mathbb{R}$ be such a non-empty subset and suppose $B \coloneqq \{-x : x \in A\} \subseteq \mathbb{R}$. Because A is non-empty, so too is B. Moreover, if $L \in \mathbb{R}$ is a lower bound of A, this means $x \geq L$ for all $x \in A$. Therefore, $-x \leq -L$ for all $x \in A$ and so $-L$ is an upper bound of B. By the Axiom of Completeness, the number $K \coloneqq \sup(B) \in \mathbb{R}$ exists. It remains to prove that $-\sup(B) = \inf(A)$. Indeed, assume that $M \in \mathbb{R}$ satisfies $M \geq -\sup(B)$. But it follows that $-M \leq \sup(B)$. But by definition of a supremum, there exists $x \in B$ such that $x \geq -M$. This is equivalent to $-x \leq M$. Because $x \in B$, by definition we have $-x \in A$ and so M is **not** a lower bound on A. Hence, $-\sup(B)$ is the greatest lower bound on A, i.e. it is precisely $\inf(A)$. $\square$

Proof (Sneaky): Let $A \subseteq \mathbb{R}$ be such a non-empty subset and suppose B is the set of lower bounds on A. Because A is assumed to be bounded below, we know that $B \neq \emptyset$. Furthermore, for all $y \in B$ and all $x \in A$, we have $x \geq y$ by definition of bounded below. Therefore, every $x \in A$ is an upper bound on B. So we can apply the Axiom of Completeness to deduce that $K \coloneqq \sup(B) \in \mathbb{R}$ exists. Since K is the **least** upper bound on B and every $x \in A$ is an upper bound on B, we know that $x \geq K$ for all $x \in A$. This tells us K is a lower bound on A. But any $L > K$ is not in the set B (since K is an upper bound on B), and so any such $L > K$ is not a lower bound on A. This establishes that K is the **greatest** lower bound on B, i.e. $K = \inf(A)$. $\square$

Note: Suppose that $A \subseteq B \subseteq \mathbb{R}$ such that their suprema and infima exist. It follows that

$$\sup(A) \leq \sup(B) \qquad \text{and} \qquad \inf(A) \geq \inf(B).$$

Definition 4.5 The diameter of a bounded set $A \subseteq \mathbb{R}$ is $\operatorname{diam}(A) = \sup\{|x-y| : x, y \in A\}$.

Proposition 4.7 *The diameter is given by* $\operatorname{diam}(A) = \sup(A) - \inf(A)$.

Proof: Consider the set $B \coloneqq \{|x-y| : x, y \in A\} = \{x - y : x, y \in A \text{ and } x \geq y\}$. Notice that Definition 4.5 declares that $\operatorname{diam}(A) = \sup(B)$. For all $x, y \in A$, we have $x \leq \sup(A)$ and $y \geq \inf(A)$ by definition of the supremum and infimum. Therefore, $x - y \leq \sup(A) - \inf(A)$. This tells us that $\sup(A) - \inf(A)$ is an upper bound on B. To prove that this is $\operatorname{diam}(A) = \sup(B)$, we now aim to show it is the **least** upper bound. Firstly, notice that $0 \in B$ so no $K < 0$ can be an upper bound on B. Let's instead restrict to $0 \leq K < \sup(A) - \inf(A)$. Define the number

$$\varepsilon \coloneqq \frac{1}{2}(\sup(A) - \inf(A) - K) > 0.$$

Note that $\sup(A) - \varepsilon$ is **not** an upper bound on A, and $\inf(A) + \varepsilon$ is **not** a lower bound on A. So, there exists $x, y \in A$ such that $x > \sup(A) - \varepsilon$ and $y < \inf(A) + \varepsilon$. Combining these yields

$$x - y > \sup(A) - \varepsilon - (\inf(A) + \varepsilon) = K \geq 0.$$

It follows from this that $x \geq y$ and so $x - y \in B$, but also that $x - y > K$ so K is **not** an upper bound on B. Consequently, we have shown that $\sup(A) - \inf(A) = \sup(B) = \operatorname{diam}(A)$. $\square$

Lemma 4.9 (Reverse Triangle Inequality) *For all* $x, y \in \mathbb{R}$, *we have* $|x - y| \geq \big||x| - |y|\big|$.

Proof: Let $a = x - y$ and $b = y$. Then, the usual Triangle Inequality tells us that

$$|x| = |a + b| \leq |a| + |b| = |x - y| + |y| \qquad \Rightarrow \qquad |x - y| \geq |x| - |y|.$$

But repeating this argument with $a = y - x$ and $b = x$ tells us that

$$|y| = |a + b| \leq |a| + |b| = |y - x| + |x| \qquad \Rightarrow \qquad |x - y| \geq |y| - |x|.$$

Combining these and using the fact $\max\{|x| - |y|, |y| - |x|\} = \big||x| - |y|\big|$ gives the result. $\square$

Lemma 4.8 *Let* $A \subseteq \mathbb{R}$ *and* $|A| \coloneqq \{|x| : x \in A\}$. *Then,* $\operatorname{diam}(|A|) \leq \operatorname{diam}(A)$.

Proof: For all $x, y \in A$, $\big||x| - |y|\big| \leq |x - y| \leq \operatorname{diam}(A)$ by the Reverse Triangle Inequality and Definition 4.5. This tells us that $\operatorname{diam}(A)$ is an upper bound on the set $\{\big||x| - |y|\big| : x, y \in A\}$. But by definition again, we have $\operatorname{diam}(|A|) = \sup\{\big||x| - |y|\big| : x, y \in A\} \leq \operatorname{diam}(A)$. $\square$

4.2 Dissections and Riemann Sums

Definition 4.10 A dissection of a closed bounded interval $[a,b]$ is a **finite** subset $\mathscr{D}$ of $[a,b]$ containing both a and b. By convention, if $\mathscr{D}$ has $n+1$ elements, we call $\mathscr{D}$ a dissection of size n and label its elements $a_0, a_1, ..., a_{n-1}, a_n$ in such a way that

$$a = a_0 < a_1 < \cdots < a_{n-1} < a_n = b.$$

Remark The size of the dissection is **one less** than the number of elements it contains since the size actually counts the number of subintervals $[a_{j-1}, a_j]$ into which the dissection divides $[a,b]$.

Note: A dissection $\mathcal{D}$ is called regular if the points in it are regularly spaced, i.e. for all j,

$$a_j - a_{j-1} = \frac{b-a}{n}.$$

Let $f : [a,b] \to \mathbb{R}$ be bounded and $\mathscr{D}$ be a dissection of size n of $[a,b]$. Throughout, we will use

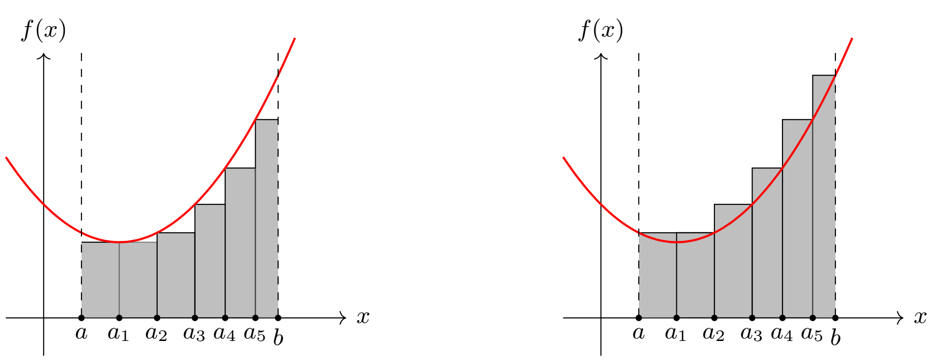
$$m_j \coloneqq \inf\{f(x) : x \in [a_{j-1}, a_j]\} \qquad \text{and} \qquad M_j \coloneqq \sup\{f(x) : x \in [a_{j-1}, a_j]\}. \tag{†}$$

Definition 4.11 Let $f : [a,b] \to \mathbb{R}$ be bounded and $\mathscr{D}$ be a dissection of size n of $[a,b]$.

(i) The lower Riemann sum of f with respect to $\mathscr{D}$ is $l_{\mathscr{D}}(f) = \sum_{j=1}^{n} m_j(a_j - a_{j-1})$.

(ii) The upper Riemann sum of f with respect to $\mathscr{D}$ is $u_{\mathscr{D}}(f) = \sum_{j=1}^{n} M_j(a_j - a_{j-1})$.

Remark If $f(x) \geq 0$ for all $x \in [a,b]$, we can visualise each of the lower and upper Riemann sums geometrically. Indeed, the lower Riemann sum gives the total area of the tallest rectangles with bases $[a_{j-1}, a_j]$ that fit under the graph $y = f(x)$. Similarly, the upper Riemann sum gives the total area of the shortest rectangles with bases $[a_{j-1}, a_j]$ that fit over the graph $y = f(x)$.



(a) The lower Riemann sum.

(b) The upper Riemann sum.

Figure 8: The geometric interpretation of the lower and upper Riemann sums.

Proposition 4.13 *Let $f : [a, b] \to \mathbb{R}$ be bounded above by M and below by m, and $\mathscr{D}$ be any dissection of $[a, b]$. Then, we have*

$$m(b-a) \leq l_{\mathscr{D}}(f) \leq u_{\mathscr{D}}(f) \leq M(b-a).$$

Proof: Let $\mathscr{D} = \{a_0, a_1, ..., a_n\}$ and m_j and M_j be defined as in (†). Then, by definition of upper and lower bounds, it is clear that $m \leq m_j \leq M_j \leq M$; the central inequality here implies that

$$l_{\mathscr{D}}(f) = \sum_{j=1}^{n} m_j(a_j - a_{j-1}) \leq \sum_{j=1}^{n} M_j(a_j - a_{j-1}) = u_{\mathscr{D}}(f).$$

As for the inequalities on the far left and far right, we respectively obtain the following:

$$l_{\mathscr{D}}(f) = \sum_{j=1}^{n} m_j(a_j - a_{j-1}) \geq \sum_{j=1}^{n} m(a_j - a_{j-1}) = m(a_n - a_0) = m(b-a)$$

and

$$u_{\mathscr{D}}(f) = \sum_{j=1}^{n} M_j(a_j - a_{j-1}) \leq \sum_{j=1}^{n} M(a_j - a_{j-1}) = M(a_n - a_0) = M(b-a). \quad \square$$

Note: Hence, Proposition 4.13 tells us that the set of lower Riemann sums is bounded above by $M(b-a)$ and the set of upper Riemann sums is bounded below by $m(b-a)$.

4.3 Definition of the Riemann Integral

Definition 4.14 Let $f : [a, b] \to \mathbb{R}$ be bounded.

(i) The lower Riemann integral of f is $l(f) = \sup\{l_{\mathscr{D}}(f) : \mathscr{D} \text{ is any dissection of } [a, b]\}$.

(ii) The upper Riemann integral of f is $u(f) = \inf\{l_{\mathscr{D}}(f) : \mathscr{D} \text{ is any dissection of } [a, b]\}$.

(iii) We say that f is Riemann integrable on $[a, b]$ if $l(f) = u(f)$.

We call the common number $l(f) = u(f)$ the Riemann integral of f on $[a, b]$ and denote it

$$\int_a^b f \qquad \text{or} \qquad \int_a^b f(x)\,\mathrm{d}x.$$

Remark Thinking geometrically, the lower Riemann integral can be thought of as the least upper bound on the collection of all underestimates of the area under the graph $y = f(x)$. Similarly, the upper Riemann integral is the greatest lower bound on the collection of all overestimates of the area under the graph $y = f(x)$. Loosely speaking, the number $\int_a^b f$ is the unique (if it exists) number which is smaller than every overestimate and larger than every underestimate.

Definition 4.15 We say that a dissection $\mathscr{D}'$ of $[a, b]$ is a refinement of a dissection $\mathscr{D}$ also of $[a, b]$ if we have $\mathscr{D}' \supseteq \mathscr{D}$. In other words, $\mathscr{D}$ is obtained from $\mathscr{D}'$ by throwing away some number of points. If $\mathscr{D}' \setminus \mathscr{D}$ contains k points, we call $\mathscr{D}'$ a k-point refinement of $\mathscr{D}$.

Note: For any dissection $\mathscr{D}$ of $[a,b]$, the unique 0-point refinement of $\mathscr{D}$ is simply $\mathscr{D}$ itself.

