Calculus

Justin Keung

April 13, 2023

Contents

Chapter 1

Sequences

1.1 Definition

A sequence is a specific type of function. A function is a specific type of binary relations. All sequences are functions, and all functions are binary relations, but not all binary relations are functions. Are all functions sequences?

Definition 1.1. A sequence is a function $f : \mathbb{N} \mapsto \mathbb{R}$. For convenience, in this course we define a sequence to be $f : \mathbb{N}^+ \mapsto \mathbb{R}$ instead.

Notation. We can also denote a sequence as $(a_n)_{n\geq 1}$, where $a_n = f(n)$. Writing a_n or $a_1, a_2, ...$ is also fine as long as it is clear enough that they denote a sequence.

1.2 Monotonicity

Definition 1.2. A sequence is increasing $\Leftrightarrow a_{n+1} \geq a_n \forall n \geq 1$.

Definition 1.3. A sequence is decreasing $\Leftrightarrow a_{n+1} \leq a_n \forall n \geq 1$.

Definition 1.4. A sequence is monotonic $\Leftrightarrow$ it is either increasing or decreasing (or both).

Definition 1.5. A sequence is strictly increasing $\Leftrightarrow a_{n+1} > a_n \forall n \geq 1$.

Definition 1.6. A sequence is decreasing $\Leftrightarrow a_{n+1} < a_n \forall n \geq 1$.

1.3 Convergence and divergence

Definition 1.7. A sequence converges if it converges to a limit $l \in \mathbb{R}$, ∞, or $-\infty$. A sequence diverges if it does not converge.

1.3.1 $\varepsilon - N$ definition of convergence

Let $(a_n)_{n\geq 1}$ be a sequence.

Definition 1.8. $(a_n)_{n\geq 1}$ converges to a limit l if and only if

$$\forall \varepsilon > 0, \exists N \in \mathbb{N}(\forall n \in \mathbb{N}(n > N \Rightarrow |a_n - l| < \varepsilon))$$

We denote this as $\lim_{n\to\infty} a_n = l$ or $(a_n)_{n\geq 1} \to l$. As above, $a_n \to l$ would also be fine.

Note that n should be strictly greater than N.

Intuition. If somebody gives us an arbitrary $\varepsilon > 0$, then we can always find an N such that every term after that will be within ε of the limit. So if ε is very close to 0, then this guarantees that after some point, all terms in the sequence will be very close to the limit.

The statement of the definition itself isn't very helpful when we try to prove limits with it. A general template for answering problems is as follows:

Template. Given some $\varepsilon > 0$, we want to find limit l and some $N \in \mathbb{N}$ such that $\forall n > N, |a_n - l| < \varepsilon$. We speculate that $l =$ [some value]. Now,

$$\begin{aligned} |a_n - l| < \varepsilon &\Leftrightarrow [\text{substitute } a_n \text{ and } l] \\ &\Leftrightarrow [\text{simplify}] \\ &\Leftrightarrow [\text{remove the absolute operators and justify}] \\ &\Leftrightarrow \ldots \\ &\Leftrightarrow n > [\text{some function of } \varepsilon, \text{ let's call this } N(\varepsilon)] \end{aligned}$$

Choose $N \in \mathbb{N}$ such that $N > N(\varepsilon)$. Then $\forall n > N, n > N(\varepsilon)$.

$$\begin{aligned} |a_n - l| &= \ldots \\ &< \varepsilon \end{aligned}$$

Since ε arbitrary, $\forall \varepsilon > 0, \exists N \in \mathbb{N}(\forall n \in \mathbb{N}(n > N \Rightarrow |a_n - l| < \varepsilon))$, where $l =$ blah. So limit is blah.

If the question gives us the limit and asks us to prove it directly, just change around the sentences a bit.

Example. (MMT 1, 6(a)) Let α be a positive, real constant. Use a direct proof to show that the sequence $(a_n)_{n\geq 1} = (n^{-\alpha})$ converges to 0 as n tends to infinity.

Proof. Given some $\varepsilon > 0$, we want to find some $N \in \mathbb{N}$ such that $\forall n > N, |a_n - l| < \varepsilon$, where $l = 0$. Now,

$$\begin{aligned}
|a_n - l| < \varepsilon &\Leftrightarrow \left|n^{-\alpha} - 0\right| < \varepsilon \\
&\Leftrightarrow \left|\frac{1}{n^\alpha}\right| < \varepsilon \\
&\Leftrightarrow \frac{1}{n^\alpha} < \varepsilon && \text{since } n > 0 \Rightarrow \frac{1}{n^\alpha} > 0 \\
&\Leftrightarrow n^\alpha > \frac{1}{\varepsilon} \\
&\Leftrightarrow n > \frac{1}{\varepsilon^\alpha} = \varepsilon^{-\alpha}
\end{aligned}$$

Choose $N \in \mathbb{N}$ such that $N > \varepsilon^{-\frac{1}{\alpha}}$. Then $\forall n > N, n > \varepsilon^{-\frac{1}{\alpha}}$.

$$\begin{aligned}
|a_n - l| &= \left|n^{-\alpha} - 0\right| \\
&= \frac{1}{n^\alpha} && \text{from above} \\
&< \frac{1}{(\varepsilon^{-\frac{1}{\alpha}})^\alpha} \\
&= \varepsilon
\end{aligned}$$

Since ε arbitrary, $\forall \varepsilon > 0, \exists N \in \mathbb{N}(\forall n \in \mathbb{N}(n > N \Rightarrow |a_n - l| < \varepsilon))$. So limit is 0. $\square$

1.3.2 Converging to ∞ and $-\infty$

Definition 1.9. $(a_n)_{n\geq 1}$ converges to ∞ if and only if

$$\forall r \in \mathbb{R}, \exists N \in \mathbb{N}(\forall n \in \mathbb{N}(n \geq N \Rightarrow a_n > r))$$

Note that this time, the requirement is n greater than or equal to N, but not strictly greater than N. I don't know why this is the case, but we'll just accept it. Maybe the slides are wrong.

Intuition. If r is arbitrarily large, then this guarantees that after some point, all the terms will be larger than the arbitrarily large r, i.e. mega large.

Example. Use a direct proof to show that the sequence $(a_n)_{n\geq 1} = n!$ converges to ∞.

Proof. Given some $r \in \mathbb{R}$, we want to find some $N \in \mathbb{N}$ such that $\forall n \geq N, a_n > r$.

$$\begin{aligned}
a_n > r &\Leftrightarrow n! > r \\
&\Leftrightarrow n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 1 > r
\end{aligned}$$

Choose $N \in \mathbb{N}$ such that $N > r$. Then $\forall n \geq N, n > r$.

$$\begin{aligned} a_n &= n! \\ &= n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 1 \\ &\geq n \cdot 1 \cdot 1 \cdot \ldots \cdot 1 \\ &= n \\ &> r \end{aligned}$$

Since r arbitrary, $\forall r \in \mathbb{R}, \exists N \in \mathbb{N}(\forall n \in \mathbb{N}(n \geq N \Rightarrow a_n > r))$. So a_n converges to ∞. $\square$

Definition 1.10. $(a_n)_{n\geq 1}$ converges to $-\infty$ if and only if

$$\forall r \in \mathbb{R}, \exists N \in \mathbb{N}(\forall n \in \mathbb{N}(n \leq N \Rightarrow a_n < r))$$

Also note that n is less than or equal to N, but not strictly less than N.

Example. Use a direct proof to show that the sequence $(a_n)_{n\geq 1} = \ln \frac{1}{n}$ converges to $-\infty$.

Proof. Given some $r \in \mathbb{R}$, we want to find some $N \in \mathbb{N}$ such that $\forall n \geq N, a_n < r$.

$$\begin{aligned} a_n < r &\Leftrightarrow \ln \frac{1}{n} < r \\ &\Leftrightarrow \frac{1}{n} < e^r \\ &\Leftrightarrow n > e^{-r} \end{aligned}$$

Choose $N \in \mathbb{N}$ such that $N > e^{-r}$. Then $\forall n \geq N, n > e^{-r}$.

$$\begin{aligned} a_n &= \ln \frac{1}{n} \\ &< \ln \frac{1}{e^{-r}} \\ &= \ln e^r \\ &= r \end{aligned}$$

Since r arbitrary, $\forall r \in \mathbb{R}, \exists N \in \mathbb{N}(\forall n \in \mathbb{N}(n \geq N \Rightarrow a_n < r))$. So a_n converges to $-\infty$. $\square$

1.3.3 Divergence

A sequence diverges if it does not converge to a limit $l \in \mathbb{R}$, does not converge to ∞, and does not converge to $-\infty$.

Example. Prove that the sequence $(a_n)_{n\geq 1} = (-1)^n$ diverges.

Proof. We will prove that a_n diverges by contradiction.

Assume a_n converges to a limit $l \in \mathbb{R}$. Then

$$\forall \varepsilon > 0, \exists N \in \mathbb{N}(\forall n \in \mathbb{N}(n > N \Rightarrow |a_n - l| < \varepsilon))$$

Suppose $\varepsilon = 1$. Choose arbitrary $n > N$. Then $2n > N$ and $2n + 1 > N$.

$$\begin{aligned}
|a_{2n} - l| < \varepsilon &\Leftrightarrow \left|(-1)^{2n} - l\right| < 1 \\
&\Leftrightarrow |1 - l| < 1 \\
&\Leftrightarrow -1 < 1 - l < 1 \\
&\Leftrightarrow -2 < -l < 0 \\
&\Leftrightarrow 0 < l < 2 \\
|a_{2n+1} - 1| < \varepsilon &\Leftrightarrow \left|(-1)^{2n+1} - l\right| < 1 \\
&\Leftrightarrow |-1 - l| < 1 \\
&\Leftrightarrow -1 < -1 - l < 1 \\
&\Leftrightarrow 0 < -l < 2 \\
&\Leftrightarrow -2 < l < 0
\end{aligned}$$

Since no $l \in \mathbb{R}$ can satisfy both $0 < l < 2$ and $-2 < l < 0$, we have a contradiction. So a_n does not converge to a limit $l \in \mathbb{R}$.

Assume a_n converges to ∞. Then

$$\forall r \in \mathbb{R}, \exists N \in \mathbb{N}(\forall n \in \mathbb{N}(n \geq N \Rightarrow a_n > r))$$

Suppose $r = 2$. Since $a_n = -1$ or 1, and $-1 < 2$ and $1 < 2$, we have a contradiction. So a_n does not converge to ∞.

Assume a_n converges to $-\infty$. Then

$$\forall r \in \mathbb{R}, \exists N \in \mathbb{N}(\forall n \in \mathbb{N}(n \geq N \Rightarrow a_n < r))$$

Suppose $r = -2$. Since $a_n = -1$ or 1, and $-1 > -2$ and $1 > -2$, we have a contradiction. So a_n does not converge to $-\infty$.

Therefore, a_n diverges. □

1.4 Sandwich theorem (squeeze theorem)

Definition 1.11. Suppose $(l_n)_{n\geq 1} \to l$ and $(u_n)_{n\geq 1} \to l$, where $l \in \mathbb{R}$.

$$\exists N \in \mathbb{N} : \forall n \geq N, l_n \leq a_n \leq u_n \Rightarrow a_n \to l$$

Again, n is greater than or equal to, but somehow not strictly greater than, N.

Proof. Given some $\varepsilon > 0$, since $l_n \to l$,

$$\exists N_1 \in \mathbb{N} : \forall n > N_1, |l_n - l| < \varepsilon$$

Also, since $u_n \to l$,

$$\exists N_2 \in \mathbb{N} : \forall n > N_2, |u_n - l| < \varepsilon$$

Let $N' = \max(N_1, N_2)$. Then $\forall n > N'$,

$$\begin{aligned}
|l_n - l| < \varepsilon &\Leftrightarrow -\varepsilon < l_n - l < \varepsilon \\
&\Leftrightarrow l - \varepsilon < l_n < l + \varepsilon
\end{aligned}$$

and similarly, $l - \varepsilon < u_n < l + \varepsilon$.

Assume that

$$\exists N \in \mathbb{N} : \forall n \geq N, l_n \leq a_n \leq u_n$$

Then $\forall n > \max(N, N')$,

$$\begin{aligned} l - \varepsilon < l_n \leq a_n \leq u_n < l + \varepsilon &\Leftrightarrow l - \varepsilon < a_n < l + \varepsilon \\ &\Leftrightarrow -\varepsilon < a_n - l < \varepsilon \\ &\Leftrightarrow |a_n - l| < \varepsilon \end{aligned}$$

□

Note that it is possible to use a constant sequence for one of l_n or u_n.

Example. (MMT 1 6(b)) Use the sandwich theorem to show that

$$\lim_{n\to\infty} \frac{n!}{n^n} = 0$$

Solution. Let $a_n = \frac{n!}{n^n}$. Let $l_n = 0$. Then $\lim\limits_{n\to\infty} l_n = 0$.

$$\forall n > 0, n! > 0 \text{ and } n^n > 0 \Rightarrow a_n > 0 \Rightarrow a_n > l_n$$

Let $u_n = \frac{1}{n}$. Then $\lim\limits_{n\to\infty} u_n = 0$. $\forall n > 0$,

$$\begin{aligned} a_n &= \frac{n!}{n^n} \\ &= \frac{n}{n} \cdot \frac{n-1}{n} \cdot \frac{n-2}{n} \cdot \ldots \cdot \frac{1}{n} \\ &= 1 \cdot (1 - \frac{1}{n}) \cdot (1 - \frac{2}{n}) \cdot \ldots \cdot \frac{1}{n} \end{aligned}$$

Since every term ≤ 1, we have $a_n \leq \frac{1}{n} = u_n$. So $\forall n > 0, l_n \leq a_n \leq u_n$. Since $l_n \to 0$ and $u_n \to 0$ as $n \to \infty$, by sandwich theorem, $a_n \to 0$.

1.5 Triangle inequality

Lemma 1.12.

$$\forall x, y > 0, x \leq y \Leftrightarrow x^2 \leq y^2$$

Proof. Assume $x \leq y$. Then

$$\begin{aligned} x \leq y \Rightarrow x^2 &\leq y * x && \text{since } x > 0 \\ &\leq y * y && \text{since } x \leq y \\ &= y^2 \end{aligned}$$

So $x \leq y \Rightarrow x^2 \leq y^2$.

Assume $x^2 \leq y^2$. Then

$$\begin{aligned} x^2 \leq y^2 &\Rightarrow x^2 - y^2 \leq 0 \\ &\Rightarrow (x-y)(x+y) \leq 0 \\ &\Rightarrow x - y \leq 0 && \text{since } x, y > 0 \Rightarrow x + y > 0 \\ &\Rightarrow x \leq y \end{aligned}$$

So $x^2 \leq y^2 \Rightarrow x \leq y$. □

Lemma 1.13 (Triangle inequality)**.**

$$|a+b| \leq |a| + |b|$$

Proof.

$$\begin{aligned} |a+b|^2 &= (a+b)^2 \\ &= a^2 + 2ab + b^2 \\ &\leq a^2 + 2|ab| + b^2 \\ &= |a|^2 + 2|ab| + |b|^2 \\ &= |a|^2 + 2|a||b| + |b|^2 \\ &= (|a| + |b|)^2 \end{aligned}$$

From lemma, $|a+b| \leq |a| + |b|$. □

Here is another proof of the triangle inequality:

Proof.

$$a + b \leq |a| + b \leq |a| + |b|$$
$$-a - b \leq |a| - b \leq |a| + |b| \Rightarrow a + b \leq -(|a| + |b|)$$

So we have

$$-(|a| + |b|) \leq a + b \leq |a| + |b| \Leftrightarrow |a+b| \leq |a| + |b|$$

□

1.6 Bounds

Suppose $X \subseteq \mathbb{R}$.

- $u \in \mathbb{R}$ is an upper bound of X if $x \leq u$ for all $x \in X$.
- $s \in \mathbb{R}$ is the supremum (least upper bound) of X if $s \leq u$ for all upper bounds u of X.
- $l \in \mathbb{R}$ is a lower bound of X if $x \geq l$ for all $x \in X$.
- $i \in \mathbb{R}$ is the infimum (greatest lower bound) of X if all $i \geq l$ for all lower bounds l of X.

To describe the bounded conditions of X, we have

- X is bounded above if X has an upper bound.
- X is bounded below if X has a lower bound.
- X is bounded if X has an upper bound and a lower bound.

1.7 Cauchy sequences

Definition 1.14. A sequence $(a_n)_{n\geq 1}$ is a Cauchy sequence (in the real numbers) if and only if

$$\forall \varepsilon > 0,\ \exists N \in \mathbb{N} : \forall n, m \in \mathbb{N},\ n, m > N \Rightarrow |a_n - a_m| < \varepsilon$$

We will use *convergent* to mean "converging to some limit $l \in \mathbb{R}$".

Proposition 1.15. All convergent sequences are Cauchy.

Proof. Suppose the sequence a_n converges to $l \in \mathbb{R}$. Then

$$\forall \varepsilon > 0,\ \exists N \in \mathbb{N} : \forall n \in \mathbb{N},\ n > N \Rightarrow |a_n - l| < \frac{\varepsilon}{2}$$

Take arbitrary $\varepsilon > 0$. Let N satisfy the above for this ε. Take arbitrary $n, m > N$. Then $|a_n - l| < \frac{\varepsilon}{2}$ and $|a_m - l| < \frac{\varepsilon}{2}$.

$$\begin{aligned} |a_n - a_m| &= |a_n - l - (a_m - l)| \\ &\leq |a_n - l| + |-(a_m - l)| \text{ by triangle inequality} \\ &= |a_n - l| + |a_m - l| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\ &= \varepsilon \end{aligned}$$

Since n, m, ε arbitrary, a_n is a Cauchy sequence. □

Using the contrapositive of this proposition, we present another explanation to why $(-1)^n$ does not converge to a limit $l \in \mathbb{R}$.

Explanation. We want to show that $(-1)^n$ is not a Cauchy sequence. Observe that

$$\forall N \in \mathbb{N}, \exists n, m > N : |a_n - a_m| = 2$$

Pick some $0 < \varepsilon < 2$. Then there does not exist $N \in \mathbb{N}$ such that $\forall n, m > N, |a_n - a_m| < \varepsilon$. So $(-1)^n$ is not a Cauchy sequence. By contrapositive, $(-1)^n$ does not converge to a limit $l \in \mathbb{R}$.

Proposition 1.16. All Cauchy sequences are bounded.

Proof. We want to show that

$$\exists M \in \mathbb{R} : |a_n| \leq M\ \forall n \in \mathbb{N}$$

Remark. We are taking the absolute value of a_n purely for convenience; equally, we could have found separate upper and lower bounds, instead of just one M, but $|x_n| < M$ encapsulates both bounds perfectly fine. We simply take the magnitude of the bound which has the larger magnitude, then the interval $[-M, M]$ covers both bounds (picture the number line), regardless if they are positive or negative.