Lemma 4.16 (Refinement Lemma) *Let $f:[a,b]\to\mathbb{R}$ be bounded and $\mathscr{D},\mathscr{D}'$ be dissections of $[a,b]$ such that $\mathscr{D}'$ is a refinement of $\mathscr{D}$. Then,*

$$l_{\mathscr{D}}(f)\le l_{\mathscr{D}'}(f)\le u_{\mathscr{D}'}(f)\le u_{\mathscr{D}}(f).$$

Proof: We first prove this result for the case where $\mathscr{D}'=\mathscr{D}\cup\{z\}$ for $z\notin\mathscr{D}$, meaning $\mathscr{D}'$ is a one-point refinement of $\mathscr{D}$. Let $\mathscr{D}=\{a_0,a_1,...,a_n\}$. Because $z\notin\mathscr{D}$, it lies in a subinterval defined by $\mathscr{D}$, i.e. there exists $k\in\{1,...,n\}$ such that $z\in(a_{k-1},a_k)$. Alongside the notation of (†) used before, we define the following:

$$\begin{aligned}
m'&=\inf\{f(x):x\in[a_{k-1},z]\}, & M'&=\sup\{f(x):x\in[a_{k-1},z]\},\\
m''&=\inf\{f(x):x\in[z,a_k]\}, & M''&=\sup\{f(x):x\in[z,a_k]\}.
\end{aligned}$$

Because $[a_{k-1},z],[z,a_k]\subseteq[a_{k-1},a_k]$, we know immediately that $m',m''\ge k_m$ and $M',M''\le M_k$. As for the lower and upper Riemann sums with respect to this one-point refinement, we have

$$\begin{aligned}
l_{\mathscr{D}'}(f)&=\sum_{k\ne j=1}^{n}m_j(a_j-a_{j-1})+m'(z-a_{k-1})+m''(a_k-z)\\
&=l_{\mathscr{D}}(f)-m_k(a_k-a_{k-1})+m'(z-a_{k-1})+m''(a_k-z)\\
&\ge l_{\mathscr{D}}(f)-m_k(a_k-a_{k-1})+m_k(z-a_{k-1})+m_k(a_k-z)\\
&=l_{\mathscr{D}}(f)
\end{aligned}$$

and

$$\begin{aligned}
u_{\mathscr{D}'}(f)&=\sum_{k\ne j=1}^{n}M_j(a_j-a_{j-1})+M'(z-a_{k-1})+M''(a_k-z)\\
&=u_{\mathscr{D}}(f)-M_k(a_k-a_{k-1})+M'(z-a_{k-1})+M''(a_k-z)\\
&\le u_{\mathscr{D}}(f)-M_k(a_k-a_{k-1})+M_k(z-a_{k-1})+M_k(a_k-z)\\
&=u_{\mathscr{D}}(f).
\end{aligned}$$

As Proposition 4.13 tells us $l_{\mathscr{D}'}(f)\le u_{\mathscr{D}'}(f)$, we can stitch together the above to conclude that

$$l_{\mathscr{D}}(f)\le l_{\mathscr{D}'}(f)\le u_{\mathscr{D}'}(f)\le u_{\mathscr{D}}(f).$$

We can now proceed inductively. Namely, let $\mathscr{D}'=\mathscr{D}\cup\{z_1,...,z_k\}$ be a k-point refinement of $\mathscr{D}$ where each $z_i\notin\mathscr{D}$. The trick is to define a chain of dissections $\mathscr{D}_i$ for $i=0,..,k$ where $\mathscr{D}_0=\mathscr{D}$ and $\mathscr{D}_i=\mathscr{D}_{i-1}\cup\{z_i\}$. Note that $\mathscr{D}_k=\mathscr{D}'$, and each $\mathscr{D}_i$ is a one-point refinement of $\mathscr{D}_{i-1}$. Hence, $l_{\mathscr{D}}(f)\le l_{\mathscr{D}_1}(f)\le\cdots\le l_{\mathscr{D}_{k-1}}(f)\le l_{\mathscr{D}'}(f)$ and $u_{\mathscr{D}}(f)\ge u_{\mathscr{D}_1}(f)\ge\cdots\ge u_{\mathscr{D}_{k-1}}(f)\ge u_{\mathscr{D}'}(f)$ by the argument above for the one-point refinement case. By again using Proposition 4.13 to see that $l_{\mathscr{D}'}(f)\le u_{\mathscr{D}'}(f)$, we can stitch these together and conclude precisely the result we are after:

$$l_{\mathscr{D}}(f)\le l_{\mathscr{D}'}(f)\le u_{\mathscr{D}'}(f)\le u_{\mathscr{D}}(f).$$

□

Corollary 4.17 *Let $f : [a, b] \to \mathbb{R}$ be bounded and $\mathscr{D}, \widehat{\mathscr{D}}$ be* **any** *dissections of $[a, b]$. Then,*

$$l_{\mathscr{D}}(f) \leq u_{\widehat{\mathscr{D}}}(f).$$

Proof: Let $\mathscr{D}' := \mathscr{D} \cup \widehat{\mathscr{D}}$, a refinement of both $\mathscr{D}$ and of $\widehat{\mathscr{D}}$. Hence, the Refinement Lemma says

$$l_{\mathscr{D}}(f) \leq l_{\mathscr{D}'}(f) \leq u_{\mathscr{D}'}(f) \leq u_{\widehat{\mathscr{D}}}(f). \qquad \square$$

Lemma 4.18 *Let $f : [a, b] \to \mathbb{R}$ be bounded. Then, $l(f) \leq u(f)$.*

Proof: Let $\mathbb{D}$ be the set of **all** dissections of $[a, b]$. In this notation, Definition 4.14 introduces

$$l(f) = \sup\{l_{\mathscr{D}}(f) : \mathscr{D} \in \mathbb{D}\} \qquad \text{and} \qquad u(f) = \inf\{u_{\mathscr{D}}(f) : \mathscr{D} \in \mathbb{D}\}.$$

Assume to the contrary that $l(f) > u(f)$. Then, $l(f)$ is **not** a lower bound on the set of all upper Riemann sums $\{u_{\mathscr{D}}(f) : \mathscr{D} \in \mathbb{D}\}$. Hence, there exists some $\mathscr{D} \in \mathbb{D}$ such that $u_{\mathscr{D}}(f) < l(f)$. But then, $u_{\mathscr{D}}(f)$ is **not** an upper bound on the set of all lower Riemann sums $\{l_{\mathscr{D}}(f) : \mathscr{D} \in \mathbb{D}\}$. Similarly, there exists $\widehat{\mathscr{D}} \in \mathbb{D}$ such that $l_{\widehat{\mathscr{D}}}(f) > u_{\mathscr{D}}(f)$. But this contradicts Corollary 4.17. $\square$

Proposition *The function $f : [0, 1] \to \mathbb{R}$ given by $f(x) = x^2$ is Riemann integrable.*

Proof: For each $n \in \mathbb{Z}^+$, let $\mathscr{D}_n$ be the regular dissection of $[0, 1]$ of size n. Explicitly, this means

$$\mathscr{D}_n = \left\{0, \frac{1}{n}, \frac{2}{n}, ..., \frac{n-1}{n}, 1\right\}.$$

Each subinterval $[a_{j-1}, a_j] = [\frac{j-1}{n}, \frac{j}{n}]$ has width $1/n$. Since f is increasing, it is straightforward to calculate explicitly the numbers m_j and M_j introduced in (†). Indeed, one can see that

$$m_j = \inf\{f(x) : x \in [a_{j-1}, a_j]\} = \frac{(j-1)^2}{n^2} \qquad \text{and} \qquad M_j = \sup\{f(x) : x \in [a_{j-1}, a_j]\} = \frac{j^2}{n^2}.$$

By using the sums of squares formula, we can compute the lower and upper Riemann sums as

$$l_{\mathscr{D}_n}(f)\frac{1}{6}\left(1 - \frac{1}{n}\right)\left(2 - \frac{1}{n}\right) \qquad \text{and} \qquad u_{\mathscr{D}_n}(f) = \frac{1}{6}\left(1 + \frac{1}{n}\right)\left(2 + \frac{1}{n}\right).$$

But by definition, we know that $l(f) \geq l_{\mathscr{D}_n}(f)$. Since we can see from above that $l_{\mathscr{D}_n}(f) \to 1/3$ by the Algebra of Limits, it follows from Proposition 1.7 that $l(f) \geq 1/3$ also. Arguing similarly, $u(f) \leq u_{\mathscr{D}_n}(f) \to 1/3$, so the same result tells us that $u(f) \leq 1/3$. Combining these produces

$$u(f) \leq \frac{1}{3} \leq l(f).$$

Finally, Lemma 4.18 tells us that $l(f) \leq u(f)$ always, so it must be that $l(f) = u(f)$, i.e. f is Riemann integrable. The Riemann integral is this common value, so

$$\int_0^1 f = \frac{1}{3}. \qquad \square$$

Note: The only dissection of the interval $[a, a] = \{a\}$ is the singleton set $\mathscr{D} = \{a\}$. Because every function $f : [a, a] \to \mathbb{R}$ is bounded above and below by $f(a)$, one can see easily that $l_{\{a\}}(f) = u_{\{a\}}(f) = 0$. In other words, $\int_a^a f = 0$. For convenience, when $a \leq b$, we define

$$\int_b^a f := -\int_a^b f.$$

4.4 A Sequential Characterisation of Integrability

Theorem 4.21 *Let $f : [a, b] \to \mathbb{R}$ be bounded. Then, f is Riemann integrable if and only if there exists a sequence $(\mathscr{D}_n)$ of dissections of $[a, b]$ such that*

$$u_{\mathscr{D}_n}(f) - l_{\mathscr{D}_n}(f) \to 0.$$

Proof: Let $\mathbb{D}$ be the set of all dissections of $[a, b]$. Similarly, we let $\mathbb{L}$ and $\mathbb{U}$ be the set of all lower and upper Riemann sums, respectively. This means we have $l(f) = \sup(\mathbb{L})$ and $u(f) = \inf(\mathbb{U})$.

($\Rightarrow$) Assume that f is Riemann integrable, so $l(f) = u(f)$. For each $n \in \mathbb{Z}^+$, notice that $u(f) + 1/n > u(f)$. Since $u(f)$ is the infimum of $\mathbb{U}$, we know that $u(f) + 1/n$ is **not** a lower bound on $\mathbb{U}$. Hence, there exists some dissection $\mathscr{D}'_n \in \mathbb{D}$ such that $u_{\mathscr{D}'_n}(f) < u(f) + 1/n$. Similarly, we have $l(f) - 1/n < l(f)$, the latter of which is the supremum of $\mathbb{L}$. Hence, $l(f) - 1/n$ is **not** an upper bound on $\mathbb{L}$. This means there exists $\mathscr{D}''_n \in \mathbb{D}$ such that $l_{\mathscr{D}''_n}(f) > l(f) - 1/n$. We now set $\mathscr{D}_n := \mathscr{D}'_n \cup \mathscr{D}''_n$, which is a refinement of both $\mathscr{D}'_n$ and $\mathscr{D}''_n$. Consequently, we have

$$u(f) \leq u_{\mathscr{D}_n}(f) \leq u_{\mathscr{D}'_n}(f) < u(f) + \frac{1}{n}$$

and

$$l(f) - \frac{1}{n} < l_{\mathscr{D}'_n}(f) \leq l_{\mathscr{D}_n}(f) \leq l(f).$$

by the Refinement Lemma. The Squeeze Rule applies: $u_{\mathscr{D}_n}(f) \to u(f)$ and $l_{\mathscr{D}_n}(f) \to l(f)$. So $u_{\mathscr{D}_n}(f) - l_{\mathscr{D}_n}(f) \to u(f) - l(f) = 0$ by the Algebra of Limits. Further, the Riemann integral is

$$\int_a^b f = \lim_{n\to\infty} l_{\mathscr{D}_n}(f) = \lim_{n\to\infty} u_{\mathscr{D}_n}(f).$$

($\Leftarrow$) Assume that a sequence $(\mathscr{D}_n)$ exists such that $u_{\mathscr{D}_n}(f) - l_{\mathscr{D}_n}(f) \to 0$. But for all $n \in \mathbb{Z}^+$, we know that $u_{\mathscr{D}_n}(f) \geq u(f)$ and $l_{\mathscr{D}_n}(f) \leq l(f)$ by definition. Hence, $u_{\mathscr{D}_n}(f) - l_{\mathscr{D}_n}(f) \geq u(f) - l(f)$. By Proposition 1.7, we can apply the limit and preserve the inequality:

$$0 = \lim_{n\to\infty} \left(u_{\mathscr{D}_n}(f) - l_{\mathscr{D}_n}(f)\right) \geq u(f) - l(f).$$

In other words, $l(f) \geq u(f)$. Since $l(f) \leq u(f)$ by Lemma 4.18, we see that $l(f) = u(f)$ and so f is Riemann integrable. To find the integral explicitly, notice that

$$0 \leq u_{\mathscr{D}_n}(f) - u(f) \leq u_{\mathscr{D}_n}(f) - u(f) + l(f) - l_{\mathscr{D}_n}(f) = u_{\mathscr{D}_n}(f) - l_{\mathscr{D}_n}(f).$$

The Squeeze Rule now shows that $u_{\mathscr{D}_n}(f) - u(f) \to 0$, that is $u_{\mathscr{D}_n}(f) \to u(f)$ as needed. □

Theorem 4.22 *Let $f : [a, b] \to \mathbb{R}$ be increasing. Then, f is Riemann integrable.*

Proof: Since f is increasing, it is bounded below by $f(a)$ and above $f(b)$; we know that $l(f)$ and $u(f)$ exist. For each $n \in \mathbb{Z}^+$, let $\mathscr{D}_n$ be the regular dissection of $[a, b]$ of size n, meaning each subinterval has width $(b-a)/n$. Then, for each $j = 1, ..., n$, the numbers introduced in (†) are

$$m_j = \inf\{f(x) : x \in [a_{j-1}, a_j]\} = f(a_{j-1}) \qquad \text{and} \qquad M_j = \sup\{f(x) : x \in [a_{j-1}, a_j]\} = f(a_j).$$