Since a_n is a Cauchy sequence, we have

$$\forall \varepsilon > 0, \; \exists N \in \mathbb{N} : \forall n, m \in \mathbb{N}, \; n, m > N \Rightarrow |a_n - a_m| < \varepsilon$$

Let $\varepsilon = 1$. Then substituting into the definition above,

$$\exists N \in \mathbb{N} : \forall n, m \in \mathbb{N}, \; n, m > N \Rightarrow |a_n - a_m| < 1$$

Let

$$M = 1 + \max\{|a_1|, |a_2|, |a_3|, ..., |a_{N+1}|\}$$

Remark. Instead of setting $M = \varepsilon + ...$, we choose the specific case when $\varepsilon = 1$ (or any other value) so that M would not depend on ε, which we would want to be arbitrarily small. Note that we are using the N when $\varepsilon = 1$, but not any N that satisfies any arbitrary ε, so none of the proof depends on the value of ε. We can choose any ε and any corresponding N, but once we have chosen a value, we fix it regardless of the value of ε. This justifies the step of substituting $|a_n - a_{N+1}| < 1$ in the next part.

Then $\forall n > N$,

$$\begin{aligned} |a_n| &= |a_n + a_{N+1} - a_{N+1}| \\ &\leq |a_n - a_{N+1}| + |a_{N+1}| \text{ by triangle inequality} \\ &< 1 + |a_{N+1}| \\ &\leq M \end{aligned}$$

Remark. We include a_{N+1} when defining M because otherwise, we wouldn't know what $|a_n - a_N|$ evaluates to, since the definition for Cauchy only applies for $n > N$ but not $n = N$. $|a_n - a_N|$ isn't directly related to M, but we want to be able to do the trick of subtracting and adding the same thing, which is a_{N+1} in this case.

So M is an upper bound for a_n. □

Although we can combine Proposition 1.15 and Proposition 1.16 to prove the following proposition transitively, we can do so directly in a fairly straightforward manner:

Proposition 1.17. All convergent sequences are bounded.

Proof. Suppose the sequence a_n converges to some limit $l \in \mathbb{R}$. Then

$$\forall \varepsilon > 0, \; \exists N \in \mathbb{N} : \forall n \in \mathbb{N}, \; n > N \Rightarrow |a_n - l| < \varepsilon$$

Let $\varepsilon = 1$ and let N satisfy the above definition when $\varepsilon = 1$. Then for all $n > N$, we have

$$\begin{aligned} |a_n| - |l| &\leq |a_n - l| \text{ by reverse triangle inequality} \\ &< 1 \\ \Leftrightarrow |a_n| &< |l| + 1 \end{aligned}$$

Let

$$M = \max\{|l| + 1, |a_1|, |a_2|, ..., |a_N|\}$$

Then $\forall n > N$, $|a_n| < |l| + 1 \leq M$. Otherwise, $|a_1|, |a_2|, ..., |a_N| \leq M$ by construction of M. So $|a_n| \leq M \ \forall n \in \mathbb{N}$. □

1.7.1 Subsequences

Definition 1.18. A subsequence is an infinite ordered subset of a sequence.

Note that a subsequence, by default, has infinitely many terms. Since it is defined to be a subset of a sequence, the sequence from which the subsequence is taken from must also have infinitely many terms.

Notation. We denote a subsequence of a sequence $(a_n)_{n\geq 1}$ as $(a_{n_i})_{i\geq 1}$, where n_i is a strictly increasing sequence in $\mathbb{N}^+$.

Proposition 1.19. For any subsequence $(a_{n_i})_{i\geq 1}$, $n_i \geq i$.

Explanation. A sequence is a subsequence of itself, so in this case we have $n_i = i$ since we are not omitting any terms to form the subsequence (which is the sequence itself). For any other subsequence, suppose the first omitted term is a_k. Then

$$\begin{cases} n_i = i & i < k \\ n_i > i & i \geq k \end{cases}$$

so in general, for any subsequence, we have $n_i \geq i$.

Theorem 1.20. Let a_n be a sequence and a_{n_i} be a subsequence of a_n. If a_n converges to limit $l \in \mathbb{R}$, then a_{n_i} also converges to l.

Proof. Since a_n converges to l, given some $\varepsilon > 0$,

$$\exists N \in \mathbb{N} : \forall i > N, |a_i - l| < \varepsilon$$

Since $n_i \geq i$, $i > N$ implies $n_i > N$, so

$$\forall i > N, |a_{n_i} - l| < \varepsilon$$

□

Theorem 1.21. Every subsequence of real numbers has a monotonic subsequence.

Proof. Let a_n be a peak if

$$a_n > a_m \qquad\qquad \forall m > n$$

If the sequence has infinitely many peaks at $n_1 < n_2 < n_3 < ...$, then $a_{n_1} > a_{n_2} > a_{n_3} > ...$ is a strictly decreasing subsequence.

Otherwise, the sequence has only finitely many peaks at $n_1 < n_2 < ... < n_k$, or no peaks. Consider a_{n_k+1}. Since $a_{n_{k+1}}$ is not a peak,

$$\exists i > k + 1 : a_{n_i} \geq a_{n_k+1}$$

and since a_{n_i} is not a peak, we can construct a (not strictly) increasing subsequence in this way. □

1.7.2 Completeness

We define two notions of completeness:

Definition 1.22. A metric space (X, d) is Cauchy complete if and only if all Cauchy sequences in (X, d) converges to an element in X.

Example. The set of rational numbers $\mathbb{Q}$ is not Cauchy complete.

Consider the sequence

$$a_1 = 1, a_{n+1} = \frac{x_n + \frac{2}{x_n}}{2}$$

Every term in the sequence is in $\mathbb{Q}$, but the sequence converges to $\sqrt{2}$, which is not in $\mathbb{Q}$.

Consider another sequence in the rationals

$$a_1 = 3.1, a_2 = 3.14, a_3 = 3.142, a_4 = 3.1416, ...$$

where a_n represents the nth decimal approximation of π. The sequence converges to $\pi \notin \mathbb{Q}$.

Definition 1.23 (Axiom of Dedekind completeness)**.** A partially ordered set X is Dedekind complete if and only if it has the least-upper-bound property, i.e. every non-empty subset of X with an upper bound has a least upper bound (i.e. supremum).

Theorem 1.24 (Fundamental theorem of analysis)**.** Every increasing sequence of real numbers a_n with an upper bound:

- has a supremum s, and
- converges to s.

Proof. A non-empty subset of the partially ordered set $\mathbb{R}$ is equivalent to an increasing sequence of real numbers. So, the axiom of Dedekind completeness tells us that every increasing sequence of real numbers with an upper bound has a supremum.

We prove the second point in two stages.

In the first stage, we claim that

$$\exists N \in \mathbb{N}(|a_N - s| < \varepsilon)$$

We prove this by contradiction. Suppose $\forall n \in \mathbb{N}, |a_n - s| \geq \varepsilon$. Since s is the supremum, we have $s \geq a_n \forall n \in \mathbb{N}$, so $s - a_n \geq \varepsilon \forall n \in \mathbb{N}$. Rearranging the inequality gives

$$a_n \leq s - \varepsilon \qquad \forall n \in \mathbb{N}$$

By definition, $s - \varepsilon$ is an upper bound. However, since $\varepsilon > 0$, $s - \varepsilon < s$, which contradicts the fact that s is the supremum, i.e. least upper bound. So the claim is true.

In the second stage, we claim that

$$\forall n \in \mathbb{N}(n > N \Rightarrow |a_n - s| < \varepsilon)$$

where N satisfies the criterion in our first claim. Since a_n increasing, $n > N \Rightarrow a_n \geq a_N$. So

$$\begin{aligned} s - a_n &\leq s - a_N \\ &< \varepsilon && \text{from above} \end{aligned}$$

Since $s - a_n \geq 0$, we have $|a_n - s| < \varepsilon$. $\square$

Chapter 2

Continuous functions

2.1 Accumulation point

Definition 2.1 (Neighbourhood)**.** A set $A \subseteq \mathbb{R}$ is a neighbourhood of a point a, if there exists an open interval I such that

$$a \in I \subseteq A$$

Note that an interval is a set.

Definition 2.2 (Accumulation point)**.** A real number ζ is an accumulation point of a set $A \in \mathbb{R}$ if every neighbourhood of ζ contains an infinite number of elements in A.

Note that ζ does not have to be an element of A.

- $\mathbb{Z}$ has no accumulation point since there exist a neighbourhood which does not contain an infinite number of members of $\mathbb{Z}$. For every $x \in \mathbb{Z}$, consider the neighbourhood

  $$(x - \frac{1}{2},\ x + \frac{1}{2})$$

  which contains no members of $\mathbb{Z}$ other than x.

- Every point $x \in \mathbb{R}$ is an accumulation point, since all neighbourhoods of x contain an infinite number of members of $\mathbb{R}$.

- Similarly, every rational number $x \in Q$ is an accumulation point of $\mathbb{Q}$. Every irrational number is also an accumulation point of $\mathbb{Q}$. So every real number is an accumulation point of $\mathbb{Q}$.

- The set

  $$\{\frac{1}{n} : n \in \mathbb{N}^+\}$$

  has only one accumulation point, 0, which is not in the set. 0 is an accumulation point because every neighbourhood of 0 must have no upper bound, so something in the form of

  $$(n, \infty)$$

  which contains an infinite number of members of the set.

- Every point in the closed interval $[a, b]$ is an accumulation point of $[a, b]$ (intervals are sets). Every point outside the closed interval is not an accumulation point of $[a, b]$, since there must exist a neighbourhood of the point which lies entirely outside of the interval, and therefore contains no elements of the interval (so does not contain an infinite number of elements in the interval).

- Every point in the open interval (a, b) is an accumulation point of (a, b). The endpoints, a and b, are also accumulation points.

2.2 $\varepsilon - \delta$ definition of limit of a function

Basically the most important thing in this chapter:

Definition 2.3 ($\varepsilon - \delta$ definition of limit of a function)**.** A function $f : A \to \mathbb{R}$, where $A \subseteq \mathbb{R}$, has a limit $l \in \mathbb{R}$ at the accumulation point x_0 of A if and only if

$$\forall \varepsilon > 0,\ \exists \delta > 0 : |x - x_0| < \delta \Rightarrow |f(x) - l| < \varepsilon$$

We denote this as

$$\lim_{x \to x_0} f(x) = l$$

An intuitive translation of the definition reads whenever the input is sufficiently close to the target input, then the output must be within the specified range of the limit.

Example. Prove that

$$\lim_{x \to 2} x^2 + x - 2 = 4$$

Solution. Let $f(x) = x^2 + x - 2$. Given some $\varepsilon > 0$, we want to find some $\delta > 0$ such that

$$|x - 2| < \delta \Rightarrow |f(x) - 4| < \varepsilon.$$

$$\begin{aligned} |f(x) - 4| < \varepsilon &\Leftrightarrow \left|(x^2 + x - 2) - 4\right| < \varepsilon \\ &\Leftrightarrow \left|x^2 + x - 6\right| < \varepsilon \\ &\Leftrightarrow |(x+3)(x-2)| < \varepsilon \\ &\Leftrightarrow |x - 2| < \frac{\varepsilon}{|x+3|} \end{aligned}$$

Suppose, arbitrarily, we let $|x - 2| < 1$. Then $|x + 3| \leq |x - 2| + 5 < 6$. We want the above inequality to be true as well, so we have

$$|x - 2| < \frac{\varepsilon}{6} < \frac{\varepsilon}{|x+3|}$$

Now, we have two constraints that we want to both satisfy:

$$|x - 2| < 1 \text{ and } |x - 2| < \frac{\varepsilon}{6}$$

so we choose $\delta = \min(1, \frac{\varepsilon}{6})$ and both will be satisfied.

The actual proof. Choose $\delta = \min(1, \frac{\varepsilon}{6})$. Then $\delta \leq 1$ and $\delta \leq \frac{\varepsilon}{6}$. Suppose $|x - 2| < \delta$. Then we have

$$|x - 2| < 1 \Rightarrow |x + 3| < 6$$

and

$$|x - 2| < \frac{\varepsilon}{6}$$

$$\begin{aligned}\left|x^2+x-6\right| &= |(x-2)(x+3)| \\ &= |x-2||x+3| \\ &< \frac{\varepsilon}{6}|x+3| \\ &< \frac{\varepsilon}{6}\cdot 6 \\ &= \varepsilon\end{aligned}$$

which completes the proof.

Example. (MMT 1.5, Q1) Prove that

$$\lim_{x\to 0}\frac{1}{x}$$

does not exist.

Solution. Suppose the limit $= l \in \mathbb{R}$. Then either $l > 0$, $l < 0$, or $l = 0$.

$\mathbf{l > 0.}$ Suppose $\varepsilon = l$. There exists $\delta > 0$ such that

$$|x| < \delta \Rightarrow \left|\frac{1}{x} - l\right| < l$$

Choose $-\delta < x < 0$. Clearly the LHS of the implication is true. But

$$\left|\frac{1}{x} - l\right| < l \Leftrightarrow 0 < \frac{1}{x} < 2l$$

so the RHS is false, and so the implication is false, contradiction.

$\mathbf{l < 0.}$ Suppose $\varepsilon = -l$ (since $\varepsilon > 0$). There exists $\delta > 0$ such that

$$|x| < \delta \Rightarrow \left|\frac{1}{x} - l\right| < -l$$

Choose $0 < x < \delta$. Clearly the LHS of the implication is true. But

$$\left|\frac{1}{x} - l\right| < -l \Leftrightarrow 2l < \frac{1}{x} < 0$$

so the RHS is false, and so the implication is false, contradiction.

$\mathbf{l = 0.}$ Suppose $\varepsilon = 1$. There exists $\delta > 0$ such that

$$|x| < \delta \Rightarrow \left|\frac{1}{x}\right| < 1$$

Choose $0 < x < \min(\delta, 1)$. The LHS of the implication is true. But

$$|x| < 1 \Leftrightarrow \left|\frac{1}{x}\right| > 1$$

so the RHS is false, and so the implication is false, contradiction.

In all three cases, we reach a contradiction. So the limit does not exist.

2.2.1 Definitions for alternative use cases

Here, we present a few definitions for some common use cases:

Definition 2.4.

$$\lim_{x \to \infty} f(x) = l$$

if and only if

$$\forall \varepsilon > 0,\ \exists c \in \mathbb{R} : \forall x > c,\ |f(x) - l| < \varepsilon$$

Definition 2.5.

$$\lim_{x \to -\infty} f(x) = l$$

if and only if

$$\forall \varepsilon > 0,\ \exists c \in \mathbb{R} : \forall x < c,\ |f(x) - l| < \varepsilon$$

Definition 2.6.

$$\lim_{x \to x_0} f(x) = \infty$$

if and only if

$$\forall r > 0,\ \exists \delta > 0 : |x - x_0| < \delta \Rightarrow f(x) > r$$

Definition 2.7.

$$\lim_{x \to \infty} f(x) = \infty$$

if and only if

$$\forall r > 0,\ \exists c \in \mathbb{R} : x > c \Rightarrow f(x) > r$$

2.3 Continuity

Definition 2.8. A function f is continuous at x_0 if

$$\lim_{x \to x_0} f(x) = f(x_0)$$

Definition 2.9. A function f is continuous on the interval $[a, b]$ if it is continuous at all $x_0 \in [a, b]$.

Example. Let f be defined by

$$f(x) = \begin{cases} x \sin \frac{1}{x}, & x \neq 0 \\ 0, & x = 0 \end{cases}$$

Prove that f is continuous at 0.

Proof. We want to show that

$$\lim_{x \to 0} f(x) = f(0) = 0$$

$$\begin{aligned}\left|x\sin\frac{1}{x}-0\right|<\varepsilon \Leftrightarrow \left|x\sin\frac{1}{x}\right| &< \varepsilon\\ \Leftrightarrow |x|\left|\sin\frac{1}{x}\right| &< \varepsilon\\ \Leftrightarrow |x| < \frac{\varepsilon}{\left|\sin\frac{1}{x}\right|} &\leq \frac{\varepsilon}{1}=\varepsilon\end{aligned}$$

Let $\delta=\varepsilon$. Then

$$\begin{aligned}|x|<\delta &\Rightarrow |x|<\varepsilon\\ &\Rightarrow |x|\left|\sin\frac{1}{x}\right|<\varepsilon\cdot\left|\sin\frac{1}{x}\right|<\varepsilon\\ &\Rightarrow \left|x\sin\frac{1}{x}\right|<\varepsilon\end{aligned}$$

□

Example. Consider the sign function $sgn:\mathbb{R}\to\mathbb{R}$ defined by

$$sgn(x)=\begin{cases}-1, & x<0\\ 0, & x=0\\ 1, & x>0\end{cases}$$

Prove that sgn is continuous at all points except 0.

Proof. Take arbitrary $x_0<0$. Take arbitrary $\varepsilon>0$. For this ε, we want to find some $\delta>0$ such that for all $x\in\mathbb{R}$,

$$|x-x_0|<\delta\Rightarrow|f(x)-f(x_0)|<\varepsilon$$

Intuition. When we have $x_0<0$, we want to make sure the x that we can choose are close enough to x_0 such that $x<0$, then both $f(x)$ and $f(x_0)$ evaluate to -1.

Let $\delta=\frac{|x_0|}{2}$. Suppose $|x-x_0|<\delta$ (because that's how we prove implications – we assume the antecedent is true, then proceed to show that the consequent is true as well). Then we have

$$\begin{aligned}-\frac{|x_0|}{2}<x-x_0&<\frac{|x_0|}{2}\\ \Leftrightarrow x_0-\frac{|x_0|}{2}<x&<x_0+\frac{|x_0|}{2}\end{aligned}$$

We focus on the upper inequality.

$$\begin{aligned}x&<x_0+\frac{|x_0|}{2}\\ &=x_0-\frac{x_0}{2}\text{ since }x_0<0\\ &=\frac{x_0}{2}\\ &<0\end{aligned}$$

so $f(x) = -1$. Then
$$|f(x) - f(x_0)| = |-1 - (-1)| = 0 < \varepsilon$$
Since $\varepsilon > 0$ arbitrary, f is continuous at x_0. Since $x_0 < 0$ arbitrary, f is continuous at all $x_0 < 0$.