Using these in the definition of the upper and lower Riemann sums, we obtain

$$\begin{aligned} u_{\mathscr{D}_n}(f) - l_{\mathscr{D}_n}(f) &= \frac{b-a}{n} \sum_{j=1}^{n} \big(f(a_j) - f(a_{j-1})\big) \\ &= \frac{b-a}{n} \big(f(b) - f(a)\big) \\ &\to 0, \end{aligned}$$

to which we can apply Theorem 4.21 and conclude that f is Riemann integrable. □

Theorem *Let $f : [a, b] \to \mathbb{R}$ be decreasing. Then, f is Riemann integrable.*

Proof: Since f is decreasing, it is bounded below by $f(b)$ and above $f(a)$; we know that $l(f)$ and $u(f)$ exist. For each $n \in \mathbb{Z}^+$, let $\mathscr{D}_n$ be the regular dissection of $[a, b]$ of size n, meaning each subinterval has width $(b-a)/n$. Then, for each $j = 1, ..., n$, the numbers introduced in (†) are

$$m_j = \inf\{f(x) : x \in [a_{j-1}, a_j]\} = f(a_j) \qquad \text{and} \qquad M_j = \sup\{f(x) : x \in [a_{j-1}, a_j]\} = f(a_{j-1}).$$

Using these in the definition of the upper and lower Riemann sums, we obtain

$$\begin{aligned} u_{\mathscr{D}_n}(f) - l_{\mathscr{D}_n}(f) &= \frac{b-a}{n} \sum_{j=1}^{n} \big(f(a_{j-1}) - f(a_j)\big) \\ &= \frac{b-a}{n} \big(f(a) - f(b)\big) \\ &\to 0, \end{aligned}$$

to which we can apply Theorem 4.21 and conclude that f is Riemann integrable. □

Theorem 4.26 *Let $f : [a, b] \to \mathbb{R}$ be continuous. Then, f is Riemann integrable.*

Proof: By the Extreme Value Theorem, we know f is bonded and so $l(f)$ and $u(f)$ exist. Consider now the sequence $(\mathscr{D}_n)$ of regular dissections of $[a, b]$ of size 2^n. Since $\mathscr{D}_n \subseteq \mathscr{D}_{n+1}$, the Refinement Lemma tells us that $(l_{\mathscr{D}_n}(f))$ is an increasing sequence and $(u_{\mathscr{D}_n}(f))$ is a decreasing sequence. Hence, the Monotone Convergence Theorem tells us that $l_{\mathscr{D}_n}(f) \to K$ and $u_{\mathscr{D}_n}(f) \to L$ for some $K, L \in \mathbb{R}$. The Algebra of Limits and Proposition 1.7 now inform us that

$$u_{\mathscr{D}_n}(f) - l_{\mathscr{D}_n}(f) \to K - L \geq 0.$$

In order to show integrability via Theorem 4.21, we must prove that $K - L = 0$. To this end, assume for a contradiction that $\varepsilon \coloneqq K - L > 0$. As usual, we use the notation m_j and M_j established in (†). Then, the difference between the upper and lower Riemann sums is

$$u_{\mathscr{D}_n}(f) - l_{\mathscr{D}_n}(f) = \frac{b-a}{2^n} \sum_{j=1}^{2^n} (M_j - m_j) \geq \varepsilon.$$

The sum consists of 2^n non-negative terms, so at least one of them must be greater than or equal to $\varepsilon/2^n$ (if not, the whole sum would be less than ε). As such, there exists $j \in \{1, ..., 2^n\}$ with

$$M_j - m_j \geq \frac{\varepsilon}{b-a}.$$

Note that $f : [a_{j-1}, a_j] \to \mathbb{R}$ is continuous; the Extreme Value Theorem once again applies, telling us that f attains a maximum and minimum on the subinterval $[a_{j-1}, a_j]$. In other words, there exist $x_n, y_n \in [a_{j-1}, a_j]$ such that $f(x_n) = M_j$ and $f(y_n) = m_j$. It follows that for each $n \in \mathbb{Z}^+$, there exist points $x_n, y_n \in [a, b]$ satisfying

$$|x_n - y_n| \leq \frac{b-a}{2^n}, \tag{\textbf{4.1}}$$

since they lie in an interval of this width, and

$$f(x_n) - f(y_n) \geq \frac{\varepsilon}{b-a}. \tag{\textbf{4.2}}$$

The sequence (x_n) is bounded, so there exists a convergent subsequence $x_{n_k} \to c \in [a, b]$ as a result of the Bolzano-Weierstrass Theorem. By (**4.1**), we know that

$$x_{n_k} - \frac{b-a}{2^{n_k}} \leq y_{n_k} \leq x_{n_k} + \frac{b-a}{2^{n_k}}.$$

The Squeeze Rule implies $y_{n_k} \to c$ also. Because f is continuous, we know that $f(x_{n_k}) \to f(c)$ and $f(y_{n_k}) \to f(c)$. The Algebra of Limits now tells us that $f(x_{n_k}) - f(y_{n_k}) \to f(c) - f(c) = 0$; this is a contradiction to both (**4.2**) and Proposition 1.7. □

Proposition *The following function $f : [0, 1] \to \mathbb{R}$ is* **not** *Riemann integrable:*

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

Proof: Let $\mathscr{D}$ be any dissection of $[0, 1]$. Then, every subinterval $[a_{j-1}, a_j]$ contains both rational and irrational numbers; the punchline to this is that $m_j = 0$ and $M_j = 1$ for all j. Consequently,

$$l_{\mathscr{D}}(f) = \sum_{j=1}^{n} 0(a_j - a_{j-1}) = 0 \qquad \text{and} \qquad u_{\mathscr{D}}(f) = \sum_{j=1}^{n} 1(a_j - a_{j-1}) = a_n - a_0 = 1.$$

Because this is true for **any** dissection, we see that $l(f) = \sup\{0\} = 0 \neq 1 = \inf\{1\} = u(f)$. □

4.5 Elementary Properties of the Riemann Integral

Proposition 4.27 *Let f be Riemann integrable on $[a,b]$ and on $[b,c]$. Then, f is Riemann integrable on $[a,c]$, and the Riemann integral is given by*

$$\int_a^c f = \int_a^b f + \int_b^c f.$$

Proof: By Theorem 4.21, there exist sequences of dissections $(\mathscr{D}_n')$ of $[a,b]$ and $(\mathscr{D}_n'')$ of $[b,c]$ with

$$l_{\mathscr{D}_n'}(f) \to \int_a^b f, \qquad u_{\mathscr{D}_n'}(f) \to \int_a^b f, \qquad l_{\mathscr{D}_n''}(f) \to \int_b^c f, \qquad u_{\mathscr{D}_n''}(f) \to \int_b^c f.$$

Let $\mathscr{D}_n = \mathscr{D}_n' \cup \mathscr{D}_n''$, which is a dissection of $[a,c]$. By Definition 4.11 and the Algebra of Limits,

$$l_{\mathscr{D}_n}(f) = l_{\mathscr{D}_n'}(f) + l_{\mathscr{D}_n''}(f) \to \int_a^b f + \int_b^c f,$$
$$u_{\mathscr{D}_n}(f) = u_{\mathscr{D}_n'}(f) + u_{\mathscr{D}_n''}(f) \to \int_a^b f + \int_b^c f.$$

Hence, the claim is immediate from Theorem 4.21; f is Riemann integrable on $[a,c]$. □

Proposition 4.28 *Let $f, g : [a,b] \to \mathbb{R}$ be Riemann integrable such that $f(x) \leq g(x)$ for all $x \in [a,b]$. Then, their integrals satisfy the inequality*

$$\int_a^b f \leq \int_a^b g.$$

Proof: By Theorem 4.21, there exist sequences of dissections $(\mathscr{D}_n')$ and $(\mathscr{D}_n'')$ of $[a,b]$ where

$$l_{\mathscr{D}_n'}(f) \to \int_a^b f \qquad \text{and} \qquad l_{\mathscr{D}_n''}(f) \to \int_a^b g.$$

Let $\mathscr{D}_n = \mathscr{D}_n' \cup \mathscr{D}_n''$, which is a dissection of $[a,b]$ and a refinement of each of the dissections given above. By applying the Refinement Lemma and the Squeeze Rule, it follows that

$$l_{\mathscr{D}_n}(f) \to \int_a^b f \qquad \text{and} \qquad l_{\mathscr{D}_n}(f) \to \int_a^b g.$$

But since $f(x) \leq g(x)$ for all $x \in [a,b]$, it follows that $\inf(f) \leq \inf(g)$ on any subset of $[a,b]$. Hence, $l_{\mathscr{D}_n}(f) \leq l_{\mathscr{D}_n}(g)$, or rather $0 \leq l_{\mathscr{D}_n}(g) - l_{\mathscr{D}_n}(f)$, for all $n \in \mathbb{Z}^+$. But now, $0 \leq \int_a^b g - \int_a^b f$ by taking the limit and using Proposition 1.7. Of course, this rearranges to what we require. □

Note: Assuming that f and g are continuous, we can replace $\leq$ by $<$ in Proposition 4.28.

Proposition 4.29 (Integral Triangle Inequality) *Let $f : [a, b] \to \mathbb{R}$ be Riemann integrable. Then, $|f| : [a, b] \to \mathbb{R}$ is Riemann integrable and its integral satisfies the inequality*

$$\left| \int_a^b f \right| \leq \int_a^b |f|.$$

Proof: Firstly, **if** $|f|$ is Riemann integrable, then since $-\big|f(x)\big| \leq f(x) \leq \big|f(x)\big|$ for all $x \in [a, b]$, it follows immediately from Proposition 4.28 that

$$-\int_a^b |f| \leq \int_a^b f \leq \int_a^b |f|.$$

This is equivalent to the inequality in the statement, so all that remains is to show Riemann integrability of $|f|$. To that end, first note that for any bounded function $g : [a, b] \to \mathbb{R}$ and a fixed dissection $\mathscr{D} = \{a_0, a_1, ..., a_k\}$ of $[a, b]$ of any size k, we can use Proposition 4.7 to see that

$$\begin{aligned}
0 &\leq u_{\mathscr{D}}(g) - l_{\mathscr{D}}(g) \\
&= \sum_{j=1}^{k} \big(\sup\{g(x) : x \in [a_{j-1}, a_j]\} - \inf\{g(x) : x \in [a_{j-1}, a_j]\}\big)\,(a_j - a_{j-1}) \\
&= \sum_{j=1}^{k} \operatorname{diam}(\{g(x) : x \in [a_{j-1}, a_j]\})(a_j - a_{j-1}).
\end{aligned}$$

Consequently, in the case $g = |f|$, the above argument can be slightly extended:

$$\begin{aligned}
0 &\leq u_{\mathscr{D}}(|f|) - l_{\mathscr{D}}(|f|) \\
&= \sum_{j=1}^{k} \operatorname{diam}(\{\big|f(x)\big| : x \in [a_{j-1}, a_j]\})(a_j - a_{j-1}) \\
&\leq \sum_{j=1}^{k} \operatorname{diam}(\{f(x) : x \in [a_{j-1}, a_j]\})(a_j - a_{j-1}) \\
&= u_{\mathscr{D}}(f) - l_{\mathscr{D}}(f),
\end{aligned}$$

where the second inequality above comes from Lemma 4.8. But this is true for **any** dissection $\mathscr{D}$. By Theorem 4.21, there exists a sequence of dissections $(\mathscr{D}_n)$ such that $u_{\mathscr{D}_n}(f) - l_{\mathscr{D}_n}(f) \to 0$. However, the estimation done above, in the case of the dissection $\mathscr{D}_n$, really tells us that

$$0 \leq u_{\mathscr{D}_n}(|f|) - l_{\mathscr{D}_n}(|f|) \leq u_{\mathscr{D}_n}(f) - l_{\mathscr{D}_n}(f).$$

The Squeeze Rule implies $u_{\mathscr{D}_n}(|f|) - l_{\mathscr{D}_n}(|f|) \to 0$, that is $|f|$ is Riemann integrable. □

Lemma 4.30 *Let $f, g : [a, b] \to \mathbb{R}$ be bounded and $\mathscr{D}$ any dissection of $[a, b]$. Then,*

$$l_{\mathscr{D}}(f + g) \geq l_{\mathscr{D}}(f) + l_{\mathscr{D}}(g) \qquad \text{and} \qquad u_{\mathscr{D}}(f + g) \leq u_{\mathscr{D}}(f) + u_{\mathscr{D}}(g).$$

Proof: We shall prove the first inequality (the second is done via an analogous argument). Suppose first that $\mathscr{D} = \{a_0, a_1, ..., a_n\}$. For any bounded function $h : [a, b] \to \mathbb{R}$, we use notation similar to that of (†), except we keep track of which function we are referring to here:

$$m_j(h) = \inf\{h(x) : x \in [a_{j-1}, a_j]\}.$$

For all $x \in [a_{j-1}, a_j]$, we have $f(x) \geq m_j(f)$ and $g(x) \geq m_j(g)$ by definition of the infimum, from which it follows that $f(x) + g(x) \geq m_j(f) + m_j(g)$. The number on the right-hand side is a lower bound on the set $\{f(x) + g(x) : x \in [a_{j-1}, a_j]\}$. Consequently, $m_j(f + g) \geq m_j(f) + m_j(g)$ since the number on the left-hand side is the **greatest** lower bound on the aforementioned set. Thus,

$$\begin{aligned} l_{\mathscr{D}}(f+g) &= \sum_{j=1}^{n} m_j(f+g)(a_j - a_{j-1}) \\ &\geq \sum_{j=1}^{n} \left(m_j(f) + m_j(g)\right)(a_j - a_{j-1}) \\ &= l_{\mathscr{D}}(f) + l_{\mathscr{D}}(g). \end{aligned}$$

□

Theorem 4.31 (Linearity of the Riemann Integral) *Let $f, g : [a, b] \to \mathbb{R}$ both be Riemann integrable, and $\alpha \in \mathbb{R}$ some constant. Then, the following are true:*

(i) *The function αf is Riemann integrable on $[a, b]$, with* $\displaystyle\int_a^b \alpha f = \alpha \int_a^b f$.