The case where $x_0 > 0$ follows similarly (but is still included here for completeness). Take arbitrary $x_0 > 0$. Take arbitrary $\varepsilon > 0$. Let $\delta = \frac{|x_0|}{2}$. Suppose $|x - x_0| < \delta$. Then we have
$$\begin{aligned}
-\frac{|x_0|}{2} &< x - x_0 < \frac{|x_0|}{2} \\
\Leftrightarrow x_0 - \frac{|x_0|}{2} &< x < x_0 + \frac{|x_0|}{2}
\end{aligned}$$
We focus on the lower inequality.
$$\begin{aligned}
x &> x_0 - \frac{|x_0|}{2} \\
&= x_0 - \frac{x_0}{2} \text{ since } x_0 > 0 \\
&= \frac{x_0}{2} \\
&> 0
\end{aligned}$$
so $f(x) = 1$. Then
$$|f(x) - f(x_0)| = |1 - 1| = 0 < \varepsilon$$
Since $\varepsilon > 0$ arbitrary, f is continuous at x_0. Since $x_0 > 0$ arbitrary, f is continuous at all $x_0 > 0$.

Let $x_0 = 0$. Since we want to show that *sgn* is not continuous at 0, we want to show
$$\exists \varepsilon > 0 : \forall \delta > 0,\ \exists x \in \mathbb{R} : |x - 0| < \delta \wedge |f(x) - f(0)| \geq \varepsilon$$

> **Intuition.** We know that all the points around 0 either evaluate to -1 or 1, so they are of distance 1 away from $f(0) = 0$. We can find points that are infinitesmally close to 0 (i.e. satisfy $|x - x_0| < \delta$), but are still of distance 1 away from $f(0)$. So we pick some $\varepsilon < 1$, then the distance must be greater than ε.

> **Remark.** It doesn't really matter if we use *take arbitrary* or *let* to show *for all*, but things can get messy if we want to show *there exists* and *for all* at the same time, so it's best to distinguish them. Also, we can't use *take arbitrary* to show *there exists*.

Let $\varepsilon = \frac{1}{2}$ (existence). Take arbitrary $\delta > 0$ (for all). Let $x = \frac{\delta}{2}$ (existence). Then
$$\begin{aligned}
|x| &= \left|\frac{\delta}{2}\right| \\
&= \frac{\delta}{2} \text{ since } \delta > 0 \\
&< \delta
\end{aligned}$$
Since $x = \frac{\delta}{2}$, we know that $x > 0$, so $f(x) = 1$. But then
$$|f(x) - f(0)| = |1 - 0| = 1 > \varepsilon$$
and we are done. $\square$

2.3.1 Bolzano-Weierstrass theorem

Definition 2.10 (Bolzano-Weierstrass theorem)**.** Every bounded sequence has a convergent subsequence.

We will see how the theorem can be used to prove several properties of continuous functions.

Proposition 2.11. If $f : [a, b] \to \mathbb{R}$ is a continuous function, then there exist $r, s \in [a, b]$ such that

$$f(r) = \sup\{f(x) : x \in [a, b]\} \in \mathbb{R}$$
$$f(s) = \inf\{f(x) : x \in [a, b]\} \in \mathbb{R}$$

In other words, we want to show two things:

- f attains its maximum and minimum in $[a, b]$ in $[a, b]$ (confusing, yes)
- The maximum and minimum are finite

Proof. We will only show the first half of both statements, i.e. the parts for maximum. The parts for minimum follow similarly. Let $M = \sup\{f(x) : x \in [a, b]\}$.

By definition of supremum, for all $n \geq 1$, there exists x_n such that

$$M \geq f(x_n) \geq M - \frac{1}{n}$$

> **Remark.** Suppose there exists $n \geq 1$ such that for all $x \in [a, b]$, the inequality does not hold. Then either $f(x_n) > M$ or $f(x_n) < M - \frac{1}{n}$. For the former, this violates the fact that M is the supremum, and hence an upper bound. For the latter, this also violates the fact that M is the supremum, since we now have a smaller upper bound (namely, $M - \frac{1}{n}$) and so M is not the least upper bound.

The sequence x_n is bounded (by $[a, b]$), so by Bolzano-Weierstrass, there exists a convergent subsequence x_{n_i}.

> **Remark.** x_n is not necessarily convergent.

Suppose $x_{n_i} \to v_M$. Since $a \leq x_{n_i} \leq b \; \forall i$, we have $v_M \in [a, b]$. By continuity of f, we have

$$\lim_{i \to \infty} f(x_{n_i}) = f(v_M)$$

But then $f(x_{n_i})$ is a subsequence of $f(x_n)$, so they converge to the same limit. By sandwich theorem, $f(x_n) \to M$, so also $f(x_{n_i}) \to M$. Therefore,

$$f(v_M) = \lim_{i \to \infty} f(x_{n_i}) = M$$

Suppose M is not finite, i.e. f is not bounded above. Then for all $n \geq 1$, there exists $x_n \in [a, b]$ such that $f(x_n) > n$. Since x_n is taken from $[a, b]$, x_n is bounded. By Bolzano-Weierstrass, there exists a convergent subsequence x_{n_i}.

Suppose $x_{n_i} \to x_0$. Since $a \leq x_{n_i} \leq b$ for all i, $x_0 \in [a, b]$, so f is continuous at x_0. Therefore, we have

$$\lim_{i \to \infty} f(x_{n_i}) = f(x_0)$$

which is finite (??????). However, since $f(x_{n_i})$ is a subsequence of $f(x_n)$, they tend to the same limit. Since (the rest of the proof seems wrong.) □

2.3.2 Intermediate value theorem

Theorem 2.12 (Intermediate value theorem)**.** Consider a continuous function $f : [a, b] \to \mathbb{R}$. If s is between $f(a)$ and $f(b)$, i.e.

$$\min\{f(a), f(b)\} < s < \max\{f(a), f(b)\}$$

then there exists $c \in (a, b)$ such that $f(c) = s$.

Proof. Suppose $f(a) < s < f(b)$. The case where $f(b) < s < f(a)$ can be handled similarly. Consider the set

$$S = \{x \in [a, b] : f(x) \leq s\}$$

Since S is bounded above and non-empty, it must have a supremum, say c.

(Why must $c \in [a, b]$?)

Since $c \in [a, b]$, f is continuous at c. By continuity, we have

$$\forall \varepsilon > 0,\ \exists \delta > 0 : |x - c| < \delta \Rightarrow |f(x) - f(c)| < \varepsilon$$

Take arbitrary $\varepsilon > 0$. Since c is the supremum, there must exist some $x_1 \in (c-\delta, c]$ such that $x_1 \in S$.

> **Remark.** Otherwise, this means that the largest element is some finite distance δ smaller than c, so the supremum would be at most $c - \delta$, but not c.

Since $x_1 \in S$, we have $f(x_1) \leq s$. As

$$|f(x) - f(c)| < \varepsilon \Leftrightarrow f(c) - \varepsilon < f(x) < f(c) + \varepsilon$$

we get

$$\begin{aligned} & f(c) - \varepsilon < f(x_1) < f(c) + \varepsilon \\ & \Rightarrow f(c) - \varepsilon < f(x_1) \leq s \\ & \Rightarrow f(c) - \varepsilon < s \\ & \Rightarrow f(c) < s + \varepsilon \end{aligned} \tag{2.1}$$

Since c is the supremum, for all $x \in (c, c + \delta)$, $x \notin S$.

> **Remark.** Otherwise, we have an element greater than c, so c is not an upper bound.

But then either $f(x) \leq s$ or $f(x) > s$. So there exists $x_2 \in (c, c + \delta)$ such that $f(x_2) > s$. Then

$$\begin{aligned} & f(c) - \varepsilon < f(x_2) < f(c) + \varepsilon \\ & \Rightarrow s < f(x_2) < f(c) + \varepsilon \\ & \Rightarrow s < f(c) + \varepsilon \\ & \Rightarrow s - \varepsilon < f(c) \end{aligned} \tag{2.2}$$

Combining Equation (2.1) and Equation (2.2), we get

$$s - \varepsilon < f(c) < s + \varepsilon$$

Since $\varepsilon > 0$ arbitrary, this is true for all $\varepsilon > 0$. So we have $f(c) = s$. □

2.3.3 Uniform continuity

Definition 2.13 (Uniform continuity). A function $f : A \to \mathbb{R}$ is uniformly continuous if

$$\forall \varepsilon > 0,\ \exists \delta > 0 : \forall x, y \in A,\ |x - y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon$$

How is this different from the definition of continuity? The difference is that for continuity,

$$\forall \varepsilon > 0\ \forall y \in A,\ \exists \delta > 0 : \forall x \in A,\ |x - y| \Rightarrow |f(x) - f(y)| < \varepsilon$$

the value of δ may depend on both ε and the accumulation point y, whereas in uniform continuity, the value of δ only depends on ε and is independent of the choice of y. We illustrate the difference in the following two examples.

Example. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^2$. Show that f is not uniformly continuous.

Proof. We want to show that

$$\exists \varepsilon > 0 : \forall \delta > 0,\ \exists x, y \in \mathbb{R} : |x - y| < \delta \wedge |f(x) - f(y)| \geq \varepsilon$$

Let $\varepsilon = 1$. Take arbitrary $\delta > 0$. How might we find the right x, y? Firstly, we want

$$|x - y| < \delta$$

so we can choose $y = x + \frac{\delta}{2}$, then the difference will be $\frac{\delta}{2} < \delta$. Then we try to expand $f(x) - f(y)$:

$$\begin{aligned} |f(x) - f(y)| &= \left| x^2 - (x + \frac{\delta}{2})^2 \right| \\ &= \left| (x + \frac{\delta}{2})^2 - x^2 \right| \\ &= \left| x^2 + \delta x + \frac{\delta^2}{4} - x^2 \right| \\ &= \left| \delta x + \frac{\delta^2}{4} \right| \\ &= \delta(x + \frac{\delta}{4}) \end{aligned}$$

With this, we can choose $x = \frac{1}{\delta}$ so that the expanded product becomes $1+$ something positive $> \varepsilon$.

Let $x = \frac{1}{\delta} + \frac{\delta}{2}$, $y = \frac{1}{\delta}$ (x and y are swapped from the above selection for easier manipulation.) Then

$$|x - y| = \left| \frac{1}{\delta} + \frac{\delta}{2} - \frac{1}{\delta} \right| = \frac{\delta}{2} < \delta$$

Also,

$$\begin{aligned}|f(x)-f(y)| &= \left|(\frac{1}{\delta}+\frac{\delta}{2})^2-(\frac{1}{\delta})^2\right| \\ &= \left|\frac{1}{\delta^2}+1+\frac{\delta^2}{4}-\frac{1}{\delta^2}\right| \\ &= \left|1+\frac{\delta^2}{4}\right| \\ &= 1+\frac{\delta^2}{4} \\ &> 1 \\ &= \varepsilon\end{aligned}$$

Since $\delta > 0$ arbitrary, we are done. □

Example. Let $f : [-M, M] \to \mathbb{R}$ be defined by $f(x) = x^2$. Show that f is uniformly continuous.

Proof. We want to show that

$$\forall\varepsilon > 0,\ \exists\delta > 0 : \forall x, y \in [-M, M],\ |x-y| < \delta \Rightarrow |f(x)-f(y)| < \varepsilon$$

Take arbitrary $\varepsilon > 0$, and take arbitrary $x, y \in [-M, M]$. For $|f(x) - f(y)| < \varepsilon$ to hold for this ε, we want

$$\begin{aligned}|f(x)-f(y)| < \varepsilon &\Leftrightarrow \left|x^2-y^2\right| < \varepsilon \\ &\Leftrightarrow |(x+y)(x-y)| < \varepsilon \\ &\Leftrightarrow |x+y||x-y| < \varepsilon \\ &\Leftrightarrow |x-y| < \frac{\varepsilon}{|x+y|} \leq \frac{\varepsilon}{2M}\end{aligned}$$

since $-2M < x+y < 2M$. Then we let $\delta = \frac{\varepsilon}{2M}$. We are basically done here because we have just found a δ that only depends on ε, but not x or y, but we will proceed with the definition just to be cautious.

$$\begin{aligned}|x-y| < \delta \Rightarrow |x-y| &< \frac{\varepsilon}{2M} \\ \Leftrightarrow |x-y||x+y| &< \frac{\varepsilon}{2M}|x+y| \\ \Leftrightarrow \left|x^2-y^2\right| &< \frac{\varepsilon}{2M}|x+y| \leq \frac{\varepsilon}{2M}\cdot 2M = \varepsilon\end{aligned}$$

□

Now, we will show a few (useful?) properties involving uniform continuity.

Proposition 2.14. If $f : A \to \mathbb{R}$ is uniformly continuous, then it is continuous on A.

Proof. We want to show that

$$\forall\varepsilon > 0\ \forall x_0 \in A,\ \exists\delta > 0 : \forall x \in A,\ |x-x_0| < \delta \Rightarrow |f(x)-f(x_0)| < \varepsilon$$

Assume f is uniformly continuous. Then we have

$$\forall \varepsilon > 0,\ \exists \delta > 0 : \forall x, y \in A,\ |x - y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon$$

Take arbitrary $\varepsilon > 0$ and take arbitrary $x_0 \in A$. By assumption, we know that

$$\exists \delta > 0 : \forall x \in A,\ |x - x_0| < \delta \Rightarrow |f(x) - f(x_0)| < \varepsilon$$

Since $\varepsilon > 0$ arbitrary, f is continuous at x_0. Since $x_0 \in A$ arbitrary, f is continuous on A.

> **Remark.** In this proof, it is important to be precise in what is arbitrary and what is not, and the implications of generalising an arbitrary variable to all cases.

$\square$

Proposition 2.15 (Characterisation of uniform continuity via sequences)**.** Let $f : A \to \mathbb{R}$. f is uniformly continuous *if and only if* for any pair of sequences $(x_n)_{n\geq 1}, (y_n)_{n\geq 1} \in A$ with $x_n - y_n \to 0$, we have $f(x_n) - f(y_n) \to 0$.

Proof. $\Rightarrow$ Suppose f is uniformly continuous. Then

$$\forall \varepsilon > 0,\ \exists \delta > 0 : \forall x, y \in A,\ |x - y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon$$

Take arbitrary $\varepsilon > 0$ and let $\delta > 0$ satisfy the above for this ε. Since $x_n - y_n \to 0$, we know that

$$\exists N \in N : n > N \Rightarrow |x_n - y_n| < \delta$$

But then

$$\exists N \in N : n > N \Rightarrow |f(x_n) - f(y_n)| < \varepsilon$$

so $f(x_n) - f(y_n) \to 0$.

$\Leftarrow$. Suppose for any pair of sequences $x_n, y_n \in A$ with $x_n - y_n \to 0$, we have $f(x_n) - f(y_n) \to 0$. Assume that f is not uniformly continuous. Then this means

$$\exists \varepsilon > 0 : \forall \delta > 0,\ \exists x, y \in A : |x - y| < \delta \wedge |f(x) - f(y)| \geq \varepsilon$$

> **Intuition.** We are assuming towards a contradiction. One possible way, which is how this proof will follow, to reach a contradiction is to contradict the initial implication assumption, i.e. $x_n - y_n \to 0$ but $f(x_n) - f(y_n) \not\to 0$.
>
> The definition of uniform continuity (or rather, its negation), as it is presented now, has nothing to do with sequences. What do we do? We construct δ specifically such that it forms a sequence that converges to 0, and so by the sandwich theorem, $x_n - y_n \to 0$. Now we have introduced n into the inequality. But then $\varepsilon > 0$, so $f(x_n) - f(y_n) \not\to 0$.

> **Remark.** When in doubt, set $\delta = \frac{1}{n}$.

Then we have

$$\forall n \in \mathbb{N},\ \exists x_n, y_n \in A : |x_n - y_n| < \frac{1}{n} \wedge |f(x_n) - f(y_n)| \geq \varepsilon$$

Since

$$|x_n - y_n| < \frac{1}{n} \; \forall n \in \mathbb{N}$$

we have $x_n - y_n \to 0$. But then for every n,

$$f(x_n) - f(y_n) \geq \varepsilon > 0$$

so $f(x_n) - f(y_n) \not\to 0$, which contradicts our initial assumption. So f must be uniformly continuous. □

This characterisation is particularly useful when we want to show that some function is *not* uniformly continuous, i.e. we find some pair of sequences x_n and y_n such that $x_n - y_n \to 0$ but $f(x_n) - f(y_n) \not\to 0$.

Example. Show that $f : (0, 1) \to \mathbb{R}$ defined by

$$f(x) = \frac{1}{x}$$

is not uniformly continuous.

Solution. Let $x_n = \frac{1}{n}$ and $y_n = \frac{1}{2n}$. We know that

$$x_n - y_n = \frac{1}{2n} \to 0$$

but

$$f(x_n) - f(y_n) = n - 2n = -n \to -\infty$$

so by Proposition 2.15, f is not uniformly continuous.

Theorem 2.16. Every continuous function $f : [a, b] \to \mathbb{R}$ is uniformly continuous.

Proof. Suppose f is not uniformly continuous. Then we have

$$\exists \varepsilon > 0 : \forall \delta > 0, \; \exists x, y \in [a, b] : |x - y| < \delta \wedge |f(x) - f(y)| \geq \varepsilon$$

which implies (again, the same trick)

$$\exists \varepsilon > 0 : \forall n \geq 1, \; \exists x_n, y_n \in [a, b] : |x_n - y_n| < \frac{1}{n} \wedge |f(x_n) - f(y_n)| \geq \varepsilon$$

Now we have two sequences x_n and y_n to work with. From the above, we get

$$-\frac{1}{n} < x_n - y_n < \frac{1}{n}$$

and so by the sandwich theorem, we get $x_n - y_n \to 0$.

Since x_n is bounded, by Bolzano-Weierstrass, there exists a convergent subsequence x_{n_i}. Suppose $x_{n_i} \to x_0$. Since $a \leq x_n \leq b$ for all $n \geq 1$, we know that the limit $x_0 \in [a, b]$. Since f is continuous on $[a, b]$, it is continuous on every point in the interval, and so it is continuous on x_0. By definition of continuity, we have $f(x_{n_i}) \to f(x_0)$.

Similarly, since y_n is bounded, there exists a convergent subsequence y_{n_i}. As $x_n - y_n \to 0$, we get $x_{n_i} - y_{n_i} \to 0$ as well.

Remark. I don't know how sound it is to directly jump to this conclusion, but loosely speaking we have

$$x_n - y_n \to 0 \equiv \lim_{n\to\infty} x_n - y_n = 0$$

Since $n_i \to \infty$ as $i \to \infty$, we have

$$\lim_{i\to\infty} x_{n_i} - y_{n_i} = 0$$

Then by the linearity of limits we get $y_{n_i} \to x_0$. By continuity, we have

$$f(y_{n_i}) \to f(x_0)$$

and again, by the linearity of limits we get

$$f(x_{n_i}) - f(y_{n_i}) \to f(x_0) - f(x_0) = 0$$

However, from our initial assumption, $f(x_n) - f(y_n) \geq \varepsilon > 0 \ \forall n \geq 1$, so $f(x_n) - f(y_n) \not\to 0$. We have reached a contradiction, so f must be uniformly continuous. □

Observe that in the theorem we have just proved, f is continuous on a *closed* interval $[a, b]$.