(ii) *The function $f + g$ is Riemann integrable on $[a, b]$, with* $\displaystyle\int_a^b (f+g) = \int_a^b f + \int_a^b g$.

Proof: (i) By Theorem 4.21, there exists a sequence of dissections $(\mathscr{D}_n)$ such that

$$l_{\mathscr{D}_n}(f) \to \int_a^b f \qquad \text{and} \qquad u_{\mathscr{D}_n}(f) \to \int_a^b f.$$

It is not too hard to see directly from Definition 4.11 that

$$l_{\mathscr{D}_n}(\alpha f) = \begin{cases} \alpha l_{\mathscr{D}_n}(f) & \text{if } \alpha \geq 0 \\ \alpha u_{\mathscr{D}_n}(f) & \text{if } \alpha < 0 \end{cases} \qquad \text{and} \qquad u_{\mathscr{D}_n}(\alpha f) = \begin{cases} \alpha u_{\mathscr{D}_n}(f) & \text{if } \alpha \geq 0 \\ \alpha l_{\mathscr{D}_n}(f) & \text{if } \alpha < 0 \end{cases}.$$

In either case, the Algebra of Limits implies that $l_{\mathscr{D}_n}(\alpha f) \to \alpha \int_a^b f$ and $u_{\mathscr{D}_n}(\alpha f) \to \alpha \int_a^b f$; the integrability and its expression follow from the sequential characterisation in Theorem 4.21.

(ii) By Theorem 4.21, there exist sequences of dissections $(\mathscr{D}'_n)$ and $(\mathscr{D}''_n)$ of $[a, b]$ such that

$$l_{\mathscr{D}'_n}(f) \to \int_a^b f, \qquad u_{\mathscr{D}'_n}(f) \to \int_a^b f, \qquad l_{\mathscr{D}''_n}(g) \to \int_a^b g, \qquad u_{\mathscr{D}''_n}(g) \to \int_a^b g.$$

Let $\mathscr{D}_n = \mathscr{D}'_n \cup \mathscr{D}''_n$, a refinement of the two dissections above. The Refinement Lemma implies

$$\begin{aligned} l_{\mathscr{D}'_n}(f) \leq l_{\mathscr{D}_n}(f) \leq \int_a^b f, &\qquad \int_a^b f \leq u_{\mathscr{D}_n}(f) \leq u_{\mathscr{D}'_n}(f), \\ l_{\mathscr{D}''_n}(g) \leq l_{\mathscr{D}_n}(g) \leq \int_a^b g, &\qquad \int_a^b g \leq u_{\mathscr{D}_n}(g) \leq u_{\mathscr{D}''_n}(g). \end{aligned}$$

We can now apply the Squeeze Rule to each of the above inequalities. Doing so yields

$$l_{\mathscr{D}_n}(f) \to \int_a^b f, \qquad u_{\mathscr{D}_n}(f) \to \int_a^b f, \qquad l_{\mathscr{D}_n}(g) \to \int_a^b g, \qquad u_{\mathscr{D}_n}(g) \to \int_a^b g.$$

By Lemma 4.30, we see also that

$$l_{\mathscr{D}_n}(f) + l_{\mathscr{D}_n}(g) \leq l_{\mathscr{D}_n}(f+g) \leq l(f+g)$$

and

$$u(f+g) \leq u_{\mathscr{D}_n}(f+g) \leq u_{\mathscr{D}_n}(f) + u_{\mathscr{D}_n}(g).$$

But the far-left and far-right (respectively) both converge to $\int_a^b f + \int_a^b g$, so applying the limit to the above inequalities and using Proposition 1.7 produces for us

$$\int_a^b f + \int_a^b g \leq l(f+g) \qquad \text{and} \qquad u(f+g) \leq \int_a^b f + \int_a^b g.$$

Combining these gives us $u(f+g) \leq l(f+g)$, but we know from Lemma 4.18 that the reverse inequality is always true. Hence, we have straight-up equality: $l(f+g) = u(f+g)$, telling us that $f+g$ is Riemann integrable. Furthermore, notice that

$$\int_a^b (f+g) = l(f+g) \geq \int_a^b f + \int_a^b g$$

and

$$\int_a^b (f+g) = u(f+g) \leq \int_a^b f + \int_a^b g,$$

which tells us that the integral is indeed given by $\int_a^b (f+g) = \int_a^b f + \int_a^b g$. $\square$

Note: The linearity in Theorem 4.31 establishes something deeper, namely that the set $L([a,b])$ of Riemann integrable functions on $[a,b]$ is an $\mathbb{R}$-**vector space**. Indeed, it is closed under the obvious addition and scalar multiplication operations. Moreover, it shows that this map, associating to each function its integral, is a linear map between vector spaces:

$$L([a,b]) \to \mathbb{R}, \qquad f \mapsto \int_a^b f.$$

5 The Fundamental Theorem of Calculus

5.1 The First Form and the Intermediate Value Theorem

Reminder: For relevant $f : [a, b] \to \mathbb{R}$, recall $\int_a^b f \in \mathbb{R}$ is a single number, **not** a function.

That said, if we fix the left-hand endpoint a of the interval and allow the right-hand endpoint b to vary, we can consider $\int_a^b f$ as a function of b. This idea is crucial for the next theorem.

Theorem 5.1 (Fundamental Theorem of Calculus Version 1) *Let $I \subseteq \mathbb{R}$ be an interval and $f : I \to \mathbb{R}$ be continuous. For any fixed $a \in I$, define the function $F : I \to \mathbb{R}$ by*

$$F(x) = \int_a^x f.$$

Then, F is differentiable with derivative $F' = f$.

Proof: First, f is continuous on $[a, x]$ (if $x \geq a$) or $[x, a]$ (if $x < a$) for all $x \in I$. This establishes F is well-defined as continuous functions are integrable (Theorem 4.26). We aim to determine

$$\lim_{y \to x} \frac{F(y) - F(x)}{y - x},$$

which we will do by using the sequential characterisation of limits (Theorem 1.35). To that end, let (y_n) be a sequence in $I \setminus \{x\}$ such that $y_n \to x$, and

$$s(y_n) \coloneqq \frac{F(y_n) - F(x)}{y_n - x}.$$

We must show $s(y_n) \to f(x)$. For each $n \in \mathbb{Z}^+$, note that either $y_n > x$ or $y_n < x$. This gives us two options for $s(y_n)$, which can be dealt with by simplifying $F(y_n) - F(x)$ via Proposition 4.27:

$$s(y_n) = \begin{cases} \dfrac{1}{y_n - x} \displaystyle\int_x^{y_n} f & \text{if } y_n > x \\ \dfrac{1}{x - y_n} \displaystyle\int_{y_n}^{x} f & \text{if } y_n < x \end{cases}.$$

In either case, the Extreme Value Theorem guarantees the existence of w_n and z_n between x and y_n such that $f(w_n)$ is the minimum value on that closed interval, and $f(z_n)$ is the maximum value on the same closed interval. Therefore, Proposition 4.13 gives us the following:

$$\begin{aligned} y_n > x \quad &\Rightarrow \quad \frac{1}{y_n - x} f(w_n)(y_n - x) \leq s(y_n) \leq \frac{1}{y_n - x} f(z_n)(y_n - x), \\ y_n < x \quad &\Rightarrow \quad \frac{1}{x - y_n} f(w_n)(x - y_n) \leq s(y_n) \leq \frac{1}{x - y_n} f(z_n)(x - y_n). \end{aligned}$$

Either way, we conclude that $f(w_n) \leq s(y_n) \leq f(z_n)$. Finally, since $y_n \to x$, the Squeeze Rule implies also that $w_n \to x$ and $z_n \to x$. But the continuity of f implies that $f(w_n) \to f(x)$ and $f(z_n) \to f(x)$. The Squeeze Rule applies to the above inequality, meaning $s(y_n) \to f(x)$. $\square$

Note: The Fundamental Theorem of Calculus Version 1 says this: any function that is continuous on an interval is the derivative of some differentiable function on said interval.

We can use the Fundamental Theorem of Calculus to give a remarkable proof of Theorem 1.47.

Theorem 1.47 (Intermediate Value Theorem) *Let $f : [a, b] \to \mathbb{R}$ be continuous and y be a number between $f(a)$ and $f(b)$. Then, there exists $c \in [a, b]$ such that $f(c) = y$.*

Proof: Let $F : [a, b] \to \mathbb{R}$ be given by $F(x) = \int_a^x f$. Then, F is differentiable with $F' = f$ by Theorem 5.1. In these terms, y is a number between $F'(a) = f(a)$ and $F'(b) = f(b)$. Hence, there exists $c \in [a, b]$ such that $F'(c) = y$, by Darboux's Theorem, but this is precisely $f(c) = y$. $\square$

Corollary 5.2 *Let $I \subseteq \mathbb{R}$ be an interval and $f : I \to \mathbb{R}$ be continuous. For any fixed $b \in I$, define the function $F : I \to \mathbb{R}$ by*

$$F(x) = \int_x^b f.$$

Then, F is differentiable with derivative $F' = -f$.

Proof: Let $a \in I$ be arbitrary and define $G : I \to \mathbb{R}$ by $G(x) = \int_a^x f$. By Proposition 4.27,

$$G(x) + F(x) = \int_a^x f + \int_x^b f = \int_a^b f \equiv K \in \mathbb{R}$$

some constant. But G is differentiable with $G' = f$ by Theorem 5.1. Hence, the above can be rearranged to $F(x) = K - G(x)$; this is differentiable by linearity, with $F' = 0 - G' = -f$. $\square$

Method – Differentiating an Integral: Suppose we are given an integral $g(x) = \int_{u(x)}^{v(x)} f$. We can differentiate this by using the Fundamental Theorem of Calculus Version 1, namely

$$g'(x) = v'(x)f(v(x)) - u'(x)f(u(x)).$$

5.2 The Second Form and a Practical Method for Computing Integrals

The next version of the Fundamental Theorem of Calculus is one we've used implicitly for years.

Theorem 5.4 (Fundamental Theorem of Calculus Version 2) *Let $f : [a, b] \to \mathbb{R}$ be continuous and $F : [a, b] \to \mathbb{R}$ be* **any** *differentiable function such that $F' = f$. Then,*

$$\int_a^b = F(b) - F(a).$$

Proof: Define the function $g : [a, b] \to \mathbb{R}$ by $g(x) = \left(\int_a^x f\right) - F(x)$. By Theorem 5.1 and the definition of F, we see that $g'(x) = f(x) - f(x) = 0$ for all $x \in [a, b]$. Therefore, we know g is constant by Proposition 3.11(iii). In particular, $g(b) = g(a)$ and this is equivalent to saying

$$\left(\int_a^b f\right) - F(b) = \left(\int_a^a f\right) - F(a) = 0 - F(a) \qquad \Leftrightarrow \qquad \int_a^b f = F(b) - F(a). \qquad \square$$

Reminder: Let $f : D \to \mathbb{R}$ be a function. We say f is odd if $f(-x) = -f(x)$ for all $x \in D$.

Proposition 5.6 *Let $f : [-a, a] \to \mathbb{R}$ be a continuous odd function. Then,*

$$\int_{-a}^a f = 0.$$

Proof: Let's define $F : [0, a] \to \mathbb{R}$ by $F(x) = \int_{-x}^x f$. Using Proposition 4.27, we can write

$$F(x) = \int_{-x}^0 f + \int_0^x f.$$

Then, the Chain Rule along with Theorem 5.1 and Corollary 5.2 gives us its derivative:

$$F'(x) = -(-f(-x)) + f(x) = f(-x) + f(x) = -f(x) + f(x) = 0,$$

using the oddness of f. Hence, we know F is constant by Proposition 3.11(iii). In particular,

$$\int_{-a}^a f = F(a) = F(0) = 0. \qquad \square$$

Remark The best part about Theorem 5.4 is that we can explicitly compute integrals of f just by conjuring up a so-called *anti-derivative* of f, that is **any** function whose derivative is f. It is common to call an anti-derivative of f an *indefinite integral*, denoted somewhat ambiguously by $\int f(x)\,\mathrm{d}x$. Really, this represents a *class* of functions (recall the involvement of "$+C$" when you have worked with indefinite integrals before) who all differentiate to f. This **is** a function of x, **not** a single number; its definition has **nothing** to do with Riemann sums. But the relation between $\int f$ and $\int_a^b f$ is exactly provided by the Fundamental Theorem of Calculus Version 2.