- If f were continuous on an *open* interval, uniform continuity is not guaranteed (e.g. $f : (0, 1) \to \mathbb{R}$ with $f(x) = \frac{1}{x}$).
- If f were bounded (with regards to its outputs) and continuous, uniform continuity is also not guaranteed (e.g. $f(x) = \cos(x^2)$).

2.4 Exercises

1. Prove the following limits using the $\varepsilon - \delta$ definition.
   a. $\lim_{x\to 4} x = 4$
   b. $\lim_{x\to\infty} \sqrt{x} = \infty$
   c. $\lim_{x\to 0} x(\cos x)(x^2 + 1) = 0$
   d. $\lim_{x\to 8} \sqrt[3]{x} = 2$
   e. $\lim_{x\to 1} x^3 - 2x = -1$
2. Using the definition of uniform continuity, show that the following functions are not uniformly continuous.
   a. $f : (0, 1] \to \mathbb{R}$ with $f(x) = \frac{1}{x}$
   b. $f(x) = x^2 \ \forall x \in \mathbb{R}$
   c. $f(x) = \cos(x^2) \ \forall x \in \mathbb{R}$
3. Using Proposition 2.15, show that the following functions are not uniformly continuous.
   a. $f(x) = x^2 \ \forall x \in \mathbb{R}$
   b. $f(x) = \cos(x^2) \ \forall x \in \mathbb{R}$

Proof. We want to show that

$$\exists \varepsilon > 0 : \forall \delta > 0,\ \exists x, y \in (0,1] : |x-y| < \delta \wedge |f(x) - f(y)| < \varepsilon$$

Let $\varepsilon = 1$. Take arbitrary δ such that $0 < \delta < 1$ (we will deal with $\delta \geq 1$ later). Since we want

$$|x-y| < \delta$$

we choose $y = x + \frac{\delta}{2}$. Then we expand:

$$\begin{aligned}
|f(x) - f(y)| \geq \varepsilon &= \left| \frac{1}{x} - \frac{1}{x + \frac{\delta}{2}} \right| \\
&= \left| \frac{1}{x} - \frac{2}{2x + \delta} \right| \\
&= \left| \frac{2x + \delta - 2x}{x(2x+\delta)} \right| \\
&= \frac{\delta}{x(2x+\delta)}
\end{aligned}$$

We want this to $\geq \varepsilon$:

$$\begin{aligned}
\frac{\delta}{x(2x+\delta)} \geq \varepsilon &\Leftrightarrow \frac{\delta}{x(2x+\delta)} \geq 1 \\
&\Leftrightarrow \delta \geq x(2x+\delta) \\
&\Leftrightarrow 2x^2 + \delta x - \delta \leq 0 \\
&\Leftrightarrow 2(x + \frac{\delta}{4}) - \delta - \frac{\delta^2}{8} \leq 0
\end{aligned}$$

So we choose $x = \frac{\delta}{4}$, $y = \frac{\delta}{4} + \frac{\delta}{2} = \frac{3\delta}{4}$. Since $0 < \delta < 1$, we know that $x, y \in (0,1]$. Then

$$|x-y| = \left| \frac{\delta}{4} - \frac{3\delta}{4} \right| = \left| -\frac{\delta}{2} \right| = \frac{\delta}{2} < \delta$$

Also,

$$\begin{aligned}
|f(x) - f(y)| &= \left| \frac{4}{\delta} - \frac{4}{3\delta} \right| \\
&= \left| \frac{8\delta}{3\delta} \right| \\
&= \frac{8}{3} \\
&> 1 \\
&= \varepsilon
\end{aligned}$$

so we are done for all $\delta \in (0,1)$.

What about $\delta \geq 1$? We simply take $\varepsilon = 1$ (as before), $x = \frac{1}{4}$ and $y = \frac{3}{4}$. Then

$$|x-y| = \left| \frac{1}{4} - \frac{3}{4} \right| = \frac{1}{2} < 1 \leq \delta$$

and the other half of the proof follows as above. $\square$

Chapter 3

Integration

In this chapter, we will introduce the Darboux and Riemann integrals and prove their equivalence. In both definitions of the integral, we consider the *limit* of the sum of the areas of (finitely many) rectangles under (or above) the curve. The notion of *limit* and how we calculate the area of each rectangle differentiate these two definitions.

3.1 Partitions

Definition 3.1 (Partition). A partition P of $[a, b]$ is given by a finite set of endpoints

$$a = x_0 < x_1 < x_2 < \cdots < x_n = b$$

By abuse of notation, we either denote P by its endpoints

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

or by its subintervals

$$P = \{I_0, I_1, \ldots, I_{n-1}\}$$

where $I_k = [x_k, x_{k+1}]$. Observe that there are $n + 1$ endpoints but n intervals.

The endpoints do not have to be evenly distributed across the interval. Therefore, we define the notion of *mesh*:

Definition 3.2 (Mesh). The mesh of a partition P is the width of the widest subinterval in P, that is,

$$\|P\| = \max\{x_{i+1} - x_i,\ i \in [0, n-1]\}$$

Another often useful notion is *refinement*:

Definition 3.3 (Refinement). Given two partitions P and Q of $[a, b]$, we say that Q refines P, or Q is a refinement of P, if $P \subseteq Q$.

Remark. Several notes on refinements:

- Any partition is a refinement of itself.
- $\|Q\| \leq \|P\|$ does not imply Q refines P. For Q to refine P, we demand that Q contains all of the

endpoints in P, and possibly some more. Consider the following partitions of $[0,6]$:

$$P = \{0, 3, 6\}$$
$$Q = \{0, 2, 4, 6\}$$

Even though Q is intuitively the "finer" partition ($\|Q\| = 2$ but $\|P\| = 3$), Q does not refine P as $3 \in P$ but $3 \notin Q$.

3.2 Darboux integral

3.2.1 Lower, upper darboux sums

Before we define the notion of lower and upper Darboux sums, we shall introduce two synonyms for ease of reading. Given a partition

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

and a bounded function $f : [a,b] \to \mathbb{R}$, let

$$M_i = \sup_{x \in [x_i, x_{i+1}]} f(x)$$
$$m_i = \inf_{x \in [x_i, x_{i+1}]} f(x)$$

or equivalently, by our abuse of notation,

$$M_i = \sup_{x \in I_i} f(x)$$
$$m_i = \inf_{x \in I_i} f(x)$$

Other literature may define them by taking the supremum or infimum over $[x_{i-1}, x_i]$ instead, so pay attention to the given definitions to avoid off-by-one errors. Using these synonyms, we now define lower and upper Darboux sums.

Definition 3.4 (Lower and upper Darboux sum). Let $f : [a,b] \to \mathbb{R}$ be a bounded function. Given a partition P of $[a,b]$, the lower Darboux sum of f with respect to P is

$$L(f, P) = \sum_{i=0}^{n-1} m_i (x_{i+1} - x_i)$$

and the upper Darboux sum of f with respect to P is

$$U(f, P) = \sum_{i=0}^{n-1} M_i (x_{i+1} - x_i)$$

We may often omit Darboux and simply say the *lower sum* or the *upper sum*, as these definitions are equivalent to lower and upper Riemann sums, which will be introduced in the next section.

We will derive several important properties involving lower and upper sums, partitions, and refinements. We will assume that $f : [a,b] \to \mathbb{R}$ is bounded.

Proposition 3.5. For any partition P of $[a,b]$,

$$L(f, P) \leq U(f, P)$$

Proof. By definition, for any $x \in I_i$,

$$m_i \leq f(x) \leq M_i$$

so for every $i \in [0, n-1]$,

$$m_i \leq M_i$$

and the result follows by definition of $L(f, P)$ and $U(f, P)$. $\square$

Proposition 3.6. Let P and Q be partitions of $[a, b]$, and suppose Q refines P. Then

$$L(f, P) \leq L(f, Q)$$
$$U(f, P) \geq U(f, Q)$$

Proof. The case where $P = Q$ is trivial. Suppose Q has one more endpoint x_j than P. (Since Q refines P, Q must contain all endpoints of P, and possibly some more.) Let $x_j \in [x_k, x_{k+1}]$ for some $k \in [0, n-1]$. (Since Q is a partition of $[a, b]$, $x_j \in [a, b]$.) Then the lower and upper sums differ at the term on the interval $[x_k, x_{k+1}]$. Observe that

$$\begin{aligned} m_k(x_{k+1} - x_k) &= m_k(x_{k+1} - x_j) + m_k(x_j - x_k) \\ &\leq \inf_{x \in [x_j, x_{k+1}]} f(x)(x_{k+1} - x_j) + \inf_{x \in [x_k, x_j]} f(x)(x_j - x_k) \end{aligned}$$

Intuitively, the infimum may fall into the left half ($[x_k, x_j]$) or the right half ($[x_j, x_{k+1}]$) of the interval. When we split the interval into two, the infimum on either half may take a larger value, since the original infimum no longer belongs to this half. Of course, the original infimum may be attained on the common endpoint of the two halves, i.e. x_j, in which case the new infima are unchanged. Similarly, we have

$$\begin{aligned} M_k(x_{k+1} - x_k) &= M_k(x_{k+1} - x_j) + M_k(x_j - x_k) \\ &\geq \sup_{x \in [x_j, x_{k+1}]} f(x)(x_{k+1} - x_j) + \sup_{x \in [x_k, x_j]} f(x)(x_j - x_k) \end{aligned}$$

Therefore,

$$\begin{aligned} L(f, P) &= \sum_{i=0}^{n-1} m_i(x_{i+1} - x_i) \\ &= \sum_{i=0}^{k-1} m_i(x_{i+1} - x_i) + m_k(x_{k+1} - x_k) + \sum_{i=k+1}^{n-1} m_i(x_{i+1} - x_i) \\ &= \sum_{i=0}^{k-1} m_i(x_{i+1} - x_i) + m_k(x_{k+1} - x_j) + m_k(x_j - x_k) + \sum_{i=k+1}^{n-1} m_i(x_{i+1} - x_i) \\ &\leq \sum_{i=0}^{k-1} m_i(x_{i+1} - x_i) + \inf_{x \in [x_j, x_{k+1}]} f(x)(x_{k+1} - x_j) + \inf_{x \in [x_k, x_j]} f(x)(x_j - x_k) + \sum_{i=k+1}^{n-1} m_i(x_{i+1} - x_i) \\ &= L(f, Q) \end{aligned}$$

and the inequality for the upper sum follows exactly the same way. If Q has n more endpoints than P, we add one endpoint at a time to get

$$P \subseteq P_1 \subseteq P_2 \subseteq \ldots \subseteq P_n = Q$$

where P_m has m more endpoints than P. Then the inequality follows:

$$L(f,P) \leq L(f,P_1) \leq \ldots \leq L(f,P_n) = L(f,Q)$$
$$U(f,P) \leq U(f,P_1) \leq \ldots \leq U(f,P_n) = U(f,Q)$$

□

We can interpret the result above intuitively.

Intuition. The lower sum should be strictly less than the area under the curve, while the upper sum should be strictly greater than the area under the curve. When we refine the partition, the lower and upper sums should both get closer and closer to the true area. Therefore, the lower sums should form a monotonically increasing (non-decreasing) sequence, while the upper sums should form a monotonically decreasing (non-increasing) sequence, as the partitions get finer and finer.

In fact, we can generalise Proposition 3.5 using Proposition 3.6.

Proposition 3.7. For any partition P and Q of $[a,b]$,

$$L(f,P) \leq U(f,Q)$$

Proof. Observe that $P \cup Q$ is a common refinement of P and Q. In other words, $P \cup Q$ refines both P and Q. By Proposition 3.5 and Proposition 3.6,

$$L(f,P) \leq L(f,P\cup Q) \leq U(f,P\cup Q) \leq U(f,Q)$$

□

3.2.2 Definition of the integral

The lower and upper Darboux integrals formalise the notion of the *limit* of lower and upper Darboux sums, which stems from the intuition that lower sums form an increasing sequence and upper sums form a decreasing sequence.

Definition 3.8 (Lower and upper Darboux integral)**.** The lower integral is given by

$$L(f) = \sup\{L(f,P) : P \text{ is a partition of } [a,b]\}$$

and the upper integral is given by

$$U(f) = \inf\{U(f,P) : P \text{ is a partition of } [a,b]\}$$

In some literature, we draw a bar below or above the definite integral to denote the lower or upper integral:

$$\underline{\int_a^b} f(x)\mathrm{d}x = L(f) \qquad \text{and} \qquad \overline{\int_a^b} f(x)\mathrm{d}x = U(f)$$

Before we define the Darboux integral, we prove one property of lower and upper integrals. This requires the following lemma:

Lemma 3.9. Let $f,g : [a,b] \to \mathbb{R}$ be functions. If $f(x) \leq g(y)$ for every $x,y \in [a,b]$, then $\sup f \leq \inf g$.

Note that the condition is different from saying $f \leq g$, i.e. $f(x) \leq g(x)$ for every $x \in [a,b]$, which does not imply the result in general.

Proof. Fix $\varepsilon > 0$. By definition of supremum and infimum, there exists $x, y \in [a, b]$ such that

$$\sup f - \frac{\varepsilon}{2} < f(x) < \sup f \qquad \text{and} \qquad \inf g < g(y) < \inf g + \frac{\varepsilon}{2}$$

Since $f(x) \leq g(y)$, it follows that

$$\sup f - \frac{\varepsilon}{2} < f(x) \leq g(y) < \inf g + \frac{\varepsilon}{2}$$

and therefore,

$$\sup f < \inf g + \varepsilon$$

Since this holds for any $\varepsilon > 0$, we have $\sup f \leq \inf g$. (We cannot conclude that $\sup f < \inf g$ since it may be the case that $f = g$ and the inequality will still hold.) □

Proposition 3.10. $L(f) \leq U(f)$.

Proof. From Proposition 3.7, we know that $L(f, P) \leq U(f, Q)$ for any partition P and Q of $[a, b]$. By lemma and definition of lower and upper integrals, the result follows. □

Definition 3.11 (Darboux integral)**.** A function $f : [a, b] \to \mathbb{R}$ is Darboux integrable if and only if it is bounded and $L(f) = U(f)$ and the common value is called the Darboux integral:

$$\int_a^b f(x)\mathrm{d}x = L(f) = U(f)$$

If f is not bounded, then either the lower or upper Darboux sums (or both) will be unbounded, so f will not be Darboux integrable.

3.2.3 Cauchy criterion

Instead of using the definition of the Darboux integral, it is sometimes more convenient to use the Cauchy criterion to show that a function is Darboux integrable. We show their equivalence:

Definition 3.12 (Cauchy criterion)**.** A function $f : [a, b] \to \mathbb{R}$ is Darboux integrable if and only if for every $\varepsilon > 0$, there exists a partition P of $[a, b]$ such that

$$U(f, P) - L(f, P) < \varepsilon$$

Proof. **Only if.** Suppose a function f is Darboux integrable. Then by definition,

$$U(f) = L(f) \qquad \Leftrightarrow \qquad \inf\{U(f, P)\} = \sup\{L(f, P)\}$$

Fix $\varepsilon > 0$. By definition of infimum and supremum, there exists partitions P and Q such that

$$U(f) < U(f, P) < U(f) + \frac{\varepsilon}{2} \qquad \text{and} \qquad L(f) - \frac{\varepsilon}{2} < L(f, Q) < L(f)$$

Consider $P \cup Q$. It follows from Proposition 3.6 that

$$U(f, P \cup Q) < U(f, P) < U(f) + \frac{\varepsilon}{2}$$
$$L(f, P \cup Q) > L(f, Q) > L(f) - \frac{\varepsilon}{2}$$

Since $L(f, P \cup Q) \leq U(f, P \cup Q)$, it follows that

$$\begin{aligned} U(f, P \cup Q) - L(f, P \cup Q) &< U(f) + \frac{\varepsilon}{2} - L(f) + \frac{\varepsilon}{2} \\ &= U(f) - L(f) + \varepsilon \\ &= \varepsilon && \text{since U(f) = L(f)} \end{aligned}$$

If. Fix $\varepsilon > 0$. Then there exists a partition P such that

$$U(f, P) - L(f, P) < \varepsilon$$

By definition, we know that

$$U(f) \leq U(f, P) \qquad \text{and} \qquad L(f) \geq L(f, P)$$

so

$$U(f) - L(f) \leq U(f, P) - L(f, P) < \varepsilon$$

From Proposition 3.10, we also know that

$$L(f) \leq U(f) \qquad \Leftrightarrow \qquad U(f) - L(f) \geq 0$$

Combining these inequalities gives

$$0 \leq U(f) - L(f) < \varepsilon$$

This holds for any $\varepsilon > 0$, so it follows that $U(f) = L(f)$. □

The following example illustrates how the Cauchy criterion can be easier to use than the definition of the Darboux integral.

Proposition 3.13. All continuous functions are Darboux integrable.

Proof. Take any continuous function $f : [a, b] \to \mathbb{R}$. Then f is uniformly continuous. So we have

$$\forall \varepsilon > 0,\ \exists \delta > 0 : \forall x, y \in [a, b],\ |x - y| < \delta \Rightarrow |f(x) - f(y)| < \frac{\varepsilon}{b - a}$$

Take any partition P of $[a, b]$ such that $\|P\| < \delta$. (Such a partition always exists since we can always subdivide $[a, b]$ into, say, roughly $(b - a)/\delta$ equal subintervals.) Then the distance between any two points in any interval will be less than δ. In particular, consider the subinterval $[x_i, x_{i+1}]$ for some i, and suppose f attains its maximum and minimum on $[x_i, x_{i+1}]$ at points c and d respectively. Since $|c - d| < \delta$, it follows that

$$|M_i - m_i| < \frac{\varepsilon}{b - a}$$

Observe that $M_i \geq m_i$, so $|M_i - m_i| = M_i - m_i$. Therefore,

$$\begin{aligned} U(f,P) - L(f,P) &= \sum M_i(x_{i+1} - x_i) + \sum m_i(x_{i+1} - x_i) \\ &= \sum (M_i - m_i)(x_{i+1} - x_i) \\ &< \frac{\varepsilon}{b-a}(x_{i+1} - x_i) \\ &= \frac{\varepsilon}{b-a} \cdot (b-a) \\ &= \varepsilon \end{aligned}$$

□

3.3 Riemann integral

3.3.1 Tagged partition

When we evaluate Darboux sums, we take the maximum and minimum values on each subinterval without caring where the maximum and minimum are attained. With Riemann sums, however, we specify at which point on each subinterval the function should be evaluated. These points are the *tags* associated with the partition. Formally,

Definition 3.14. A tagged partition (P, C) of $[a, b]$ is a partition

$$P = \{I_0, I_1, \ldots, I_{n-1}\}$$

together with a set of tags

$$C = \{c_0, c_1, \ldots, c_{n-1}\}$$

such that $c_k \in I_k$ for every $k \in [0, n-1]$.

3.3.2 Riemann sum

The Riemann sum is defined similarly to the Darboux sums.

Definition 3.15 (Riemann sum)**.** The Riemann sum of a function $f : [a, b] \to \mathbb{R}$ with respect to the tagged partition (P, C) is defined by

$$S(f, P, C) = \sum_{i=0}^{n-1} f(c_i)(x_{i+1} - x_i)$$

Clearly, we may choose the tags such that the function attains its maximum, minimum, or somewhere in between on each subinterval. Therefore,

$$L(f, P) \leq S(f, P, C) \leq U(f, P)$$

3.3.3 Riemann integral

The reason why we don't see this definition as frequently, or why many people define the Riemann integral using the definition of the Darboux integral, is because Darboux and Riemann integrals are equivalent, and that the Darboux integral is arguably easier to use.

Definition 3.16. A function $f : [a, b] \to \mathbb{R}$ is Riemann integrable if there exists $R \in \mathbb{R}$ such that for every $\varepsilon > 0$, there exists $\delta > 0$ such that for every partition P with $\|P\| < \delta$,

$$|S(f, P, C) - R| < \varepsilon$$

If such R exists, R is the Riemann integral of f.

Proposition 3.17. A function is Darboux integrable if and only if it is Riemann integrable.

There will be a lot of seemingly convoluted choices of ε in the following proof. We will first sketch out the proof below simply using ε in any definition we encounter, and then backtrack to find cleaner versions of ε.

Proof. Riemann implies Darboux. Fix $\varepsilon > 0$. Let P be a partition of $[a, b]$ that satisfies

$$|S(f, P, C) - R| < \varepsilon$$

for every set of tags C. We apply the definition of absolute value bars to get

$$R - \varepsilon < S(f, P, C) < R + \varepsilon \tag{3.1}$$

We want to 1) make use of the chosen ε and 2) relate $S(f, P, C)$ to $U(f, P)$ and $L(f, P)$. Applying the definitions of supremum and infimum achieves both goals.

We first consider $U(f, P)$. By definition of supremum, there exists a set of tags C_1 such that for every $k \in [0, n-1]$

$$M_k - \varepsilon < f(c_k) < M_k$$

Therefore,

$$\begin{aligned} S(f, P, C_1) &= \sum_{k=0}^{n-1} f(c_k)(x_{k+1} - x_k) \\ &> \sum_{k=0}^{n-1} (M_k - \varepsilon)(x_{k+1} - x_k) \\ &= \sum_{k=0}^{n-1} M_k(x_{k+1} - x_k) - \sum_{k=0}^{n-1} \varepsilon(x_{k+1} - x_k) \\ &= U(f, P) - \varepsilon(b - a) \end{aligned} \tag{3.2}$$

Similarly for $L(f, P)$, by definition of infimum, there exists a set of tags C_2 such that for every $k \in [0, n-1]$,

$$m_k < f(c_k) < m_k + \varepsilon$$

Therefore,

$$\begin{aligned}
S(f, P, C_2) &= \sum_{k=0}^{n-1} f(c_k)(x_{k+1} - x_k) \\
&< \sum_{k=0}^{n-1} (m_k + \varepsilon)(x_{k+1} - x_k) \\
&= \sum_{k=0}^{n-1} m_k(x_{k+1} - x_k) + \sum_{k=0}^{n-1} \varepsilon(x_{k+1} - x_k) \\
&= L(f, P) + \varepsilon(b - a) \qquad (3.3)
\end{aligned}$$

However, since C_1 and C_2 may not necessarily be equal, we need some way to relate Equation (3.2) and Equation (3.3). We apply Equation (3.1) to get

$$\begin{aligned}
R + \varepsilon &> U(f, P) - \varepsilon(b - a) \\
R - \varepsilon &< L(f, P) + \varepsilon(b - a)
\end{aligned}$$

and a little bit of rearranging gives

$$\begin{aligned}
R + \varepsilon + \varepsilon(b - a) &> U(f, P) \qquad (3.4) \\
R - \varepsilon - \varepsilon(b - a) &< L(f, P) \qquad (3.5)
\end{aligned}$$

Recall that our choice of P depends on ε, so we cannot generalise this to all $\varepsilon > 0$ yet. But then

$$U(f, P) \geq U(f) \qquad \text{and} \qquad L(f, P) \leq L(f)$$

for every partition P. Substituting into Equation (3.4) and Equation (3.5) gives

$$\begin{aligned}
R + \underbrace{\varepsilon}_{\varepsilon/2} + \underbrace{\varepsilon(b - a)}_{\varepsilon/2} &> U(f) \\
R - \varepsilon - \varepsilon(b - a) &< L(f)
\end{aligned}$$

Since this holds for every $\varepsilon > 0$, we have

$$R \geq U(f) \qquad \text{and} \qquad R \leq L(f)$$

so it follows that $U(f) \leq L(f)$. But then $L(f) \leq U(f)$ by Proposition 3.10. So $L(f) = U(f) = R$.

Darboux implies Riemann. Fix $\varepsilon > 0$. Since f is Darboux integrable:

- By the Cauchy criterion, there exists a partition Q of $[a, b]$ such that

$$U(f, Q) - L(f, Q) < \varepsilon \qquad (3.6)$$

- f is bounded, say $|f| \leq M$.

Let P be a partition of $[a, b]$, and suppose $\|P\| < \delta$ for some $\delta > 0$ which we will find later. Suppose Q consists of m subintervals, and therefore $m + 1$ endpoints. Since a and b are common to both P and Q, there are at most $m - 1$ endpoints in Q that are not in P. So there are at most $m - 1$ subintervals in P which contain these endpoints (that are in Q but not in P).

Consider $P' = P \cup Q$. Suppose I_k is a subinterval in P which contains at least one endpoint in Q not in P. Then the terms corresponding to I_k in $U(f, P')$ and $U(f, P)$ differ by at most $2M|I_k|$.

(Some remarks here:

- *Corresponding* here means the terms that *span* I_k. Since P' contains all the subintervals in P, and definitely some more, since it includes endpoints in Q not in P, the term in $U(f, P')$ *corresponding* to I_k might be
$$M_j|I_j| + M_{j+1}|I_{j+1}| + \cdots + M_{j'}|I_{j'}|$$
where $I_j \cup I_{j+1} \cup \ldots \cup I_{j'} = I_k$.
- $-M \leq f \leq M$, hence the difference $2M|I_k|$.)

There are at most $m-1$ of these differing terms. Also, we have assumed that $\|P\| < \delta$, so any of these differing subintervals satisfy $|I_k| < \delta$. It follows that

$$U(f, P) - U(f, P') \leq 2(m-1)M\delta$$

By exactly the same reasoning,

$$L(f, P') - L(f, P) \leq 2(m-1)M\delta$$

(Notice the P and P' swapping positions in the two inequalities. This is because $U(f, P) \geq U(f, P')$ but $L(f, P) \leq L(f, P')$.)

Now that we have related the sums with δ, what about ε? The only place where we have encountered ε so far is Equation (3.6). This means we have to relate these inequalities to $U(f, Q)$ and $L(f, Q)$. This isn't too hard: since P' refines Q,

$$U(f, Q) \geq U(f, P') \qquad \text{and} \qquad L(f, Q) \leq L(f, P')$$

A little bit of rearranging gives

$$\begin{aligned} U(f, P) &\leq U(f, Q) + 2(m-1)M\delta \\ L(f, P) &\geq L(f, Q) - 2(m-1)M\delta \end{aligned}$$

Hence,

$$U(f, P) - L(f, P) \leq \underbrace{U(f, Q) - L(f, Q)}_{\text{e.g. } \varepsilon/4} + \underbrace{4(m-1)M\delta}_{\text{e.g. } \varepsilon/2} \leq \frac{3\varepsilon}{4} < \varepsilon$$

Observe that

$$L(f, P) \leq S(f, P, C) \leq U(f, P)$$

for any tagged partition (P, C). So

$$|S(f, P, C) - D| < \varepsilon$$

for every tagged partition (P, C) with $\|P\| < \delta$, where

$$\delta = \frac{\varepsilon}{8(m-1)M}$$

□

Here is the same proof after cleaning up the ε values and less verbose.

Proof. Riemann implies Darboux. Fix $\varepsilon > 0$. Let P be a partition of $[a, b]$ that satisfies

$$|S(f, P, C) - R| < \varepsilon$$

for every set of tags C. We apply the definition of absolute value bars to get

$$R - \frac{\varepsilon}{2} < S(f, P, C) < R + \frac{\varepsilon}{2} \tag{3.7}$$

We want to 1) make use of the chosen ε and 2) relate $S(f, P, C)$ to $U(f, P)$ and $L(f, P)$. Applying the definitions of supremum and infimum achieves both goals.

We first consider $U(f, P)$. By definition of supremum, there exists a set of tags C_1 such that for every $k \in [0, n-1]$

$$M_k - \frac{\varepsilon}{2(b-a)} < f(c_k) < M_k$$

Therefore,

$$\begin{aligned}
S(f, P, C_1) &= \sum_{k=0}^{n-1} f(c_k)(x_{k+1} - x_k) \\
&> \sum_{k=0}^{n-1} (M_k - \varepsilon)(x_{k+1} - x_k) \\
&= \sum_{k=0}^{n-1} M_k(x_{k+1} - x_k) - \sum_{k=0}^{n-1} \varepsilon(x_{k+1} - x_k) \\
&= U(f, P) - \varepsilon(b-a) \qquad (3.8)
\end{aligned}$$

Similarly for $L(f, P)$, by definition of infimum, there exists a set of tags C_2 such that for every $k \in [0, n-1]$,

$$m_k < f(c_k) < m_k + \frac{\varepsilon}{2(b-a)}$$

Therefore,

$$\begin{aligned}
S(f, P, C_2) &= \sum_{k=0}^{n-1} f(c_k)(x_{k+1} - x_k) \\
&< \sum_{k=0}^{n-1} (m_k + \frac{\varepsilon}{2(b-a)})(x_{k+1} - x_k) \\
&= \sum_{k=0}^{n-1} m_k(x_{k+1} - x_k) + \sum_{k=0}^{n-1} \frac{\varepsilon}{2(b-a)}(x_{k+1} - x_k) \\
&= L(f, P) + \frac{\varepsilon}{2(b-a)} \cdot (b-a) \\
&= L(f, P) + \frac{\varepsilon}{2} \qquad (3.9)
\end{aligned}$$

However, since C_1 and C_2 may not necessarily be equal, we need some way to relate Equation (3.8) and Equation (3.9). We apply Equation (3.7) to get

$$\begin{aligned}
R + \varepsilon > U(f, P) - \frac{\varepsilon}{2} \\
R - \varepsilon < L(f, P) + \frac{\varepsilon}{2}
\end{aligned}$$

and a little bit of rearranging gives

$$R + \frac{\varepsilon}{2} + \frac{\varepsilon}{2} > U(f, P) \tag{3.10}$$
$$R - \frac{\varepsilon}{2} - \frac{\varepsilon}{2} < L(f, P) \tag{3.11}$$

Recall that our choice of P depends on ε, so we cannot generalise this to all $\varepsilon > 0$ yet. But then

$$U(f, P) \geq U(f) \qquad \text{and} \qquad L(f, P) \leq L(f)$$

for every partition P. Substituting into Equation (3.10) and Equation (3.11) gives

$$R + \varepsilon > U(f)$$
$$R - \varepsilon < L(f)$$

Since this holds for every $\varepsilon > 0$, we have

$$R \geq U(f) \qquad \text{and} \qquad R \leq L(f)$$

so it follows that $U(f) \leq L(f)$. But then $L(f) \leq U(f)$ by Proposition 3.10. So $L(f) = U(f) = R$.

Darboux implies Riemann. Fix $\varepsilon > 0$. Since f is Darboux integrable:

- By the Cauchy criterion, there exists a partition Q of $[a, b]$ such that

$$U(f, Q) - L(f, Q) < \frac{\varepsilon}{4} \tag{3.12}$$

- f is bounded, say $|f| \leq M$.

Let P be a partition of $[a, b]$, and suppose $\|P\| < \delta$ for some $\delta > 0$ which we will find later. Suppose Q consists of m subintervals, and therefore $m + 1$ endpoints. Since a and b are common to both P and Q, there are at most $m - 1$ endpoints in Q that are not in P. So there are at most $m - 1$ subintervals in P which contain these endpoints (that are in Q but not in P).

Consider $P' = P \cup Q$. Suppose I_k is a subinterval in P which contains at least one endpoint in Q not in P. Then the terms corresponding to I_k in $U(f, P')$ and $U(f, P)$ differ by at most $2M|I_k|$.

There are at most $m - 1$ of these differing terms. Also, we have assumed that $\|P\| < \delta$, so any of these differing subintervals satisfy $|I_k| < \delta$. It follows that

$$U(f, P) - U(f, P') \leq 2(m - 1)M\delta$$

By exactly the same reasoning,

$$L(f, P') - L(f, P) \leq 2(m - 1)M\delta$$

Now that we have related the sums with δ, what about ε? The only place where we have encountered ε so far is Equation (3.12). This means we have to relate these inequalities to $U(f, Q)$ and $L(f, Q)$. This isn't too hard: since P' refines Q,

$$U(f, Q) \geq U(f, P') \qquad \text{and} \qquad L(f, Q) \leq L(f, P')$$

A little bit of rearranging gives

$$U(f,P) \le U(f,Q) + 2(m-1)M\delta$$
$$L(f,P) \ge L(f,Q) - 2(m-1)M\delta$$

Hence,

$$U(f,P) - L(f,P) \le U(f,Q) - L(f,Q) + \underbrace{4(m-1)M\delta}_{\text{e.g. } \varepsilon/2} \le \frac{3\varepsilon}{4} < \varepsilon$$

Observe that

$$L(f,P) \le S(f,P,C) \le U(f,P)$$

for any tagged partition (P,C). So

$$|S(f,P,C) - D| < \varepsilon$$

for every tagged partition (P,C) with $\|P\| < \delta$, where

$$\delta = \frac{\varepsilon}{8(m-1)M}$$

□

3.4 Analogue version of triangle inequality

We can generalise, in some sense, the triangle inequality to integrals.

Lemma 3.18. If $f \le g$, then

$$\int_a^b f(x)\mathrm{d}x \le \int_a^b g(x)\mathrm{d}x$$

Proof. We use the linearity of integrals without proof.

$$\begin{aligned}\int_a^b g - \int_a^b f &= \int_a^b (f-g) \quad \text{by linearity} \\ &\ge \int_a^b 0 \\ &= 0\end{aligned}$$

and the result follows. □

Proposition 3.19 (Analogue version of triangle inequality)**.** For every function $f : [a,b] \to \mathbb{R}$,

$$\left|\int_a^b f(x)\ \mathrm{d}x\right| \le \int_a^b |f(x)|\ \mathrm{d}x$$

Proof. Observe that

$$-|f| \le f \le |f|$$

By Lemma 3.18,

$$\int_a^b -|f| \le \int_a^b f \le \int_a^b |f|$$

and the result follows. $\square$

Chapter 4

Differentiation

4.1 Definition

Definition 4.1 (Newton's difference quotient). Let $f : (a, b) \to \mathbb{R}$. The Newton's difference quotient at x for f is given by

$$\frac{\Delta f(x)}{\Delta x} = \frac{f(x+h) - f(x)}{h}$$

Definition 4.2 (Derivative). Suppose $f : (a, b) \to \mathbb{R}$ and let c be an interior point, i.e. $a < c < b$. Then f is differentiable at c with derivative $f'(c)$ if

$$\lim_{h \to 0} \frac{f(c+h) - f(c)}{h} = f'(c)$$

exists.

Definition 4.3 (Left and right derivatives). Suppose $f : [a, b] \to \mathbb{R}$. Then f is right-differentiable at $a \leq c < b$ with right derivative $f'(c^+)$ if

$$\lim_{h \to 0^+} \frac{f(c+h) - f(c)}{h} = f'(c^+)$$

exists. Similarly, f is left-differentiable at $a < c \leq b$ with left derivative $f'(c^-)$ if

$$\lim_{h \to 0^-} \frac{f(c+h) - f(c)}{h} = f'(c^-)$$

exists.

In Definition 4.2, c must be an interior point because we want to evaluate a two-sided limit as the argument of the function approaches c, which requires the existence of an open (two-sided) neighbourhood of c. On the other hand, Definition 4.3 can be useful for points of discontinuity or at the boundaries, since we only need a one-sided neighbourhood to evaluate the one-sided limits.

Using the fact that a two-sided limit exists if and only if both of its one-sided limits exist and are equal, we can now provide an alternative, often more fool-proof, definition of differentiability:

Definition 4.4. A function is differentiable if and only if its left and right derivatives exist and are equal.

Example. Find the derivative of $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} x^2, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

Solution. For $x > 0$, the derivative $f'(x)$ is given by

$$\begin{aligned} f'(x) &= \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \\ &= \lim_{h \to 0} \frac{(x+h)^2 - x^2}{h} \\ &= \lim_{h \to 0} \frac{2hx + h^2}{h} \\ &= \lim_{h \to 0} 2x + h \\ &= 2x \end{aligned}$$

> **Remark.** We are allowed to substitute $f(x+h)$ with $(x+h)^2$ here. This works because $x > 0$, so regardless if h approaches 0 from the left or right, as long as h is sufficiently close to 0, $f(x+h)$ will evaluate to $(x+h)^2$.

For $x < 0$, the derivative $f'(x)$ is given by

$$\begin{aligned} f'(x) &= \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \\ &= \lim_{h \to 0} \frac{0 - 0}{h} \\ &= 0 \end{aligned}$$

> **Remark.** Similar to above, as long as h is sufficiently close to x, $f(x+h)$ will evaluate to 0.