5.3 The Natural Logarithm

Definition 5.8 The (natural) logarithm is the function $\ln : (0, \infty) \to \mathbb{R}$ defined by

$$\ln(x) = \int_1^x \frac{1}{t}\,\mathrm{d}t.$$

Remark The integrand $f(t) = 1/t$ is continuous on $(0, \infty)$ and hence Riemann integrable on $[1, x]$ (if $x \geq 1$) or $[x, 1]$ (if $0 < x < 1$). Thus, we know that the natural logarithm is well-defined. Moreover, the Fundamental Theorem of Calculus Version 1 implies it is differentiable with

$$\ln'(x) = \frac{1}{x}.$$

Proposition 5.9 *For all $x, y \in (0, \infty)$, we have $\ln(xy) = \ln(x) + \ln(y)$.*

Proof: Fix $y \in (0, \infty)$ and define the function $f : (0, \infty) \to \mathbb{R}$ by $f(x) = \ln(xy) - \ln(x) - \ln(y)$. By applying the Chain Rule and linearity, we know f is differentiable with derivative

$$f'(x) = \frac{y}{xy} - \frac{1}{x} - 0.$$

Hence, f is constant by Proposition 3.11(iii). In particular, $f(x) = f(1) = \ln(y) - 0 - \ln(y) = 0$ for all $x \in (0, \infty)$, so the definition of f rearranges to precisely the identity we wished to prove. $\square$

Proposition 5.10 *For all $x \in (0, \infty)$ and $n \in \mathbb{Z}$, we have $\ln(x^n) = n \ln(x)$.*

Proof: We proceed by induction, starting with the base case $n = 0$ which is certainly true: $\ln(x^0) = \ln(1) = \int_1^1 1/t \, \mathrm{d}t = 0 = 0 \ln(x)$. Moreover, Proposition 5.9 in the case $y = 1/x$ tells us

$$\ln\left(x \cdot \frac{1}{x}\right) = \ln(x) + \ln\left(\frac{1}{x}\right),$$

But the left-hand side is zero which means the above becomes $\ln\bigl(1/x\bigr) = -\ln(x)$, which is precisely the statement when $n = -1$. Let's assume the claim holds for some $n = k \in \mathbb{Z}$. Then, applying Proposition 5.9 and the inductive hypothesis gives us

$$\begin{aligned}
\ln(x^{k+1}) &= \ln(x^k x) \\
&= \ln(x^k) + \ln(x) \\
&= k \ln(x) + \ln(x) \\
&= (k + 1) \ln(x)
\end{aligned}$$

and

$$\begin{aligned}
\ln(x^{k-1}) &= \ln(x^k(1/x)) \\
&= \ln(x^k) + \ln\bigl(1/x\bigr) \\
&= k \ln(x) - \ln(x) \\
&= (k - 1) \ln(x).
\end{aligned}$$

Therefore, by induction, the formula holds for all $n \in \mathbb{Z}$. $\square$

Proposition 5.11 *The function* $\ln : (0, \infty) \to \mathbb{R}$ *is smooth, strictly increasing and bijective.*

Proof: Let $f : (0, \infty) \to \mathbb{R}$ be given by $f(x) = 1/x$. We know ln is differentiable with derivative f. But f is smooth, which means ln is smooth. Also, for all $x \in (0, \infty)$, we have $f(x) > 0$. By Proposition 3.11(iv), ln is strictly increasing and hence injective. Next, for $n \in \mathbb{Z}^+$ with $n \geq 2$, let $\mathscr{D}_n = \{1, 2, ..., n\}$ be the regular dissection of $[1, n]$ of size $n - 1$. Then, we see that

$$\begin{aligned}
\ln(n) &= \int_1^n f \\
&\geq l_{\mathscr{D}_n}(f) \\
&= \sum_{j=1}^{n-1} f(j+1)((j+1)-1) \\
&= \sum_{j=1}^{n-1} \frac{1}{j+1} \\
&= \sum_{k=2}^{n} \frac{1}{k} \\
&=: s_n,
\end{aligned}$$

changing the summation variable for convenience. Note that the sequence (s_n) is unbounded above because it is a partial sum of the geometric series. Therefore, the sequence $(\ln n)$ is also unbounded above. Hence, for any $K \geq 0$, there exists $n \in \mathbb{Z}^+$ with $\ln(n) > K$. But $\ln(1) = 0$ and ln is continuous because it is differentiable (Proposition 2.8). We can thus apply the Intermediate Value Theorem: there exists $x \in [1, n]$ such that $\ln(x) = K$. As such, ln takes all non-negative values. We now let $L \leq 0$; we have just showed that there exists $x \in [1, \infty)$ such that $\ln(x) = -L$. But notice that $x \in [1, \infty)$ implies $1/x \in (0, 1]$ and, by Proposition 5.10, we have

$$L = -\ln(x) = \ln\big(1/x\big).$$

As such, ln also takes all negative values; this concludes surjectivity and thus bijectivity. $\square$

Note: It is common to define the natural logarithm as the inverse of the exponential function. Note that Definition 5.8 has no obvious connection with the exponential function, so this coincidence of terminology must be justified (and will indeed be, a bit later).

6 Uniform Convergence

6.1 Swapping Limits

There are occasions where we must compute a *double limit*, e.g. define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \lim_{k\to\infty} \sum_{n=0}^{k} \frac{x^n}{n!}$$

and compute f' (assuming it exists; it does since this is the exponential function). So, at $a \in \mathbb{R}$,

$$f'(a) = \lim_{x\to a} \frac{f(x) - f(a)}{x - a} = \lim_{x\to a} \lim_{k\to\infty} \frac{1}{x-a} \sum_{n=0}^{k} \frac{x^n - a^n}{n!}.$$

Note: The major problem is that we generally **can't** interchange two different limits.

6.2 Pointwise Convergence vs Uniform Convergence

Definition 6.4 A sequence of functions $f_n : D \to \mathbb{R}$ converges pointwise to $f : D \to \mathbb{R}$ if, for each fixed $x \in D$, the real sequence $(f_n(x))$ converges to $f(x)$ in the usual sense.

Lemma *This sequence converges pointwise to $f : [0,1] \to \mathbb{R}$ given by $f(x) = 0$:*

$$f_n : [0,1] \to \mathbb{R}, \qquad f_n(x) = \begin{cases} 2n^2 x & \text{if } 0 \le x \le \frac{1}{2n} \\ 2n - 2n^2 x & \text{if } \frac{1}{2n} \le x \le \frac{1}{n} \\ 0 & \text{if } \frac{1}{n} \le x \le 1 \end{cases}.$$

Proof: Fix $x \in (0,1]$ and notice that, for all $n > 1/x$ we have $x > 1/n$ and so $f_n(x) = 0$ by the third case in its definition above. Hence, for each $x \in (0,1]$, the real sequence $f_n(x) \to 0$. Furthermore, $f_n(0) = 2n^2 \cdot 0 = 0$ for all $n \in \mathbb{Z}^+$. Hence, $f_n(x) \to 0$ for all $x \in [0,1]$. $\square$

Remark We can interpret f_n for fixed $n \in \mathbb{Z}^+$ geometrically: its graph is an isosceles triangle based at $x = 0$ to $x = 1/n$, followed by a horizontal line from $x = 1/n$ to $x = 1$ along the x-axis.

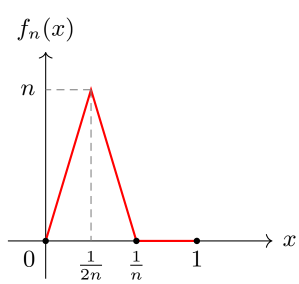


Figure 9: The geometric interpretation of $f_n : [0,1] \to \mathbb{R}$ from the above lemma.

Lemma *The sequence $f_n : [0,1] \to \mathbb{R}$ where $f_n(x) = x^n$ converges pointwise to*

$$f : [0,1] \to \mathbb{R} \qquad f(x) = \begin{cases} 0 & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x = 1 \end{cases}.$$

Proof: If $x \in [0,1)$, we have $f_n(x) = x^n \to 0$. However, if $x = 1$, we see that $f_n(1) = 1^n = 1 \to 1$ for all $n \in \mathbb{Z}^+$. Therefore, (f_n) converges pointwise to the discontinuous function on $[0,1]$ which outputs zero everywhere, except at $x = 1$ where it outputs one. This is precisely f as above. □

Remark We can interpret the sequence (f_n) geometrically: it consists of curves with endpoints $(0,0)$ and $(1,1)$ which, as n increases, get steeper and steeper. The idea is that in the pointwise limit $n \to \infty$, the curve will be flat to the x-axis except for the single point at $(1,1)$; this is the red curve (the graph of the discontinuous function f from above) in Figure 10 below.

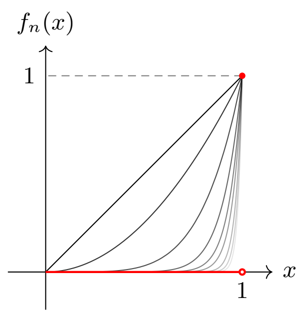


Figure 10: The geometric interpretation of $f_n : [0,1] \to \mathbb{R}$ given by $f_n(x) = x^n$.

Note: This shows that, even if each f_n is continuous, the pointwise limit f may **not** be.

Proposition 6.7 *If a sequence of functions (f_n) converges pointwise, then its limit is unique.*

Proof: Suppose to the contrary that (f_n) converges pointwise to f and $g \neq f$. Then, there exists $c \in D$ such that $f(c) \neq g(c)$. However, $f_n(c) \to f(c)$ and $f_n(c) \to g(c)$ by Definition 6.4. But this here is convergence of a real sequence, so it contradicts the usual Uniqueness of Limits (Proposition 1.4). □

Proposition 6.8 *Let (f_n) and (g_n) be sequences that, on the domain D, converge pointwise to f and g, respectively. Then, we have the following:*

(i) *$(f_n + g_n)$ converges pointwise to $f + g$ on D.*

(ii) *$(f_n g_n)$ converges pointwise to fg on D,*

Sketch of Proof: Just apply Definition 6.4 and the usual Algebra of Limits. □

Definition 6.9 The sup norm of a bounded function $f : D \to \mathbb{R}$ is defined as

$$\|f\| := \sup\{|f(x)| : x \in D\}.$$

Remark Suppose we have two continuous functions $f, g : [a, b] \to \mathbb{R}$. We can give a neat geometric interpretation to $\|f - g\|$. Indeed, this is the maximum separation between the points $(x, f(x))$ and $(x, g(x))$ as x varies in the interval $[a, b]$. This is a sort-of *distance* between f and g.

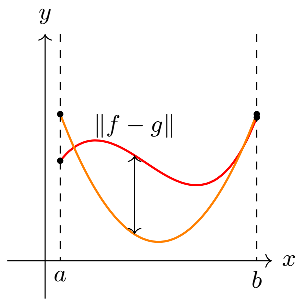


Figure 11: The geometric interpretation of $\|f - g\|$.

The geometric intuition behind the sup norm of a continuous bounded function is now clear: this is the special case where g is identically zero, so it is the maximum distance the graph $y = f(x)$ is away from the x-axis.

Definition 6.11 A sequence of functions $f_n : D \to \mathbb{R}$ converges uniformly to $f : D \to \mathbb{R}$ if the real sequence $\|f_n - f\| \to 0$ in the usual sense.

Lemma *The sequence* $f_n : [0, 1] \to \mathbb{R}$ *where* $f_n(x) = x^n$ **doesn't** *converge uniformly to*

$$f : [0, 1] \to \mathbb{R} \qquad f(x) = \begin{cases} 0 & \textit{if } 0 \le x < 1 \\ 1 & \textit{if } x = 1 \end{cases}.$$

Proof: For each $n \in \mathbb{Z}^+$, we see that

$$|f_n(x) - f(x)| = |x^n - f(x)| = \begin{cases} x^n & \text{if } 0 \le x < 1 \\ 0 & \text{if } x = 1 \end{cases}.$$

It is clear from the above calculation that

$$\{|f_n(x) - f(x)| : x \in [0, 1]\} = \{x^n : 0 \le x < 1\} \cup \{0\} = [0, 1).$$

Hence, $\|f_n - f\| = \sup[0, 1) = 1 \not\to 0$; the sequence (f_n) doesn't converge uniformly to f. □

Note: This establishes that pointwise convergence does **not** imply uniform convergence.