For $x = 0$, we consider the left and right derivatives of f at 0. The right derivative is

$$\begin{aligned} f'(0^+) &= \lim_{h \to 0^+} \frac{f(h) - f(0)}{h} \\ &= \lim_{h \to 0^+} \frac{h^2 - 0}{h} \\ &= \lim_{h \to 0^+} h \\ &= 0 \end{aligned}$$

The left derivative is

$$\begin{aligned} f'(0^-) &= \lim_{h\to 0^-} \frac{f(h)-f(0)}{h} \\ &= \lim_{h\to 0^-} \frac{0-0}{h} \\ &= 0 \end{aligned}$$

Since $f'(0^-) = 0 = f'(0^+)$, the derivative $f'(0)$ exists and $f'(0) = 0$. Therefore, f is differentiable on $\mathbb{R}$ with derivative f' given by

$$f'(x) = \begin{cases} 2x, & x > 0 \\ 0, & x \le 0 \end{cases}$$

Example. Determine if $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = |x|$ is differentiable at 0.

Solution. Since $|x| = x$ for $x > 0$, the right derivative at 0 is

$$f'(0^+) = \lim_{h\to 0^+} \frac{f(h)-f(0)}{h} = \lim_{h\to 0^+} \frac{|h|}{h} = \lim_{h\to 0^+} \frac{h}{h} = 1$$

Since $|x| = -x$ for $x < 0$, the left derivative at 0 is

$$f'(0^-) = \lim_{h\to 0^-} \frac{f(h)-f(0)}{h} = \lim_{h\to 0^-} \frac{|h|}{h} = \lim_{h\to 0^-} \frac{-h}{h} = -1$$

As $f'(0^+) \neq f'(0^-)$, $f'(0)$ does not exist and f is not differentiable at 0.

Example. For which values of x is $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} x \sin \frac{1}{x}, & x \neq 0 \\ 0, & x = 0 \end{cases}$$

differentiable? For these values of x, find $f'(x)$.

Solution. For $x \neq 0$, the derivative is given by the product and chain rules:

$$\begin{aligned} f'(x) &= x \cdot \cos \frac{1}{x} \cdot \left(-\frac{1}{x^2}\right) + \sin \frac{1}{x} \cdot 1 \\ &= \sin \frac{1}{x} - \frac{1}{x} \cos \frac{1}{x} \end{aligned}$$

$f'(0)$ does not exist since the limit

$$\lim_{h\to 0} \frac{f(h)-f(0)}{h} = \lim_{h\to 0} \frac{h \sin \frac{1}{h}}{h} = \lim_{h\to 0} \sin \frac{1}{h}$$

does not exist, so f is not differentiable at 0.

Example. For which values of x is $f : \mathbb{R} \to \mathbb{R}$ defined by

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

differentiable? For these values of x, find $f'(x)$.

Solution. For $x \neq 0$, the derivative is given by

$$f'(x) = 2x \sin \frac{1}{x} - \cos \frac{1}{x}$$

For $x = 0$,

$$\lim_{h \to 0} \frac{f(h) - f(0)}{h} = \lim_{h \to 0} \frac{h^2 \sin \frac{1}{h}}{h} = \lim_{h \to 0} h \sin \frac{1}{h} = 0$$

so $f'(0)$ exists and $f'(0) = 0$.

Remark. Observe that $\lim_{x \to 0} f'(x)$ does not exist because of the $\cos \frac{1}{x}$ term, but $f'(0)$ exists. So f' is not continuous at 0, and we say that f is not continuously differentiable.

4.2 Properties of the derivative

4.2.1 Differentiability and continuity

Theorem 4.5 (Continuity)**.** If $f : (a, b) \to \mathbb{R}$ is differentiable at $c \in (a, b)$, then f is continuous at c.

Proof. Suppose f is differentiable at c. Then

$$\lim_{h \to 0} \frac{f(c + h) - f(c)}{h} = f'(c)$$

exists.

We want to show that

$$\lim_{x \to c} f(x) = f(c)$$

or equally,

$$\lim_{h \to 0} f(c + h) = f(c)$$

or equally,

$$\lim_{h \to 0} f(c + h) - f(c) = 0$$

If we divide the expression inside the limit by h then we get the definition of the derivative. So the proof follows:

$$\begin{aligned} \lim_{h \to 0} f(c + h) - f(c) &= \lim_{h \to 0} h \cdot \frac{f(c + h) - f(c)}{h} \\ &= \lim_{h \to 0} h \cdot \lim_{h \to 0} \frac{f(c + h) - f(c)}{h} \\ &= f'(c) \lim_{h \to 0} h \\ &= 0 \end{aligned}$$

so by the linearity of limits, we get

$$\lim_{h\to 0} f(c+h) = f(c)$$

□

4.2.2 Chain rule

Since we're not JMC students, we'll not care too much about the domains and codomains of the functions:

Theorem 4.6 (Chain rule)**.** If g is differentiable at c and f is differentiable at $g(c)$, then $f \circ g$ is differentiable at c with derivative given by

$$(f \circ g)'(c) = f'(g(c)) \cdot g'(c)$$

The following proof is adapted from the first proof on the Wikipedia page, with a bit more explanation.

Proof. By definition of the derivative, the LHS is

$$(f \circ g)'(c) = \lim_{x\to c} \frac{f(g(x)) - f(g(c))}{x - c}$$

and the RHS is

$$f'(g(c)) \cdot g'(c) = \lim_{g(x)\to g(c)} \frac{f(g(x)) - f(g(c))}{g(x) - g(c)} \cdot \lim_{x\to c} \frac{g(x) - g(c)}{x - c}$$

There are two issues with the first term in the product. The first issue is that the limit now goes from $g(x) \to g(c)$ instead of $x \to c$, but this can be easily fixed, as we will see later. The second issue is there may exist $g(x) - g(c)$ such that $x \neq c$ but $g(x) - g(c)$. Why is this an issue? Suppose $c = 0$ and g is defined by

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

The $\sin \frac{1}{x}$ term will keep oscillating around 0: no matter how close x is to 0, there must exist some $x_0 < x$ such that $g(x_0) = 0$. Then, recalling the definition of limits:

$$\forall \varepsilon > 0,\ \exists \delta > 0 : 0 < |x| < \delta \Rightarrow \left| \frac{f(g(x)) - f(g(c))}{g(x) - g(c)} - \text{blah} \right| < \varepsilon$$

For all $\delta > 0$, this x_0 will make the implication false because

$$\frac{f(g(x_0)) - f(g(c))}{g(x_0) - g(c)}$$

is undefined and so the limit is undefined. But we know that $g'(0)$ exists (even though g' is not continuous at 0), so g should be continuous at 0, and the chain rule should work in this case. To work around this issue, we introduce another function Q defined by

$$Q(y) = \begin{cases} \dfrac{f(y) - f(g(c))}{y - g(c)}, & y \neq g(c) \\ f'(g(c)), & y = g(c) \end{cases}$$

Proposition 4.7.

$$\frac{f(g(x)) - f(g(c))}{x - c} \equiv Q(g(x)) \cdot \frac{g(x) - g(c)}{x - c}$$

Proof. When $g(x) \neq g(c)$,

$$Q(g(x)) \cdot \frac{g(x) - g(c)}{x - c} \equiv \frac{f(g(x)) - f(g(c))}{g(x) - g(c)} \cdot \frac{g(x) - g(c)}{x - c} \equiv \frac{f(g(x)) - f(g(c))}{x - c}$$

When $g(x) = g(c)$, the RHS becomes

$$Q(g(c)) \cdot \frac{g(x) - g(c)}{x - c} = 0 \cdot \frac{g(x) - g(c)}{x - c} = 0$$

and the LHS also evaluates to 0 since $f(g(x)) = f(g(c))$. $\square$

It follows from Proposition 4.7 that it is sufficient to show that

$$\lim_{x \to c} \left[Q(g(x)) \cdot \frac{g(x) - g(c)}{x - c} \right] = f'(g(c)) \cdot g'(c)$$

The limit exists if the limits of its two factors exist (i.e. do not shoot off to infinity), in which case it is equal to the product of the limits of the two factors. So we find the limits of the two factors separately, and if they both exist, then we have found the limit of the product.

Since g is differentiable at c,

$$\lim_{x \to c} \frac{g(x) - g(c)}{x - c} = g'(c)$$

exists.

Since f is differentiable at $g(c)$,

$$\lim_{y \to g(c)} \frac{f(y) - f(g(c))}{y - g(c)} = f'(g(c))$$

exists, and since $Q(g(c)) = f'(g(c))$, we get

$$\lim_{y \to g(c)} Q(y) = Q(g(c))$$

so Q is continuous at $g(c)$.

Remark. Technically, Q doesn't have to be continuous so we can define $Q(g(c))$ to be anything, but we define it to be $f'(g(c))$ so that everything can fall into place a bit more nicely, and there's no reason not to make Q continuous when we can easily do so.

As g is continuous at c, we have $x \to c$ implies $g(x) \to g(c)$ (resolving the first issue mentioned above), and so

$$\begin{aligned} \lim_{x \to c} Q(g(x)) &= \lim_{y \to g(c)} Q(y) && \text{by continuity of } g \text{ at } c \\ &= Q(g(c)) && \text{by continuity of } Q \text{ at } g(c) \\ &= f'(g(c)) \end{aligned}$$

Since both limits exist, we are done. $\square$

4.2.3 Product and quotient rules

Theorem 4.8 (Product rule). If $f, g : (a, b) \to \mathbb{R}$ are differentiable at $c \in (a, b)$, then fg is differentiable at c with derivative

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

Proof.

$$\begin{aligned}
(fg)'(c) &= \lim_{h\to 0} \frac{f(c+h)g(c+h) - f(c)g(c)}{h} \\
&= \lim_{h\to 0} \frac{f(c+h)g(c+h) - f(c)g(c+h) + f(c)g(c+h) - f(c)g(c)}{h} \\
&= \lim_{h\to 0} \frac{f(c+h) - f(c)}{h} \cdot \lim_{h\to 0} g(c+h) + f(c) \lim_{h\to 0} \frac{g(c+h) - g(c)}{h} \\
&= f'(c) \lim_{h\to 0} g(c+h) + f(c)g'(c) && \text{by differentiability of } f \text{ and } g \\
&= f'(c)g(c) + f(c)g'(c) && \text{by continuity of } g
\end{aligned}$$

□

Theorem 4.9 (Quotient rule). If $f, g : (a, b) \to \mathbb{R}$ are differentiable at $c \in (a, b)$, then $\left(\frac{f}{g}\right)$ is differentiable at c with derivative

$$\left(\frac{f}{g}\right)'(c) = \frac{f'(c)g(c) - f(c)g'(c)}{[g(c)]^2}$$

The first proof is similar to the proof for the product rule.

Proof.

$$\begin{aligned}
\left(\frac{f}{g}\right) &= \lim_{h\to 0} \frac{\frac{f(c+h)}{g(c+h)} - \frac{f(c)}{g(c)}}{h} \\
&= \lim_{h\to 0} \frac{f(c+h)g(c) - f(c)g(c+h)}{h \cdot g(c)g(c+h)} \\
&= \lim_{h\to 0} \frac{1}{g(c)g(c+h)} \cdot \lim_{h\to 0} \frac{f(c+h)g(c) - f(c)g(c) + f(c)g(c) - f(c)g(c+h)}{h} \\
&= \frac{1}{[g(c)]^2} \lim_{h\to 0} \frac{f(c+h)g(c) - f(c)g(c) + f(c)g(c) - f(c)g(c+h)}{h} && \text{by continuity of } g \\
&= \frac{1}{[g(c)]^2} \left[g(c) \lim_{h\to 0} \frac{f(c+h) - f(c)}{h} - f(c) \lim_{h\to 0} \frac{g(c+h) - g(c)}{h} \right] \\
&= \frac{1}{[g(c)]^2} [g(c)f'(c) - f(c)g'(c)] && \text{by differentiability of } f \text{ and } g \\
&= \frac{f'(c)g(c) - f(c)g'(c)}{[g(c)]^2}
\end{aligned}$$

□

The second proof combines the product and chain rules.

Proof. By the chain rule,

$$\frac{\mathrm{d}}{\mathrm{d}x}\frac{1}{g(x)} = \frac{\mathrm{d}}{\mathrm{d}[g(x)]}\left(\frac{1}{g(x)}\right) \cdot \frac{\mathrm{d}}{\mathrm{d}x}g(x) = -\frac{1}{[g(x)]^2} \cdot g'(x) = -\frac{g'(x)}{[g(x)]^2}$$

Substituting g with $\left(\frac{1}{g}\right)$ into the product rule,

$$\begin{aligned}\left(\frac{f}{g}\right)'(c) &= f'(c)\left(\frac{1}{g}\right) + f(c)\left(\frac{1}{g}\right)' \\ &= \frac{f'(c)}{g(c)} + f(c) \cdot \left(-\frac{g'(c)}{[g(c)]^2}\right) \\ &= \frac{f'(c)}{g(c)} - \frac{f(c)g'(c)}{[g(c)]^2} \\ &= \frac{f'(c)g(c) - f(c)g'(c)}{[g(c)]^2}\end{aligned}$$

□

4.2.4 Extreme values

Suppose $f : A \to \mathbb{R}$.

Definition 4.10 (Global and local maxima). f has a global maximum at $c \in A$ if

$$f(x) \leq f(c) \quad \forall x \in A$$

and f has a local maximum at $c \in A$ if there exists a neighbourhood U of c such that

$$f(x) \leq f(c) \quad \forall x \in A \cap U$$

Definition 4.11 (Global and local minima). f has a global minimum at $c \in A$ if

$$f(x) \geq f(c) \quad \forall x \in A$$

and f has a local minimum at $c \in A$ if there exists a neighbourhood U of c such that

$$f(x) \geq f(c) \quad \forall x \in A \cap U$$

Theorem 4.12 (Fermat's theorem). Suppose $f : A \to \mathbb{R}$ has a local extreme value at an interior point $c \in A$ and f is differentiable at c. Then $f'(c) = 0$.

Proof. We consider the case where f has a local minimum at c. The case where c is a local maximum follows similarly, only with some inequalities reversed.

If f has a local minimum at c, then there exists a δ-neighbourhood $(c-\delta, c+\delta)$ at c such that

$$f(x) \geq f(c) \quad \forall x \in (c-\delta, c+\delta)$$

For every $x \in (c, c+\delta)$ we have $x - c > 0$ and $f(x) - f(c) \geq 0$. So

$$\frac{f(x) - f(c)}{x - c} \geq 0$$

and therefore

$$f'(c) = \lim_{x \to c^+} \frac{f(x) - f(c)}{x - c} \geq 0$$

Also, for every $x \in (c - \delta, c)$ we have $x - c < 0$ and $f(x) - f(c) \geq 0$. So

$$\frac{f(x) - f(c)}{x - c} \leq 0$$

and therefore

$$f'(c) = \lim_{x \to c^-} \frac{f(x) - f(c)}{x - c} \leq 0$$

Remark. Since $f'(c)$ exists,

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

exists and is equal to $f'(c)$. Since the two-sided limit exists, the left and right hand limits exist and are equal to the two-sided limit, which is equal to $f'(c)$.

Combining the two inequalities, we get

$$0 \leq f'(c) \leq 0$$

so $f'(c) = 0$. □

Remark. Note that c must be an interior point, since the proof involves the two-sided limit of the difference quotient at c. Otherwise, if c is at a boundary point, then only one of the inequalities will be defined. For instance, if a is a local maximum, since only the right hand limit exists, we get $f'(a^+) \leq 0$.

Theorem 4.12 tells us that the global and local extrema of $f : A \to \mathbb{R}$ must be either:

(1) Boundary points of A

(2) (Interior) points where f is not differentiable

(3) Stationary points of f

Examining the statement of Theorem 4.12, we first try to negate the antecedent. Negating the first conjunct (c is an interior point) gives us (1). Negating the second conjunct (f is differentiable at c) gives us (2). It is sufficient to only consider interior points here, since (1) also covers the cases where the boundary points are not differentiable. If the antecedent is true, then the consequent must be true as well, which gives us (3). As global extrema must be local extrema, Theorem 4.12 applies to global extrema as well.

4.3 Mean value theorem

Theorem 4.13 (Rolle's theorem)**.** Suppose that $f : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$ and differentiable on (a, b), and $f(a) = f(b)$. Then there exists $c \in (a, b)$ such that $f'(c) = 0$.

Proof. By the Weierstrass extreme value theorem, f attains its global maximum and minimum on $[a, b]$. If they are both attained at the same endpoint, then f is constant, so $f'(c) = 0$ for every $c \in (a, b)$. Otherwise, at least one global extremum is attained at an interior point c. From the definition of global and local extrema, this extremum must also be a local extremum. By Theorem 4.12, $f'(c) = 0$.

Remark. If the extremum is a minimum, for example, then $f(x) \geq f(c)$ for all $x \in [a,b]$, then any arbitrary neighbourhood U of c would satisfy the definition of local minima, namely $f(x) \leq f(c)$ for all $x \in [a,b] \cap U$.

$\square$

Note that we require continuity on the closed interval $[a,b]$ but differentiability only on the open interval (a,b). This is because the boundary points are only relevant when both of the global extrema are at the boundary points, in which case we don't care about the derivative as f is constant.

We generalise Theorem 4.13 to the mean value theorem:

Theorem 4.14 (Mean value theorem)**.** Suppose $f : [a,b] \to \mathbb{R}$ is continuous on $[a,b]$ and differentiable on (a,b). Then there exists $c \in (a,b)$ such that

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

Intuition. We want to generate a new function $g : [a,b] \to \mathbb{R}$ based on f that satisfies $g(a) = g(b)$ (as well as continuity and differentiability, of course). This allows us to apply Theorem 4.13 and get $g'(c) = 0$. We want to extract $f'(c)$ from the fact that $g'(c) = 0$. So we start by reverse engineering g':

$$f'(c) = \frac{f(b) - f(a)}{b - a} \quad \Leftrightarrow \quad f'(c) - \frac{f(b) - f(a)}{b - a} = 0 = g'(c)$$

Let

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

Integrating both sides, we get

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

We want $g(a) = g(b)$. To make our lives easier, suppose $g(a) = g(b) = 0$. Consider

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

We can easily get rid of $f(a)$ by subtracting it from $g(a)$. The other part is a rather hairy constant times a, so instead of subtracting it we will replace x with $x - a$. So we have this new definition of g:

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

and we can check that $g(b) = 0$ as well. As g is just a linear combination of continuous and differentiable functions, g has the same continuity and differentiability properties as f.

Proof. The function $g : [a,b] \to \mathbb{R}$ defined by

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

is continuous on $[a,b]$ and differentiable on (a,b) with

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

Also, $g(a) = g(b) = 0$. By Theorem 4.13, there exists an interior point $c \in (a, b)$ such that $g'(c) = 0$. Rearranging the equality gives us the result. $\square$

4.3.1 Applications

Theorem 4.15. Suppose $f : (a, b) \to \mathbb{R}$ is differentiable. Then f is increasing if and only if $f'(x) \geq 0$ for every $x \in (a, b)$. Moreover, if $f'(x) > 0$ for every $x \in (a, b)$, then f is strictly increasing.

Proof. **Only if.** Suppose f is increasing. We want to show that

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \geq 0$$

For $h > 0$, $f(x+h) - f(x) > 0$ as f increasing, so the difference quotient > 0. For $h < 0$, $f(x+h) - f(x) < 0$ as f increasing, so the difference quotient also > 0. So the difference quotient > 0 for all $h \neq 0$. Then we are done.

If. Suppose $f'(x) \geq 0$ for every $x \in (a, b)$. Take arbitrary $x, y \in (a, b)$ such that $a < x < y < b$. By the mean value theorem, there exists some c between x and y such that

$$f'(c) = \frac{f(y) - f(x)}{y - x}$$

But then $f'(c) \geq 0$. Since $y - x > 0$, it must be that $f(y) - f(x) \geq 0$, so f is increasing. If $f'(c) > 0$, then $f(y) - f(x) > 0$, so f is strictly increasing. $\square$

Note that if f is strictly increasing, it does not follow that $f'(x) > 0$ for every $x \in (a, b)$. For example, the function $f(x) = x^3$.

4.4 Taylor's theorem

Definition 4.16 (Taylor polynomial)**.** Let $f : (a, b) \to \mathbb{R}$ and suppose f is n times differentiable, i.e. f', f'', f''', $\ldots$, $f^{(n)}$ exist on (a, b). The Taylor polynomial of degree n of f at $c \in (a, b)$ is

$$\begin{aligned} P_n(x) &= f(c) + f'(c)(x-c) + \frac{1}{2!}f''(c)(x-c)^2 + \cdots + \frac{1}{n!}f^{(n)}(c)(x-c)^n \\ &= \sum_{k=0}^{n} \frac{1}{k!} f^{(k)}(c)(x-c)^k \end{aligned}$$

If f is infinitely differentiable and the derivatives at c do not vanish (are not equal to 0), then the Taylor polynomial will only ever be an approximation of f. The error between this approximation and the true value of the function is denoted by R_n with

$$f(x) = P_n(x) + R_n(x)$$

Theorem 4.17 (Taylor's theorem)**.** Suppose $f : (a, b) \to \mathbb{R}$ has $n+1$ derivatives on (a, b) and let $c \in (a, b)$. For every $x \in (a, b)$, there exists ξ between c and x (either $c < \xi < x$ or $x < \xi < c$) such that

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

with

$$R_n(x) = \frac{1}{(n+1)!} f^{(n+1)}(\xi)(x-c)^{n+1}$$

This form of the remainder is called the Lagrange form, or the Lagrange error term.

The remainder term can take many other forms, but we'll only cover the Lagrange form here.

Intuition. Since we only want to show the existence of such ξ, this is a big hint that we will need to use either Rolle's theorem (Theorem 4.13) or the mean value theorem (Theorem 4.14) at some point. We want to show that the remainder, expressed in terms of f, is equal to the given form, so we define another function g to disambiguate it from R_n:

$$g(?) = f(x) - f(c) - f'(c)(x-c) - \frac{1}{2!}f''(c)(x-c)^2 - \cdots - \frac{1}{n!}f^{(n)}(c)(x-c)^n$$

What should the variable be? This expression can either be $g(x)$ or $g(c)$ (probably not $g(n)$). We choose it to be $g(c)$, so the general function is defined by

$$g(t) = f(x) - f(t) - f'(t)(x-t) - frac12!f''(t)(x-t)^2 - \cdots - \frac{1}{n!}f^{(n)}(t)(x-t)^n$$

Now we try differentiating g to see what we get out of it. Before we do that, we rewrite g using a summation:

$$g(t) = f(x) - f(t) - \sum_{k=1}^{n} \frac{1}{k!} f^{(k)}(t)(x-t)^k$$

It will soon become clear why we don't put $f(t)$ into the summation. We differentiate this form of g using the product and chain rules:

$$\begin{aligned} g'(t) &= -f'(t) - \sum_{k=1}^{n} \left[\frac{1}{k!} f^{(k+1)}(t)(x-t)^k + \frac{1}{k!} f^{(k)}(t) \cdot k(x-t)^{k-1} \cdot (-1) \right] \\ &= -f'(t) - \sum_{k=1}^{n} \left[\frac{1}{k!} f^{(k+1)}(t)(x-t)^k - \frac{1}{(k-1)!} f^{(k)}(t) \cdot (x-t)^{k-1} \right] \end{aligned}$$

Notice that the sum is telescoping, so only the first and last terms remain:

$$\begin{aligned} g'(t) &= -f'(t) - \left[\frac{1}{n!} f^{(n+1)}(t)(x-t)^n - f'(t) \right] \\ &= -\frac{1}{n!} f^{(n+1)}(t)(x-t)^n \end{aligned}$$

We have the precious $f^{(n+1)}$ term. Remember we want to show that $g(c)$ is equal to the Lagrange form, but $g'(\xi)$ still needs some reworking. We force $g'(\xi)$ to take the required form:

$$\begin{aligned} g(c) &= \frac{1}{(n+1)!} f^{(n+1)}(\xi)(x-c)^{n+1} \\ &= -\frac{1}{n+1} \frac{(x-c)^{n+1}}{(x-\xi)^n} g'(\xi) \end{aligned}$$

To avoid mess from integration by parts, we offload all the mess in front of $g'(\xi)$ to the $g(c)$ as it is a constant:

$$g'(\xi) = -\frac{(n+1)(x-\xi)^n}{(x-c)^{n+1}} g(c)$$

We can apply Rolle's theorem here if we move everything to one side and set that to be the derivative of another function h.

$$h'(\xi) = g'(\xi) + \frac{(n+1)(x-\xi)^n}{(x-c)^{n+1}} g(c) = 0$$

More generally,

$$h'(t) = g'(t) + \frac{(n+1)(x-t)^n}{(x-c)^{n+1}} g(c)$$

Integrating both sides,

$$h(t) = g(t) - \left(\frac{x-t}{x-c}\right)^{n+1} g(c)$$

We can easily check that $h(x) = h(c) = 0$.

Proof. Fix $x, c \in (a, b)$. These will form the endpoints when we apply Rolle's theorem in due course. Define $g : (a, b) \to \mathbb{R}$ by

$$g(t) = f(x) - f(t) - f'(t)(x-t) - \frac{1}{2!} f''(t)(x-t)^2 - \cdots - \frac{1}{n!} f^{(n)}(t)(x-t)^n$$

with derivative

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

Define $h : (a, b) \to \mathbb{R}$ by

$$h(t) = g(t) - \left(\frac{x-t}{x-c}\right)^{n+1} g(c)$$

with derivative

$$h'(t) = g'(t) + \frac{(n+1)(x-t)^n}{(x-c)^{n+1}} g(c)$$

Then $h(x) = h(c) = 0$. By Rolle's theorem, there exist ξ between c and x such that $h'(\xi) = 0$. Substituting this into the equalities above gives the desired reuslt. □

4.5 L'Hospital's rule

The following definition is taken from the Wikipedia page on L'Hospital's rule. The definition in our slides only cover the special case where f and g are continuously differentiable, which should be more than sufficient for any problem we will encounter as a Computing student; but for the sake of completeness, here is the full definition.

Definition 4.18 (L'Hospital's rule)**.** Suppose f and g are functions satisfying the following criteria:

- Either $\lim_{x \to c} f(x) = \lim_{x \to c} g(x) = 0$ or $\lim_{x \to c} |f(x)| = \lim_{x \to c} |g(x)| = \infty$
- f and g are differentiable on an open interval $\mathcal{I}$ containing c, except possibly at c

- $g'(x) \neq 0$ for all x in $\mathcal{I}$ with $x \neq c$
- $\displaystyle\lim_{x \to c} \frac{f'(x)}{g'(x)}$ exists

Then
$$\lim_{x \to c} \frac{f(x)}{g(x)} = \lim_{x \to c} \frac{f'(x)}{g'(x)}$$

Refer to the Wikipedia page for the proof. We will instead focus on applications of the rule.

Example. Evaluate
$$\lim_{x \to \infty} \frac{e^x}{x^2}$$

Solution. We apply l'Hospital's rule twice.
$$\begin{aligned} \lim_{x \to \infty} \frac{e^x}{x^2} &= \lim_{x \to \infty} \frac{e^x}{2x} \\ &= \lim_{x \to \infty} \frac{e^x}{2} \\ &= \infty \end{aligned}$$

Example. Evaluate
$$\lim_{x \to 0^+} x \ln x$$

Solution. This is the limit of the product of two terms: x approaches 0 while $\ln x$ approaches $-\infty$. We need a division of two terms (i.e. a fraction) to apply l'Hospital's rule, so we convert one of them to 1 over its reciprocal. In this case we choose to convert x to $1/(1/x)$ so that the limit is in the form ∞/∞:
$$\begin{aligned} \lim_{x \to 0^+} x \ln x &= \lim_{x \to 0^+} \frac{\ln x}{\frac{1}{x}} \\ &= \lim_{x \to 0^+} \frac{\frac{1}{x}}{-\frac{1}{x^2}} \\ &= \lim_{x \to 0^+} (-x) \\ &= 0 \end{aligned}$$

Example. Evaluate
$$\lim_{x \to -\infty} xe^x$$

Solution. Just like the previous example, the limit is in the indeterminate form $\infty \cdot 0$ (we ignore the sign of ∞). We try converting x to $1/(1/x)$:

$$\begin{aligned}\lim_{x\to-\infty} xe^x &= \lim_{x\to-\infty} \frac{e^x}{\frac{1}{x}} \\ &= \lim_{x\to-\infty} \frac{e^x}{-\frac{1}{x^2}} \\ &= \lim_{x\to-\infty} \frac{e^x}{\frac{2}{x^3}} \\ &= \cdots\end{aligned}$$

We are a bit stuck here. We try converting e^x to $1/(e^{-x})$ instead:

$$\begin{aligned}\lim_{x\to-\infty} xe^x &= \lim_{x\to-\infty} \frac{x}{e^{-x}} \\ &= \lim_{x\to-\infty} \frac{1}{-e^{-x}} \\ &= \lim_{x\to-\infty} -e^x \\ &= 0\end{aligned}$$

Sometimes, the choice of which term in the product to convert makes a difference.

Example. Evaluate

$$\lim_{x\to\infty} \frac{e^x + e^{-x}}{e^x - e^{-x}}$$

Solution. If we try to apply l'Hospital's rule directly, we will run into a cycle:

$$\lim_{x\to\infty} \frac{e^x + e^{-x}}{e^x - e^{-x}} = \lim_{x\to\infty} \frac{e^x - e^{-x}}{e^x + e^{-x}} = \lim_{x\to\infty} \frac{e^x + e^{-x}}{e^x - e^{-x}} = \cdots$$

We must do some preprocessing before applying the rule.

$$\begin{aligned}\lim_{x\to\infty} \frac{e^x + e^{-x}}{e^x - e^{-x}} &= \lim_{x\to\infty} \frac{e^{2x} + 1}{e^{2x} - 1} \\ &= \lim_{x\to\infty} \frac{2e^{2x}}{2e^{2x}} \\ &= 1\end{aligned}$$

Example. Evaluate

$$\lim_{x\to\infty} \frac{x^{1/2} + x^{-1/2}}{x^{1/2} - x^{-1/2}}$$

Solution. Like the previous example, we will run into issues if we apply l'Hospital's rule directly. Unlike the previous example, we will run into a neverending sequence of derivatives instead of a cycle:

$$\lim_{x\to\infty} \frac{x^{1/2} + x^{-1/2}}{x^{1/2} - x^{-1/2}} = \lim_{x\to\infty} \frac{\frac{1}{2}x^{-1/2} - \frac{1}{2}x^{-3/2}}{\frac{1}{2}x^{-1/2} + \frac{1}{2}x^{-3/2}} = \lim_{x\to\infty} \frac{-\frac{1}{4}x^{-3/2} + \frac{3}{4}x^{-5/2}}{-\frac{1}{4}x^{-3/2} - \frac{3}{4}x^{-5/2}} = \cdots$$

We preprocess the expression similarly.

$$\begin{aligned}\lim_{x\to\infty} \frac{x^{1/2}+x^{-1/2}}{x^{1/2}-x^{-1/2}} &= \lim_{x\to\infty} \frac{x+1}{x-1} \\ &= \lim_{x\to\infty} \frac{1}{1} \\ &= 1\end{aligned}$$

The following example concerns a limit in the indeterminate form ∞^0. In general, limits in any indeterminate form involving exponentials (1^∞, 0^∞, 0^0) can be tackled in a similar way, so we will only do one example.

Example. Evaluate

$$\lim_{x\to\infty} x^{1/x}$$

Solution. When we see any weird exponentials in differentiation, the first thing that should come to mind is implicit differentiation and taking the log of both sides. This is exactly what we are going to do.

Let

$$y = x^{1/x} \quad \Rightarrow \quad \ln y = \frac{1}{x} \ln x = \frac{\ln x}{x}$$

Then

$$\begin{aligned}\lim_{x\to\infty} \ln y &= \lim_{x\to\infty} \frac{\ln x}{x} \\ &= \lim_{x\to\infty} \frac{\frac{1}{x}}{1} \\ &= 0\end{aligned}$$

Observe that

$$y = e^{\ln y}$$

So

$$\begin{aligned}\lim_{x\to\infty} y &= \lim_{x\to\infty} e^{\ln y} \\ &= e^{\left(\lim_{x\to\infty} \ln y\right)} \\ &= e^0 \\ &= 1\end{aligned}$$

Chapter 5

Series

5.1 Definitions

Definition 5.1. A series is defined to be an infinite sum of real numbers:

$$\sum_{i=1}^{\infty} a_i$$

This definition isn't very insightful. Instead,

Definition 5.2. For a series (an infinite series) $\sum_{i=1}^{\infty} a_i$, the partial sum S_n is defined as

$$S_n = \sum_{i=1}^{n} a_i$$

We will use partial sums to define the convergence and divergence of a series.

Definition 5.3. A sequence converges to some $l \in \mathbb{R}$ if and only if

$$\lim_{n \to \infty} S_n = l$$

where $(S_n)_{n \geq 1}$ represents the sequence of partial sums.

A sequence diverges if it does not converge to some $l \in \mathbb{R}$.

It is very important to note the differences in the definitions of convergence and divergence for sequences and series:

Limit	$l \in \mathbb{R}$	∞	$-\infty$	No limit
Sequence a_n	Converges	Converges	Converges	Diverges
Series $\sum a_n$	Converges	Diverges	Diverges	Diverges

Nobody knows why we are defining them like this. Anyway, we proceed.

5.2 Tests

The following section is not strictly laid out according to the slides, but how a student outside of DoC would most likely be learning them.

5.2.1 Divergence test

Test 5.4 (Divergence test).

$$\lim_{n\to\infty} a_n \neq 0 \Rightarrow \sum_{n=1}^{\infty} a_n \text{ diverges}$$

In the notes, the contrapositive of this test is stated instead (and it is not stated as a test). Note that unlike in sequences, if a series does not diverge, then it converges to some $l \in \mathbb{R}$, which enables this contrapositive statement in the slides. We will prove the contrapositive here.

Proof. We want to prove that

$$\sum_{n=1}^{\infty} a_n \text{ converges} \Rightarrow \lim_{n\to\infty} a_n = 0$$

Assume $\sum_{n=1}^{\infty} a_n$ converges. Then the sequence S_n converges to some limit $l \in \mathbb{R}$. So S_n is a Cauchy sequence.

Given some $\varepsilon > 0$, using the $\varepsilon - N$ definition of convergence, we want find $N \in \mathbb{N}$ such that for all $n > N$, $|a_n| < \varepsilon$. Since S_n is a Cauchy sequence,

$$\exists N' \in \mathbb{N} : \forall n, m > N, |S_m - S_n| < \varepsilon$$

We pick $m = n+1$. Then $|S_{n+1} - S_n| < \varepsilon$. Since $S_{n+1} - S_n = a_{n+1}$, we have $|a_{n+1}| < \varepsilon$ for all $n > N'$. So let $N = N' + 1$. Then for all $n > N, n - 1 > N'$, so $|a_n| < \varepsilon$. □

5.2.2 Comparison test

We write the following notations:

- $\sum_{i=1}^{\infty} c_i$ represents a (c)onverging series.
- $\sum_{i=1}^{\infty} d_i$ represents a (d)iverging series.

Test 5.5 (Comparison test). If there exists $N \in \mathbb{N}$ such that $\forall i > N$,

- $a_i \leq c_i$, then $\sum_{i=1}^{\infty} a_i$ converges.
- $d_i \geq a_i$, then $\sum_{i=1}^{\infty} a_i$ diverges.

Note that there is no need for every term of $a_i \leq c_i$, or $a_i \geq d_i$ – we just need to show that *after some point*, all terms in the sequence will fulfill the requirement. This enables us to apply the comparison test to determine convergence of series that we would otherwise not be able to determine using the more commonly seen version, which necessitates that *every* term fulfill the requirement.

Example. (MMT 2, 1(b), adapted) Use the comparison test to determine whether the following series converges or diverges:

$$\sum_{n=1}^{\infty} \frac{4}{3n^2 - 4}$$

Solution. Let $a_n = \frac{4}{5n^2-4}$ and $c_n = \frac{2}{n^2}$.

$$\begin{aligned} a_n \leq c_n &\Leftrightarrow \frac{4}{3n^2-4} \leq \frac{2}{n^2} \\ &\Leftrightarrow 4n^2 \leq 6n^2 - 4 \\ &\Leftrightarrow 2n^2 \geq 4 \end{aligned}$$

So $\forall n \geq 2$, $2n^2 \geq 4$, $a_n \leq c_n$.

$$\sum_{n=1}^{\infty} c_n = \sum_{n=1}^{\infty} \frac{2}{n^2} = 2\sum_{n=1}^{\infty} \frac{1}{n^2}$$

Since $\sum \frac{1}{n^2}$ converges, $\sum_{n=1}^{\infty} c_n$ also converges. So $\sum_{n=1}^{\infty} a_n$ converges.

Example. (MMT 2, 1(c)) Use the comparison test to determine whether the following series converges or diverges:

$$\sum_{n=1}^{\infty} \frac{1}{n-4}$$

Solution. Let $a_n = \frac{1}{n-4}$ and $d_n = \frac{1}{n}$.

$$\begin{aligned} d_n \leq a_n &\Leftrightarrow \frac{1}{n} \leq \frac{1}{n-4} \\ &\Leftrightarrow n - 4 \leq n \\ &\Leftrightarrow -4 \leq 0 \text{ which is true } \forall n. \end{aligned}$$

So $\forall n \geq 1$, $d_n \leq a_n$. Since d_n diverges, a_n diverges.