Theorem 6.12 *If a sequence $f_n : D \to \mathbb{R}$ converges uniformly to $f : D \to \mathbb{R}$, then it converges pointwise to f.*

Proof: Suppose that (f_n) converges uniformly to f. Then, for each fixed $x \in D$, we have

$$0 \leq \left|f_n(x) - f(x)\right| \leq \sup\{\left|f_n(y) - f(y)\right| : y \in D\} = \|f_n - f\| \to 0.$$

By the Squeeze Rule, we see that $\left|f_n(x) - f(x)\right| \to 0$. Hence, $f_n(x) \to f(x)$ for each $x \in D$. $\square$

Corollary 6.13 *If a sequence of functions (f_n) converges uniformly, then its limit is unique.*

Proof: Suppose (f_n) converges uniformly to f and g say. Then, f_n converges pointwise to each of f and g by Theorem 6.12, but pointwise limits are unique by Proposition 6.7. Thus, $f = g$. $\square$

Theorem 6.14 *Let $f_n : D \to \mathbb{R}$ be a sequence of continuous functions that converges uniformly to $f : D \to \mathbb{R}$. Then, f is itself continuous.*

Proof: We must show that f is continuous at a for every $a \in D$. To this end, we use Theorem 1.42 and show that for each $\varepsilon > 0$, there exists $\delta > 0$ such that, for all $x \in D$ with $|x - a| < \delta$, we have $\left|f(x) - f(a)\right| < \varepsilon$. Let $\varepsilon > 0$ be given. Since (f_n) converges uniformly to f, we have $\|f_n - f\| \to 0$, i.e. there exists $N \in \mathbb{Z}^+$ such that, for all $n \geq N$, $\|f_n - f\| < \varepsilon/3$. In particular,

$$\|f_N - f\| = \sup\{\left|f_N(x) - f(x)\right| : x \in D\} < \frac{\varepsilon}{3} \qquad \Rightarrow \qquad \left|f_N(x) - f(x)\right| < \frac{\varepsilon}{3} \text{ for each } x \in D.$$

Furthermore, f_N is continuous by assumption; Theorem 1.42 guarantees the existence of $\delta > 0$ such that, for all $x \in D$ with $|x - a| < \delta$, we have

$$\left|f_N(x) - f_N(a)\right| < \frac{\varepsilon}{3}.$$

Therefore, for all $x \in D$ with $|x - a| < \delta$, we have

$$\begin{aligned}
\left|f(x) - f(a)\right| &= \left|f(x) - f_N(x) + f_N(x) - f_N(a) + f_N(a) - f(a)\right| \\
&\leq \left|f_N(x) - f(x)\right| + \left|f_N(x) - f_N(a)\right| + \left|f_N(a) - f(a)\right| \\
&\leq \|f_N - f\| + \left|f_N(x) - f_N(a)\right| + \|f_N - f\| \\
&< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} \\
&= \varepsilon.
\end{aligned}$$

$\square$

Method – Showing Non-Uniform Convergence of Continuous Functions: If we want to show that a sequence of continuous functions (f_n) does **not** converge uniformly on its domain, assume to the contrary it does converge uniformly to some f. Then, (f_n) converges pointwise to f by Theorem 6.12. If we can find a discontinuous pointwise limit g, we know that $f = g$ by Proposition 6.7, a contradiction to Theorem 6.14.

6.3 Uniform Convergence and Calculus

Theorem 6.16 *Let $f_n : [a,b] \to \mathbb{R}$ be a sequence of continuous functions that converges uniformly to $f : [a,b] \to \mathbb{R}$. Then, we have*

$$\int_a^b f_n \to \int_a^b f.$$

Proof: Because f is continuous by Theorem 6.14, it is Riemann integrable by Theorem 4.26; this means that the limit $\int_a^b f$ certainly exists. Now, for each $n \in \mathbb{Z}^+$, we have

$$\begin{aligned}
0 &\leq \left| \int_a^b f_n - \int_a^b f \right| \\
&= \left| \int_a^b (f_n - f) \right|, && \text{by Theorem 4.31(ii)}, \\
&\leq \int_a^b |f_n - f|, && \text{by Proposition 4.29}, \\
&\leq (b-a)\|f_n - f\|, && \text{by Proposition 4.28}.
\end{aligned}$$

By the Squeeze Rule, we conclude that $\int_a^b f_n \to \int_a^b f$, as required. □

Lemma *On the domain $[0,1]$, we have*

$$\lim_{n\to\infty} \int_0^1 \frac{n}{n+x^n}\, \mathrm{d}x = 1$$

Proof: Let $f_n : [0,1] \to \mathbb{R}$ define the sequence of integrands, that is $f_n(x) = n/(n+x^n)$. We first consider the pointwise limit of this sequence of functions. Indeed, for each $x \in [0,1)$, we see that

$$f_n(x) = \frac{n}{n+x^n} = \frac{1}{1+x^n/n} \to \frac{1}{1+0} = 1$$

by the Algebra of Limits. On the other hand, when $x = 1$, we can also see that

$$f_n(1) = \frac{n}{n+1} = \frac{1}{1+1/n} \to \frac{1}{1+0} = 1$$

again by the Algebra of Limits. Thus, (f_n) converges pointwise to the function $f : [0,1] \to \mathbb{R}$ where $f(x) = 1$. It remains to show this is also the uniform limit. For fixed $x \in [0,1]$ and $n \in \mathbb{Z}^+$,

$$\left|f_n(x) - f(x)\right| = \left| \frac{n}{n+x^n} - 1 \right| = \left| \frac{x^n}{n+x^n} \right| \leq \frac{1}{n+x^n} \leq \frac{1}{n}.$$

Consequently, $0 \leq \|f_n - f\| = \sup\{\left|f_n(x) - f(x)\right| : x \in [0,1]\} \leq 1/n$. By the Squeeze Rule, it follows that $\|f_n - f\| \to 0$. And by Theorem 6.16, we get the result we are after:

$$\int_0^1 f_n \to \int_0^1 f = \int_0^1 1 = 1.$$

□

Theorem 6.19 *Let $f_n : [a,b] \to \mathbb{R}$ be a sequence of continuously differentiable functions that converges pointwise to $f : [a,b] \to \mathbb{R}$, where the sequence of derivatives $f_n' : [a,b] \to \mathbb{R}$ converges uniformly to $g : [a,b] \to \mathbb{R}$. Then, f is continuously differentiable with $f' = g$.*

Proof: For each $n \in \mathbb{Z}^+$, let $F_n : [a,b] \to \mathbb{R}$ be defined by

$$F_n(x) = \int_a^x f_n'.$$

This exists because each f_n' is continuous. By the Fundamental Theorem of Calculus Version 2,

$$F_n(x) = f_n(x) - f_n(a).$$

By Theorem 6.16, it is true for each fixed $x \in [a,b]$ that

$$f_n(x) - f_n(a) = F_n(x) \to \int_a^x g,$$

using the fact that (f_n') converges uniformly to g on the interval $[a,x]$. But (f_n) also converges pointwise to f, so $f_n(x) \to f(x)$ and $f_n(a) \to f(a)$. Therefore, the Algebra of Limits implies $f_n(x) - f_n(a) \to f(x) - f(a)$. But (pointwise) limits are unique, meaning

$$f(x) - f(a) = \int_a^x g.$$

By the Fundamental Theorem of Calculus Version 1, f is differentiable with $f' = g$. $\square$

Note: We view Theorem 6.19 as being the analogue of Theorem 6.16 but for derivatives.

6.4 Completeness of the Set of Bounded Functions

Reminder: A real sequence (a_n) is Cauchy if, for each $\varepsilon > 0$, there exists $N \in \mathbb{Z}^+$ where, for all $n, m \geq N$, we have $|a_n - a_m| < \varepsilon$. It is equivalent to convergence (Theorem 1.19).

Definition 6.20 A sequence of bounded functions $f_n : D \to \mathbb{R}$ is uniformly Cauchy if, for each $\varepsilon > 0$, there exists $N \in \mathbb{Z}^+$ such that, for all $n, m \geq N$, we have $\|f_n - f_m\| < \varepsilon$.

Lemma 6.23 *For all bounded functions $f, g : D \to \mathbb{R}$, we have $\|f + g\| \leq \|f\| + \|g\|$.*

Proof: For all $x \in D$, we have $\big|f(x) + g(x)\big| \leq \big|f(x)\big| + \big|g(x)\big| \leq \|f\| + \|g\|$ by the usual Triangle Inequality, and using the fact that $\|h\|$ is always an upper bound on the set $\{\big|h(x)\big| : x \in D\}$ (used in the cases where $h = f$ and $h = g$). Therefore, $\|f\| + \|g\|$ is an upper bound on the set $\{\big|f(x) + g(x)\big| : x \in D\}$. But by definition, $\|f + g\|$ is the supremum of this set (the **least** upper bound), which means $\|f + g\| \leq \|f\| + \|g\|$. $\square$

Theorem 6.22 *Let $f_n : D \to \mathbb{R}$ be a sequence of bounded functions that converges uniformly to the bounded function $f : D \to \mathbb{R}$. Then, (f_n) is uniformly Cauchy.*

Proof: Let $\varepsilon > 0$ be given. Since (f_n) converges uniformly to f, there exists $N \in \mathbb{Z}^+$ such that, for all $n \geq N$, we have $\|f_n - f\| < \varepsilon/2$. Hence, for all $n, m \geq N$, Lemma 6.23 implies that

$$\|f_n - f_m\| = \|f_n - f + f - f_m\| \leq \|f_n - f\| + \|f_m - f\| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \qquad \square$$

Lemma 6.24 *Let $f_n : D \to \mathbb{R}$ be a uniformly Cauchy sequence of bounded functions. Then, (f_n) converges pointwise to some function $f : D \to \mathbb{R}$.*

Proof: Fix $x \in D$; we must show that the real sequence $(f_n(x))$ is Cauchy. Indeed, let $\varepsilon > 0$ be given. Since (f_n) is uniformly Cauchy, there exists $N \in \mathbb{Z}^+$ such that, for all $n, m \geq N$, we have $\|f_n - f_m\| < \varepsilon$. Hence, for all $n, m \geq N$, it follows that

$$\big|f_n(x) - f_m(x)\big| \leq \|f_n - f_m\| < \varepsilon.$$

By Theorem 1.19, it follows that $f_n(x) \to f(x) \in \mathbb{R}$. But if we now let x vary in D, we obtain a function $f : D \to \mathbb{R}$ to which (f_n) converges pointwise. $\square$

Lemma 6.25 *Let $f_n : D \to \mathbb{R}$ be a uniformly Cauchy sequence of bounded functions that converges pointwise to $f : D \to \mathbb{R}$. Then, f is bounded.*

Proof: Because (f_n) is uniformly Cauchy, there exists $N \in \mathbb{Z}^+$ such that, for all $n, m \geq N$, we have $\|f_n - f_m\| < 1$ (we have chosen $\varepsilon = 1$ in the definition). Furthermore, $f_N : D \to \mathbb{R}$ is bounded, so there exists $K > 0$ such that $\big|f_N(x)\big| \leq K$ for all $x \in D$. But now, for all $n \geq N$ and fixed $x \in D$, we have

$$\begin{aligned}
\big|f(x)\big| &= \big|f(x) - f_n(x) + f_n(x) - f_N(x) + f_N(x)\big| \\
&\leq \big|f(x) - f_n(x)\big| + \big|f_n(x) - f_N(x)\big| + \big|f_N(x)\big| \\
&\leq \big|f(x) - f_n(x)\big| + \|f_n - f_N\| + K \\
&\leq \big|f(x) - f_n(x)\big| + 1 + K.
\end{aligned}$$

But we assume pointwise convergence, i.e. $f_n(x) \to f(x)$. Therefore, the sequence in the final inequality above converges to $K + 1$. By Proposition 1.7, we see that $\big|f(x)\big| \leq K + 1$; this is precisely the claim that f is bounded. $\square$

Theorem 6.26 *A sequence $f_n : D \to \mathbb{R}$ of bounded functions converges uniformly if and only if it is uniformly Cauchy.*

Proof: The "only if" is Theorem 6.22. Conversely, assume that (f_n) is uniformly Cauchy. Then, it converges pointwise (Lemma 6.24) to some bounded (Lemma 6.25) function $f : D \to \mathbb{R}$. It remains to prove that this convergence is uniform, i.e. $\|f_n - f\| \to 0$. To that end, let $\varepsilon 0$ be given. Since (f_n) is uniformly Cauchy, there exists $N \in \mathbb{Z}^+$ such that, for all $n, m \geq N$, we have $\|f_n - f_m\| < \varepsilon/4$. Next, for fixed $x \in D$ and $n, m \geq N$, we have

$$\begin{aligned} \big|f_n(x) - f(x)\big| &= \big|f_n(x) - f_m(x) + f_m(x) - f(x)\big| \\ &\leq \big|f_n(x) - f_m(x)\big| + \big|f_m(x) - f(x)\big| \\ &\leq \|f_n - f_m\| + \big|f_m(x) - f(x)\big| \\ &< \frac{\varepsilon}{4} + \big|f_m(x) - f(x)\big|. \end{aligned}$$

But (f_n) converges pointwise to f, meaning $f_n(x) \to f(x)$. Therefore, there exists $N_1 \in \mathbb{Z}^+$ such that, for all $m \geq N_1$, we have $\big|f_m(x) - f(x)\big| < \varepsilon/4$. If we set $m = \max\{N, N_1\}$, the above says

$$\big|f_n(x) - f(x)\big| < \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \frac{\varepsilon}{2}$$

for all $n \geq N$. This is true for every $x \in D$, and because N is independent of x, we see that

$$\|f_n - f\| = \sup\{\big|f_n(x) - f(x)\big| : x \in D\} \leq \frac{\varepsilon}{2} < \varepsilon$$

for all $n \geq N$. Since $\varepsilon > 0$ was arbitrary, this implies $\|f_n - f\| \to 0$. □

Reminder: The set of continuous functions on a domain $D \subseteq \mathbb{R}$ is denoted $C^0(D)$.

Definition The set of bounded functions on a domain $D \subseteq \mathbb{R}$ is denoted $B(D)$.

In general, neither $B(D)$ and $C^0(D)$ is a subset of the other. However, the Extreme Value Theorem tells us that, in the case our domain is a closed bounded interval $D = [a, b]$, we have

$$C^0([a, b]) \subseteq B([a, b]).$$

Note: Recall that *completeness* means the Cauchy property is equivalent to convergence. We have shown, in Theorem 6.26, the set $B(D)$ is complete with respect to the sup norm.

It turns out we can rather easily show that $C^0([a, b])$ is compete with respect to the sup norm.

Theorem 6.27 *If a sequence $f_n : [a, b] \to \mathbb{R}$ is uniformly Cauchy, then it converges uniformly to a continuous function $f : [a, b] \to \mathbb{R}$.*

Proof: Each f_n is bounded by the Extreme Value Theorem, so (f_n) converges uniformly to a bounded function $f : [a, b] \to \mathbb{R}$ by Theorem 6.26. But f is then continuous by Theorem 6.14. □

7 Power Series

Reminder: A series is an infinite sum $\sum_{n=0}^{\infty} a_n = a_0 + a_1 + a_2 + \cdots$. The k^{th} partial sum is

$$s_k := \sum_{n=0}^{k} a_n = a_0 + a_1 + \cdots + a_k.$$

We say that a series converges precisely if the sequence (s_k) converges in the usual sense.

Definition A series $\sum_{n=0}^{\infty} a_n$ converges absolutely if the series $\sum_{n=0}^{\infty} |a_n|$ converges.

Proposition *If $\sum_{n=0}^{\infty} a_n$ converges absolutely, then it converges.*

Proof: Omitted, but proved in MATH1026. □

Theorem (Divergence Test) *If $\sum_{n=0}^{\infty} a_n$ converges, then $a_n \to 0$.*

Proof: By assumption, the sequence (s_k) of partial sums converges to some limit $L \in \mathbb{R}$. Hence, the subsequence $s_{k+1} \to L$ by Proposition 1.11. But $a_{k+1} = s_{k+1} - s_k \to L - L = 0$ by the Algebra of Limits. Since this is true for all $k \in \mathbb{Z}^+$, it follows that $a_n \to 0$. □

Note: This is called the Divergence Test because its **contrapositive** allows us to see when a series diverges. Indeed, the contrapositive says "if $a_n \not\to 0$, then $\sum_{n=0}^{\infty} a_n$ diverges".

7.1 Definition and Radius of Convergence

Definition A power series is a series of the form

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

where x is a real variable. The terms in the series are $a_n x^n$ and the k^{th} partial sum is

$$f_k(x) = \sum_{n=0}^{k} a_n x^n = a_0 + a_1 x + a_2 x^2 + \cdots + a_k x^k.$$

Note: For $x = 0$, each partial sum is $f_k(0) = a_0 \to a_0$, so the power series converges to a_0.

Definition 7.1 The radius of convergence of a power series $\sum_{n=0}^{\infty} a_n x^n$ is

$$R := \sup\{|x| : \sum_{n=0}^{\infty} |a_n x^n| \text{ converges}\}.$$

Theorem 7.2 *Let $R > 0$ be the radius of convergence of a power series $\sum_{n=0}^{\infty} a_n x^n$. Then, it converges absolutely for $|x| < R$ and diverges for $|x| > R$.*

Proof: Omitted, but proved in MATH1026. □

Much information rests on knowing the radius of convergence of a power series (but convergence or divergence when $x = \pm R$ is unclear). But one can find R by applying the Ratio Test.

Theorem 7.3 (Ratio Test) *Let $b_n > 0$ for all $n \in \mathbb{Z}^+$ such that $b_{n+1}/ab_n \to L$.*

(i) *If $L < 1$, then $\sum_{n=0}^{\infty} b_n$ converges.*
(ii) *If $L > 1$, then $\sum_{n=0}^{\infty} b_n$ diverges.*

Proof: Omitted, but also proved in MATH1026. □

Method – Finding the Radius of Convergence: Consider the power series $\sum_{n=0}^{\infty} a_n x^n$.

(i) Let $b_n := a_n x^n$ be the sequence of power series terms.
(ii) Compute $|b_{n+1}|/|b_n|$ and determine its limit as $n \to \infty$ in terms of $|x|$.
(iii) Use the Ratio Test, in particular that $|b_{n+1}|/|b_n| \to L < 1$ means the power series converges (absolutely) and $|b_{n+1}|/|b_n| \to L > 1$ means the power series diverges, to determine an upper bound on $|x|$ to ensure the limit in Step (ii) is less than one.

Many important functions are conveniently defined by power series, as we can see below.

Definition 7.4 The exponential, sine and cosine functions are defined as follows:

$$\exp : \mathbb{R} \to \mathbb{R}, \qquad \exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!},$$

$$\sin : \mathbb{R} \to \mathbb{R}, \qquad \sin(x) = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!},$$

$$\cos : \mathbb{R} \to \mathbb{R}, \qquad \cos(x) = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!}.$$

7.2 Differentiability of Power Series

The k^{th} partial sums of a power series $f(x) = \sum_{n=0}^{\infty} a_n x^n$ are polynomials of degree k. As such, Theorem 7.2 says that the sequence of partial sums (f_k) converges **pointwise** on the interval $(-R, R)$ to the power series function $f : (-R, R) \to \mathbb{R}$, where R is the radius of convergence.

Note: Uniform convergence on $(-R, R)$ is too much to ask for, but we get something close.

Theorem 7.5 *Let $f(x) = \sum_{n=0}^{\infty} a_n x^n$ have radius of convergence $R > 0$. Then, for any $\rho \in (0, R)$, the sequence of partial sums $(f_k(x))$ converges uniformly to f on $[-\rho, \rho]$.*

Proof: We will show that the sequence $f_k : [-\rho, \rho] \to \mathbb{R}$ is uniformly Cauchy (and consequently convergent by Theorem 6.26). First, since each f_k is a polynomial, it is continuous and hence bounded by the Extreme Value Theorem. Let $\varepsilon > 0$ be given. Since $\rho \in (0, R) \subseteq (-R, R)$, we know from Theorem 7.2 that $f(\rho) = \sum_{n=0}^{\infty} a_n \rho^n$ converges absolutely. Therefore, the sequence

$$b_k := \sum_{n=0}^{k} |a_n| \rho^n$$

converges, and is therefore Cauchy by Theorem 1.19. This means there exists $N \in \mathbb{Z}^+$ such that, for all $k > m \geq N$, we have $|b_k - b_m| < \varepsilon$. Hence, for all $k > m \geq N$,

$$\begin{aligned} \|f_k - f_m\| &= \sup \left\{ \left| \sum_{n=m+1}^{k} a_n x^n \right| : x \in [-\rho, \rho] \right\} \\ &\leq \sup \left\{ \sum_{n=m+1}^{k} |a_n| |x^n| : x \in [-\rho, \rho] \right\} \\ &= \sum_{n=m+1}^{k} |a_n| \rho^n \\ &= b_k - b_m \\ &< \varepsilon. \end{aligned}$$

□

Corollary 7.6 *The power series $f(x) = \sum_{n=0}^{\infty} a_n x^n$ is continuous on $(-R, R)$, for $R > 0$.*

Proof: The sequence of partial sums (f_k) is a sequence of continuous functions that converge uniformly on $[-\rho, \rho]$ to f for all $\rho \in (0, R)$, by Theorem 7.5. Hence, f is continuous on every closed bounded symmetric interval $[-\rho, \rho]$, meaning it is continuous at every $x \in (-R, R)$. □

Note: We now know that the functions introduced in Definition 7.4 are continuous!

Lemma 7.7 *The following power series have the same radius of convergence:*

$$f(x) = \sum_{n=0}^{\infty} a_n x^n \qquad \text{and} \qquad g(x) = \sum_{n=1}^{\infty} n a_n x^{n-1}.$$

Proof: For notation, we will introduce the sets

$$A := \{|x| : \sum_{n=0}^{\infty} |a_n x^n| \text{ converges}\} \qquad \text{and} \qquad B := \{|x| : \sum_{n=1}^{\infty} \left|n a_n x^{n-1}\right| \text{ converges}\}.$$

Hence, the radii of convergence of f and g are $\sup(A)$ and $\sup(B)$, respectively. The strategy is to show that each of these suprema is less than or equal to the other, from which equality will follow. Indeed, let $x \in \mathbb{R}$ with $|x| < \sup(B)$. Then, the sequence $s_k := \sum_{n=1}^{k} \left|n a_n x^{n-1}\right|$ of partial sums converges by Theorem 7.2, and is thus bounded (above) by Proposition 1.6. Hence,

$$t_k := \sum_{n=0}^{k} |a_n x^n| = |a_0| + |x| \sum_{n=1}^{k} \left|a_n x^{n-1}\right| \leq |a_0| + |x| s_k.$$

This shows that (t_k) is increasing and bounded above; it converges by the Monotone Convergence Theorem. As such, $|x| \in A$ and we have $\sup(B) \leq \sup(A)$. On the other hand, let $x \in \mathbb{R}$ with $|x| < \sup(A)$. If $x = 0$, $g(x)$ converges; we therefore assume that $|x| > 0$. Choose $\rho \in (|x|, \sup A)$ and notice that $t_k := \sum_{n=0}^{k} |a_n| \rho^n$ converges by Theorem 7.2, and is therefore bounded (above) by Proposition 1.6 once again. Now,

$$s_k = \sum_{n=1}^{\infty} \left|n a_n x^{n-1}\right| = |x|^{-1} \sum_{n=1}^{k} n \left(\frac{|x|}{\rho}\right)^n |a_n| \rho^n.$$

Since $|x|/\rho < 1$, it is true that $n(|x|/\rho)^n \to 0$ and thus it too is bounded, by $K > 0$ say. Hence,

$$s_k \leq |x|^{-1} \sum_{n=1}^{k} K |a_n| \rho^n = \frac{K}{|x|} t_k.$$

This shows that (s_k) is increasing and bounded above; it converges by the Monotone Convergence Theorem. As such, $|x| \in B$ and we have $\sup(A) \leq \sup(B)$. $\square$

Note: In summary, we showed that if $x \in \mathbb{R}$ such that g converges at x, so $|x| < \sup(B)$, then it follows that f converges at x, i.e. $|x| \in A$. We repeat the argument but flipped.

Theorem 7.8 *Let $R > 0$ be the radius of convergence of the power series*

$$f : (-R, R) \to \mathbb{R}, \qquad f(x) = \sum_{n=0}^{\infty} a_n x^n.$$

Then, f is differentiable with $f' = g$, where g is from Lemma 7.7.

Proof: Fix $\rho \in (0, R)$ and consider the sequence of continuously differentiable partial sums

$$f_k : [-\rho, \rho] \to \mathbb{R}, \qquad f_k(x) = \sum_{n=0}^{k} a_n x^n,$$

which converges pointwise to the function f (Theorem 7.2). Then, the sequence of derivatives

$$f_k' : [-\rho, \rho] \to \mathbb{R}, \qquad f_k'(x) = \sum_{n=1}^{k} n a_n x^{n-1}$$

converges uniformly (Theorem 7.5) to the function g (Lemma 7.7). By Theorem 6.19, it follows that f is differentiable with derivative $f' = g$. This is true on every closed bounded symmetric interval $[-\rho, \rho]$, meaning it is true at every $x \in (-R, R)$. □

Note: In words, the derivative of a power series is obtained by term-wise differentiation.

Corollary 7.9 *Let $R > 0$ be the radius of convergence of the power series*

$$f : (-R, R) \to \mathbb{R}, \qquad f(x) = \sum_{n=0}^{\infty} a_n x^n.$$

Then, f is smooth and its k^{th} derivative is given by the formula

$$f^{(k)}(x) = \sum_{n=k}^{\infty} \frac{n!}{(n-k)!} a_n x^{n-k}.$$

Sketch of Proof: This follows by applying Theorem 7.8 inductively. □

Note: In particular, we see that for $k \geq 0$, the coefficients of the power series are given by

$$a_k = \frac{f^{(k)}(0)}{k!}.$$

Corollary 7.10 *Let $\sum_{n=0}^{\infty} a_n x^n$ and $\sum_{n=0}^{\infty} b_n x^n$ both converge to $f : (-\rho, \rho) \to \mathbb{R}$ for some $\rho > 0$. Then, the power series are term-wise equal, that is $a_n = b_n$ for all $n \geq 0$.*

Proof: By Corollary 7.9, f is smooth. So, $f^{(n)}(0) = a_n/n! = b_n/n!$ for all $n \geq 0$, and $a_n = b_n$. □

Proposition 7.11 *The functions* exp, sin *and* cos *from* Definition 7.4 *are smooth, with*

$$\exp' = \exp, \qquad \sin' = \cos, \qquad \cos' = -\sin.$$

Proof: Smoothness is an immediate consequence of Corollary 7.9. As for their derivatives, we can compute them by referring to Theorem 7.8. Indeed, we see that for all $x \in \mathbb{R}$,

$$\exp'(x) = \sum_{n=1}^{\infty} n \left(\frac{1}{n!}\right) x^{n-1} = \sum_{n=1}^{\infty} \frac{x^{n-1}}{(n-1)!} = \sum_{m=0}^{\infty} \frac{x^m}{m!} = \exp(x).$$

Near-identical computations can be made for $\sin'$ and $\cos'$, completing the proof. $\square$

7.3 Properties of the Exponential Function

Lemma 7.12 *For all $x, y \in \mathbb{R}$, we have $\exp(x+y) = \exp(x)\exp(y)$.*

Proof: Fix $b \in \mathbb{R}$ and define the function $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = \exp(x)\exp(b-x)$. By the Product Rule and Proposition 7.11, f is differentiable. Along with the Chain Rule, we obtain

$$f'(x) = \exp'(x)\exp(b-x) - \exp(x)\exp'(b-x) = 0.$$

Proposition 3.11(iii) implies that f is constant. In particular, $f(x) = f(b)$ for all $x \in \mathbb{R}$, that is

$$\exp(x)\exp(b-x) = \exp(b).$$

But b was totally arbitrary, so setting $b = x + y$ then establishes the result. $\square$

Proposition 7.13 *The function $\exp : \mathbb{R} \to (0, \infty)$ is a bijection.*

Proof: First, we show that $\exp(x) \neq 0$ for all $x \in \mathbb{R}$. Indeed, if there was such an x, then Lemma 7.12 in the case $y = -x$ implies that

$$1 = \exp(0) = 0 \cdot \exp(-x) = 0,$$

which is obviously a contradiction. In particular, we have $\exp(0) = 1 > 0$ and exp is differentiable (and therefore continuous). It follows from the Intermediate Value Theorem that $\exp(x) > 0$ for all $x \in \mathbb{R}$. If this was not the case, and there exists $x \in \mathbb{R}$ such that $\exp(x) < 0$, we could find $c \in \mathbb{R}$ between x and 0 such that $\exp(c) = 0$, a contradiction to the first argument above. All of this tells us that exp really does map $\mathbb{R}$ to $(0, \infty)$. Next, $\exp'(x) = \exp(x) > 0$ for all $x \in \mathbb{R}$, which means exp is strictly increasing and hence injective by Proposition 3.11(iv). To show surjectivity, note that $\exp(x) = y$ implies by way of Lemma 7.12 that

$$\exp(-x)y = \exp(0) = 1 \qquad \Rightarrow \qquad \exp(-x) = \frac{1}{y}.$$

Thus, it is sufficient for surjectivity to show that exp takes all values in $[1, \infty)$. For any $y \in [1, \infty)$,

$$\exp(y) = \sum_{n=0}^{\infty} \frac{y^n}{n!} > 1 + \frac{y}{1!} > y.$$

Because $\exp(0) = 1 \leq y$, we see that y is a number between $\exp(0)$ and $\exp(y)$. But exp is continuous, so the Intermediate Value Theorem gives us some $x \in [0, y]$ such that $\exp(x) = y$. $\square$

Reminder: The natural logarithm is the function $\ln : (0, \infty) \to \mathbb{R}$ given by $\ln(x) = \displaystyle\int_1^x \frac{1}{t}\,\mathrm{d}t$.

Proposition 7.14 *The function* $\ln : (0, \infty) \to \mathbb{R}$ *is the inverse function of* $\exp : \mathbb{R} \to (0, \infty)$.

Proof: We must show that $\ln\big(\exp(x)\big) = x$ for all $x \in \mathbb{R}$, and $\exp\big(\ln(y)\big) = y$ for all $y \in (0, \infty)$. Let's consider the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = \ln\big(\exp(x)\big) - x$. By the Chain Rule (and Propositions 5.11 and 7.11), f is differentiable with derivative

$$f'(x) = \frac{\exp(x)}{\exp(x)} - 1 = 0.$$

Again, f is constant by Proposition 3.11(iii). In particular, $f(x) = f(0)$ for all $x \in \mathbb{R}$, that is

$$\ln\big(\exp(x)\big) - x = \ln\big(\exp(0)\big) - 0 = \ln(1) = 0 \qquad \Rightarrow \qquad \ln\big(\exp(x)\big) = x.$$

Next, for all $y \in (0, \infty)$, we can set $x = \ln(y)$ in the above equation to see that

$$\log\Big(\exp\big(\log(y)\big)\Big) = \log(y).$$

But log is injective, so the above implies $\exp\big(\log(y)\big) = y$, as required. $\square$

Note: The value of the exponential at one $e := \exp(1) = \displaystyle\sum_{n=0}^{\infty} \frac{1}{n!}$ is called Euler's number.

Remark People often denote $\exp(x)$ by the symbol e^x, and this is explained by Lemma 7.12. Indeed, when writing the statement of this lemma using the notation of Euler's number, it looks like one of the standard algebraic rules of integer exponents:

$$e^{x+y} = e^x \times e^y.$$

Now that we know ln is the inverse function to exp, it follows that $\ln(e) = 1$. However, there is an alternate way to define the number e via sequences, which will now be explored.

Proposition 7.15 *The sequence* $x_n = (1 + \frac{1}{n})^n$ *converges to* e.

Proof: It is clear that $x_n > 0$ for all $n \in \mathbb{Z}^+$. By Proposition 5.10, we see that

$$\ln(x_n) = \ln\left(1 + \frac{1}{n}\right)^n = n \ln\left(1 + \frac{1}{n}\right) = \frac{\ln\Big(1 + \frac{1}{n}\Big) - \ln(1)}{(1 + \frac{1}{n}) - 1} =: \frac{\ln(y_n) - \ln(1)}{y_n - 1}$$

by defining $y_n := 1 + \frac{1}{n}$. We can see from the Algebra of Limits that $y_n \to 1$. Because ln is differentiable (at one), and the above is its sequence of difference quotients at one, we have

$$\frac{\ln(y_n) - \ln(1)}{y_n - 1} \to \ln'(1) = 1.$$

In other words, $\ln(x_n) \to 1$. But exp is continuous, so the limit is preserved under taking the exponential, that is $\exp\big(\ln(x_n)\big) \to \exp(1) = e$. That said, we just established in Proposition 7.14 that $\exp\big(\ln(x_n)\big) = x_n$. Hence, we have $x_n \to e$ as required. $\square$

Definition 7.16 Let $a > 0$ and $x \in \mathbb{R}$. We say that a to the power x is the number

$$a^x := \exp(x \ln a).$$

Note: In the instance that $x \in \mathbb{Z}$, Definition 7.16 agrees with the usual definition of a^x. Working just a little bit harder, we can show that for any $p, q \in \mathbb{Z}$ with $q \neq 0$, we have

$$a^{p/q} = (\sqrt[q]{a})^p,$$

that is if b is the unique positive real number such that $b^q = a$, then $b^p = \exp\big((p/q)\ln a\big)$.

Proposition 7.17 *For any $r \in \mathbb{R}$, the function $f : (0, \infty) \to \mathbb{R}$ given by $f(x) = x^r$ is differentiable, with $f'(x) = rx^{r-1}$.*

Proof: Notice that $f(x) = \exp(r \ln x)$, so differentiability follows from the Chain Rule, with

$$f'(x) = r \exp'(r \ln x) \ln'(x) = r \exp(r \ln x) \frac{1}{x} = r \frac{x^r}{x} = rx^{r-1}. \qquad \square$$

7.4 Analyticity vs Smoothness

Reminder: A subset $U \subseteq \mathbb{R}$ is open if every point in U is contained in a symmetric open subinterval of U, that is for all $a \in U$, there exists $\delta > 0$ such that $(a - \delta, a + \delta) \subseteq U$.

Definition 7.18 Let $U \subseteq \mathbb{R}$ be open. We say $f : U \to \mathbb{R}$ is analytic if, for each $x_0 \in U$, there exists $\varepsilon > 0$ and a power series of this form which converges to f for all $x \in (x_0 - \varepsilon, x_0 + \varepsilon)$:

$$\sum_{n=0}^{\infty} a_n (x - x_0)^n.$$

Note that if $f : U \to \mathbb{R}$ is analytic, it is smooth by Corollary 7.9. However, the converse is false: there exist smooth functions that are **not** analytic.

Definition 7.19 Let $p(x) = a_0 + a_1 x + \cdots + a_k x^k$ and define the function $f_p : \mathbb{R} \to \mathbb{R}$ by

$$f_p(x) = \begin{cases} p(1/x)\exp\big(-1/x\big) & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}.$$

Lemma 7.20 (Exponentials Beat Powers for Sequences) *For any $k \in \mathbb{Z}^+$, the sequence*

$$a_n = n^k \exp(-n) \to 0.$$

Proof: Consider the ratio of successive terms and apply the Algebra of Limits to see that

$$\frac{a_{n+1}}{a_n} = \left(1 + \frac{1}{n}\right)^k \exp(-1) \to \frac{1}{e} < 1.$$

Hence, $\sum_{n=1}^{\infty} a_n$ converges by the Ratio Test, and thus $a_n \to 0$ by the Divergence Test. □

Lemma 7.21 (Exponentials Beat Powers for Functions) *For any $k \in \mathbb{Z}^+$, we have*

$$\lim_{x\to\infty} x^k \exp(-x) = 0.$$

Proof: Let $g(x) = x^k \exp(-x)$. We must show that, for each $\varepsilon > 0$, there exists $K \in \mathbb{R}$ such that, for all $x > K$, we have $\big|g(x) - 0\big| < \varepsilon$. Indeed, let $\varepsilon > 0$ be given. Since the sequence $g(n) = n^k \exp(-n) \to 0$ by Lemma 7.20, there exists $N \in \mathbb{Z}^+$ such that, for all $n \geq N$, we have $\big|g(n) - 0\big| < \varepsilon$. Consequently, $0 < g(n) < \varepsilon$. By taking $K = \max\{k, N\}$, for all $x > K$, we have

$$g'(x) = (kx^{k-1} - x^k)\exp(-x) = -x^{k-1}(x-k)\exp(-x) < 0$$

by the Chain and Product Rules. Therefore, g is strictly decreasing by Proposition 3.11(v). Because $K \geq N$, we see that for all $x > K$, $0 < g(x) < g(K) < \varepsilon$, so we are done. □

Lemma 7.22 *Let $f_p : \mathbb{R} \to \mathbb{R}$ be as in* Definition 7.19*. Then,*

$$\lim_{x\to 0} f_p(x) = 0.$$

Proof: Let $p(x) = a_0 + a_1 x + \cdots + a_k x^k$ and denote by m_j the monomial of degree j for each $j \in \mathbb{N}$, that is $m_j(x) = x^j$. Then, we can write $f_p = a_0 f_{m_0} + a_1 f_{m_1} + \cdots + a_k f_{m_k}$. By the Algebra of Limits, it suffices to prove the claim in the case $p(x) = m_k(x) = x^k$. To that end, let $\varepsilon > 0$ be given. By Lemma 7.21, there exists $K > 0$ such that, for all $x > K$, we have

$$\Big|x^k \exp(-x) - 0\Big| < \varepsilon.$$

Define $\delta = 1/K > 0$. Then, for all $x \in (0, \delta)$,

$$0 \leq f_p(x) = (1/x)^k \exp\big(-1/x\big) < \varepsilon,$$

since $1/x > 1/\delta = K$. But for all $x \in (-\delta, 0)$,

$$0 < f_p(x) < \varepsilon.$$

Therefore, for all $x \in \mathbb{R}$ with $0 < |x - 0| < \delta$, we have $\big|f_p(x) - 0\big| < \varepsilon$, as required. □

Theorem 7.23 *Let* $f_p : \mathbb{R} \to \mathbb{R}$ *be as in* Definition 7.19. *Then,* f_p *is differentiable with*

$$f_p' = f_q, \qquad \text{where } q(x) = x^2(p(x) - p'(x)).$$

Proof: On the open set $(0, \infty)$, f_p coincides with the differentiable function $p(1/x)\exp\bigl(-1/x\bigr)$. Therefore, the Localisation Lemma (Lemma 2.24) along with the Chain and Product Rules imply that it is differentiable on $(0, \infty)$ with

$$f'(x) = p'(1/x)(-x^{-2})\exp\bigl(1/x\bigr) + p(1/x)\exp\bigl(-1/x\bigr)(x^{-2}) = q(1/x)\exp\bigl(-1/x\bigr).$$

Similarly, f_p coincides with the differentiable function 0 on the open set $(-\infty, 0)$; it is differentiable on this open set (again by the Localisation Lemma) with derivative 0. It remains to show that f_p is differentiable at zero with $f_p'(0) = 0$. In other words, we must prove that

$$\lim_{x\to 0} \frac{f_p(x) - f_p(0)}{x - 0} = \lim_{x\to 0} \frac{f_p(x)}{x} = 0.$$

But $f_p(x)/x = f_s(x)$, where $s(x) = xp(x)$, so this follows immediately from Lemma 7.22. □

Corollary 7.24 *The function* $f_p : \mathbb{R} \to \mathbb{R}$ *is smooth, with* $f_p^{(n)}(0) = 0$ *for all* $n \in \mathbb{N}$.

Proof: Let $X \coloneqq \{f_p : p \text{ is a polynomial}\}$ be set set of such functions. Then, for all $f_p \in X$, we have $f_p(0) = 0$ by definition. But f_p is differentiable and $f_p' \in X$ (Theorem 7.23). Hence, every element of X is infinitely differentiable and has all its derivatives zero. □

Note: Almost every f_p is smooth but **not** analytic. The only exception is f_0, that is where we take the polynomial to be $p(x) = 0$. In this case, $f_0 \equiv 0$ which is trivially analytic.

Proposition *The function* f_1*, as defined in* Definition 7.19 *with* $p(x) = 1$*, is* **not** *analytic.*

Proof: Explicitly, notice that $f_1 : \mathbb{R} \to \mathbb{R}$ is the function given by

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

By Corollary 7.24, it is smooth and all its derivatives at zero are zero. Assume to the contrary that f_1 is analytic. Then, there exist $\varepsilon > 0$ and coefficients (a_n) such that, for all $x \in (-\varepsilon, \varepsilon)$,

$$f_1(x) = \sum_{n=0}^{\infty} a_n x^n.$$

Using Corollaries 7.9 and 7.24, we know that $a_n = f_1^{(n)}(0)/n! = 0$. Therefore, $f_1(x) = 0$ for all $x \in (-\varepsilon, \varepsilon)$. However, $f_1(\varepsilon/2) = \exp\bigl(-2/\varepsilon\bigr) > 0$ by the definition of f_1, a contradiction. □