5.2.3 Limit comparison test

Test 5.6 (Limit comparison test)**.**

$$\lim_{n\to\infty} \frac{a_n}{c_n} = l \in \mathbb{R} \Rightarrow \sum_{n=1}^{\infty} a_n \text{ converges.}$$

$$\lim_{n\to\infty} \frac{d_n}{a_n} = l \in \mathbb{R} \Rightarrow \sum_{n=1}^{\infty} a_n \text{ diverges.}$$

Note that in this statement of the test, a_n is in the numerator when we want to show convergence, but a_n is in the denominator when we want to show divergence. This is very odd as putting a_n in the denominator does not change anything, but we will accept it for the purposes of passing this module. The more commonly seen version of the test is as follows:

Test 5.7 (Limit comparison test, better)**.** If $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ are series such that

$$\lim_{n\to\infty} \frac{a_n}{b_n} = l \in \mathbb{R}$$

then $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ both converge or both diverge.

Example. (MMT 2, 2(c)) Use the limit comparison test to determine whether the following series converges or diverges:

$$\sum_{n=1}^{\infty} \frac{1}{3n^2 + 4n - 2}$$

Solution. Let $a_n = \frac{1}{3n^2+4n-2}$ and $c_n = \frac{1}{3n^2}$.

$$\begin{aligned}
\lim_{n\to\infty} \frac{a_n}{c_n} &= \frac{\frac{1}{3n^2+4n-2}}{\frac{1}{3n^2}} \\
&= \lim_{n\to\infty} \frac{3n^2}{3n^2 + 4n - 2} \\
&= \lim_{n\to\infty} \frac{3}{3 + \frac{4}{n} - \frac{2}{n^2}} \\
&= \frac{3}{3 + 0 - 0} \\
&= 1
\end{aligned}$$

Also,

$$\sum_{n=1}^{\infty} c_n = \sum_{n=1}^{\infty} \frac{1}{3n^2} = \frac{1}{3} \sum_{n=1}^{\infty} \frac{1}{n^2}$$

Since $\sum_{n=1}^{\infty} \frac{1}{n^2}$ converges, $\sum_{n=1}^{\infty} c_n$ also converges. Since $\lim\limits_{n\to\infty} \frac{a_n}{c_n} \in \mathbb{R}$ exists and $\sum_{n=1}^{\infty} c_n$ converges, a_n converges.

5.2.4 D'Alembert's (limit) ratio test

Test 5.8 (D'Alembert's ratio test)**.** If there exists some $N \in \mathbb{N}$ such that

- $\frac{a_{n+1}}{a_n} \geq 1$ for all $n \geq N$, then $\sum_{n=1}^{\infty} a_n$ diverges.
- there exists some $k \in \mathbb{R}$ such that $k < 1$ and $\frac{a_{n+1}}{a_n} \leq k$ for all $n \geq N$, then $\sum_{n=1}^{\infty} a_n$ converges.

This test isn't very useful, but it can be used to prove the correctness of the more useful *limit* ratio test:

Test 5.9 (D'Alembert's limit ratio test)**.** Suppose $\lim\limits_{n\to\infty} \frac{a_{n+1}}{a_n} = l \in \mathbb{R}$.

- $l > 1 \Rightarrow \sum_{n=1}^{\infty} a_n$ diverges.
- $l < 1 \Rightarrow \sum_{n=1}^{\infty} a_n$ converges.
- $l = 1 \Rightarrow$ inconclusive.

For $l = 1$, consider $a_n = \frac{1}{n}$ and $b_n = \frac{1}{n^2}$. The limits $\lim\limits_{n\to\infty} \frac{a_{n+1}}{a_n}$ and $\lim\limits_{n\to\infty} \frac{b_{n+1}}{a_b}$ are both 1, but a_n diverges and b_n converges.

Also, it is very painful that we do not introduce the absolute value version (i.e. the correct version) of d'Alembert's (limit) ratio test, since as we will show later, absolute convergence implies unconditional convergence. I am of the strong opinion that this non-absolute value version is wrong, but for the purposes of passing this module, this incorrect version is in this set of notes, as aligned with the slides. The correct version is stated in the next section.

Example. (MMT 2, 3(d)) Using d'Alembert's limit ratio test and $\lim_{n\to\infty}(1+\frac{1}{n})^n = e$, determine whether the following series converges or diverges:

$$\sum_{n=1}^{\infty} \frac{n!}{n^n}$$

Solution. Let $a_n = \frac{n!}{n^n}$. Since $a_n > 0$ for all $n > 0$, we can apply d'Alembert's limit ratio test.

$$\begin{aligned}
\lim_{n\to\infty} \frac{a_{n+1}}{a_n} &= \lim_{n\to\infty} \frac{(n+1)!}{(n+1)^{n+1}} \cdot \frac{n^n}{n!} \\
&= \lim_{n\to\infty} \frac{(n+1)!}{n!} \cdot \frac{n^n}{(n+1)^{n+1}} \\
&= \lim_{n\to\infty} (n+1) \cdot \frac{n^n}{(n+1)^{n+1}} \\
&= \lim_{n\to\infty} \frac{n^n}{(n+1)^n} \\
&= \lim_{n\to\infty} (\frac{n}{n+1})^n \\
&= \lim_{n\to\infty} \frac{1}{(1+\frac{1}{n})^n} \\
&= \frac{1}{e} \\
&< 1
\end{aligned}$$

So $\sum_{n=1}^{\infty} \frac{n!}{n^n}$ converges.

5.2.5 Integral test

Test 5.10 (Integral test). Let $f : \mathbb{R} \to \mathbb{R}$ be continuous, decreasing, and positive on the interval $[1, \infty)$, such that $a_n = f(n)$. If there exists $N \in \mathbb{N}$ such that

- $\int_N^\infty f(x)\,\mathrm{d}x$ converges, then $\sum_{n=1}^{\infty} a_n$ converges.
- $\int_N^\infty f(x)\,\mathrm{d}x$ diverges, then $\sum_{n=1}^{\infty} a_n$ diverges.

Example. We consider the upper Riemann approximation to show that $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges.

It is clear from Figure 5.1 that $\sum_{n=1}^{\infty} \frac{1}{n} > \int_1^\infty \frac{1}{n}$. Then,

$$\begin{aligned}
\int_1^\infty \frac{1}{x}\,\mathrm{d}x &= \lim_{b\to\infty} [\ln x]_1^b \\
&= \lim_{b\to\infty} [\ln b - \ln 1] \\
&= \lim_{b\to\infty} \ln b
\end{aligned}$$

which diverges since $\ln x$ is strictly increasing and not bounded above. So $\sum_{n=1}^{\infty} \frac{1}{n}$ also diverges.

Example. We consider the lower Riemann approximation to show that $\sum_{n=2}^{\infty} \frac{1}{n^2}$ converges (note that we are summing from 2, not 1).

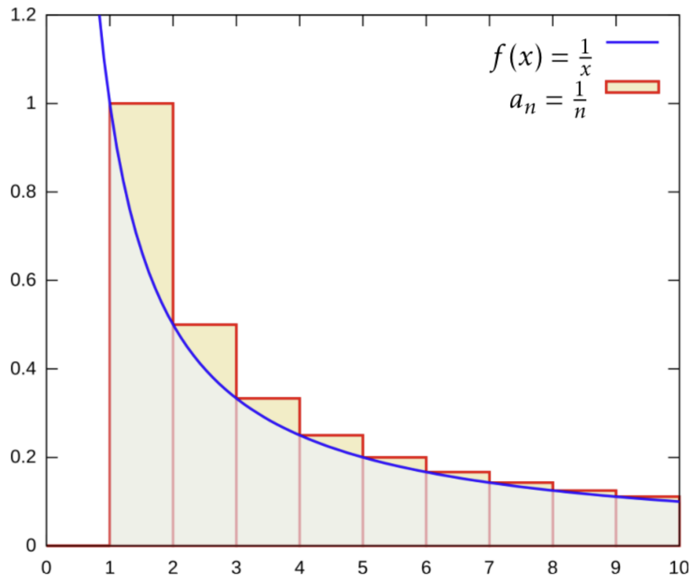


Figure 5.1: Upper Riemann approximation of $\frac{1}{n}$

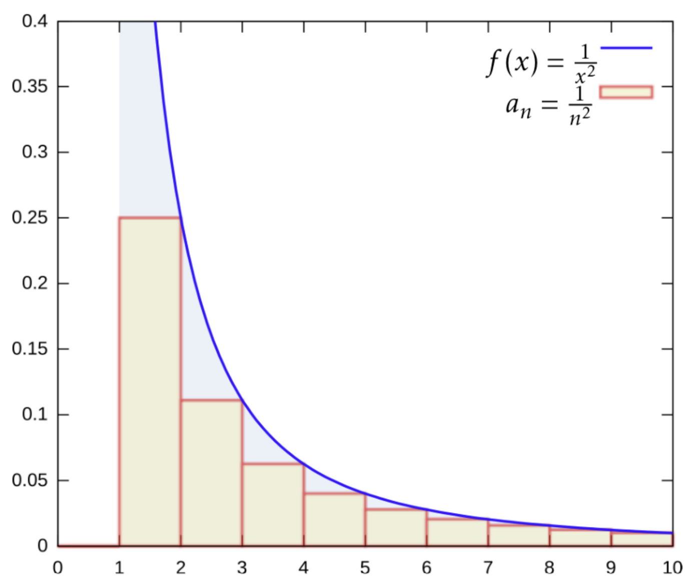


Figure 5.2: Lower Riemann approximation of $\frac{1}{n^2}$

It is clear from Figure 5.2 that $\sum_{n=2}^{\infty} \frac{1}{n} < \int_1^{\infty} \frac{1}{n^2}$. Then,

$$\begin{aligned}
\int_1^{\infty} \frac{1}{x^2}\,\mathrm{d}x &= \lim_{b\to\infty} \left[-\frac{1}{x}\right]_1^b \\
&= \lim_{b\to\infty} (-\frac{1}{b} + \frac{1}{1}) \\
&= \lim_{b\to\infty} (1 - \frac{1}{b}) \\
&= 1 - 0 \\
&= 1
\end{aligned}$$

so the integral converges. Then $\sum_{n=2}^{\infty} \frac{1}{n^2}$ also converges, so $\sum_{n=1}^{\infty} \frac{1}{n^2}$ converges.

Example. (MMT 2, 4) Use the integral test to determine whether the following series converges or diverges:

$$\sum_{n=1}^{\infty} \frac{1}{e^n}$$

Solution. Let $f(x) = \frac{1}{e^x}$. $f(x)$ is continuous, decreasing, and positive on the interval $[1, \infty)$.

$$\begin{aligned}
\int_1^{\infty} f(x)\,\mathrm{d}x &= \lim_{b\to\infty} \int_1^b \frac{1}{e^x}\,\mathrm{d}x \\
&= \lim_{b\to\infty} \left[-\frac{1}{e^x}\right]_1^b \\
&= \lim_{b\to\infty} (-\frac{1}{e^b} + \frac{1}{e}) \\
&= 0 + \frac{1}{e} \\
&= \frac{1}{e}
\end{aligned}$$

so the integral converges. So $\sum_{n=1}^{\infty} \frac{1}{e^n}$ also converges.

5.3 Absolute convergence

Before we write down the correct version of the (limit) ratio test, we need to define a few things.

Definition 5.11 (Absolute convergence)**.** A series $\sum_{n=1}^{\infty} a_n$ is absolutely convergent if and only if

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

converges also.

Definition 5.12 (Unconditional convergence)**.** We define a permutation π over the natural numbers by $\pi : \mathbb{N} \to \mathbb{N}$, such that π is bijective. A series $\sum_{n=1}^{\infty} a_n$ is unconditionally convergent if and only if it converges, and

$$\sum_{n=1}^{\infty} a_{\pi(n)}$$

converges to the same limit, for all $\pi : \mathbb{N} \to \mathbb{N}$.

Lemma 5.13. Absolute convergence implies (unconditional) convergence.

5.3.1 Absolute value comparison test

Test 5.14 (Absolute value comparison test). Let b_n be a non-decreasing sequence such that $\sum_{n=1}^{\infty} b_n$ converges. Let a_n be a sequence such that $|a_n| \leq b_n$ for all $n \geq 1$. Then $\sum_{n=1}^{\infty} a_n$ converges.

Proof. Since $|a_n| \leq b_n$ for all $n \geq 1$, we know that b_n is a non-negative, non-decreasing (given) sequence. Let S_n and S'_n denote the partial sums for $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$, respectively. Since $\sum_{n=1}^{\infty} b_n$ converges, by contrapositive of divergence test, $\lim\limits_{n\to\infty} b_n = 0$. Since the sequence b_n converges to some real number, S'_n is a Cauchy sequence. Then given some $\varepsilon > 0$, there exists some $N \in \mathbb{N}$ such that for all $m, n > N$, $|S'_m - S'_n| < \varepsilon$. We let $m \geq n > N$. Since b_n is non-decreasing, we have $S'_m \geq S'_n$, so $S'_m - S'_n < \varepsilon$.

$$\begin{aligned}
|S_m - S_n| &= \left| \sum_{i=n+1}^{m} a_i \right| \\
&\leq \sum_{i=n+1}^{m} |a_i| \text{ by triangle inequality} \\
&\leq \sum_{i=n+1}^{m} b_i \\
&= S'_m - S'_n \\
&< \varepsilon
\end{aligned}$$

So S_n is a Cauchy sequence as well, and converges to some $l \in \mathbb{R}$. So $\sum_{n=1}^{\infty} a_n$ converges. □

5.3.2 D'Alembert's (limit) ratio test, correct

Test 5.15 (D'Alembert's (limit) ratio test, correct). Suppose $\lim\limits_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| = l$.

- $l > 1 \Rightarrow \sum_{n=1}^{\infty} a_n$ diverges.
- $l < 1 \Rightarrow \sum_{n=1}^{\infty} a_n$ converges absolutely, so by lemma, converges unconditionally.
- $l = 1 \Rightarrow$ inconclusive.

Proof. $\mathbf{l > 1.}$ By $\varepsilon - N$ definition of convergence, given some $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $n > N$,

$$\left| \left|\frac{a_{n+1}}{a_n}\right| - l \right| < \varepsilon$$

Suppose $\varepsilon = \frac{l-1}{2}$. Since $l > 1$, $\varepsilon > 0$. Then we have

$$\begin{aligned}\left|\left|\frac{a_{n+1}}{a_n}\right| - l\right| < \frac{l-1}{2} &\Leftrightarrow \frac{1-l}{2} < \left|\frac{a_{n+1}}{a_n}\right| - l < \frac{l-1}{2} \\ &\Leftrightarrow \frac{1+l}{2} < \left|\frac{a_{n+1}}{a_n}\right| < \frac{3l-1}{2}\end{aligned}$$

So

$$\begin{aligned}\left|\frac{a_{n+1}}{a_n}\right| &> \frac{1+l}{2} \\ &> 1 \text{ since } l > 1\end{aligned}$$

Let $c = \frac{1+l}{2}$. Then $c > 1$.

$$\begin{aligned}\left|\frac{a_{n+1}}{a_n}\right| > c &\Leftrightarrow \frac{|a_{n+1}|}{|a_n|} > c \\ &\Leftrightarrow |a_{n+1}| > c|a_n| \\ &\Rightarrow |a_{m+N+2}| > c^m|a_{N+1}| \text{ for all } m \geq 1\end{aligned}$$

Since $c > 1$, the geometric sequence $c^m|a_n|$ converges to ∞, so the sequence a_n also converges to ∞. (Sound reasoning?) By divergence test, since $\lim\limits_{n\to\infty} |a_n| \neq 0$, $\sum_{n=1}^{\infty} |a_n|$ diverges. So $\sum_{n=1}^{\infty} a_n$ diverges.

l < 1. Given some $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $n > N$,

$$\left|\left|\frac{a_{n+1}}{a_n}\right| - l\right| < \varepsilon$$

Suppose $\varepsilon = \frac{1-l}{2}$. Since $l < 1$, $\varepsilon > 0$. Then we have

$$\begin{aligned}\left|\left|\frac{a_{n+1}}{a_n}\right| - l\right| < \frac{1-l}{2} &\Leftrightarrow \frac{l-1}{2} < \left|\frac{a_{n+1}}{a_n}\right| - l < \frac{1-l}{2} \\ &\Leftrightarrow \frac{3l-1}{2} < \left|\frac{a_{n+1}}{a_n}\right| < \frac{1+l}{2}\end{aligned}$$

So

$$\begin{aligned}\left|\frac{a_{n+1}}{a_n}\right| &< \frac{1+l}{2} \\ &< 1 \text{ since } l < 1\end{aligned}$$

Let $c = \frac{1+l}{2}$. Then $c < 1$.

$$\begin{aligned}\left|\frac{a_{n+1}}{a_n}\right| < c &\Leftrightarrow |a_{n+1}| < c|a_n| \\ &\Rightarrow |a_{m+N+2}| < c^m|a_{N+1}| \text{ for all } m \geq 1\end{aligned}$$

Since $c < 1$, the geometric series

$$\sum_{n=1}^{\infty} c^m|a_{N+1}| = |a_{N+1}| \sum_{n=1}^{\infty} c^m$$

converges. Then by comparison,

$$\sum_{n=N+3}^{\infty} |a_n|$$

converges as well, so $\sum_{n=1}^{\infty} |a_n|$ converges, and $\sum_{n=1}^{\infty} a_n$ converges too. □

5.3.3 n^{th} root test

Definition 5.16 (Limit superior). Let $(a_n)_{n\geq 1}$ be a sequence. Let $(b_n)_{n\geq 1}$ be a sequence defined by

$$b_n = \sup\{a_m \mid m \geq n\}$$

Then

$$\limsup_{n\to\infty} a_n = \lim_{n\to\infty} b_n$$

Semantically, b_n represents the supremum of the set of all the terms after and including a_n. Note that b_n is a decreasing (non-increasing) sequence, since we may be removing the maximum from the set, in which case the supremum may decrease; otherwise, the supremum remains unchanged.

Definition 5.17 (Limit inferior). Let $(a_n)_{n\geq 1}$ be a sequence. Let $(b_n)_{n\geq 1}$ be a sequence defined by

$$b_n = \inf\{a_m \mid m \geq n\}$$

Then

$$\liminf_{n\to\infty} a_n = \lim_{n\to\infty} b_n$$

Note that b_n is an increasing (non-decreasing) sequence, since we may be removing the smallest number in the set, and therefore increase the infimum; or we may not be removing the smallest number, in which case the infimum remains unchanged.

Lemma 5.18. For any sequence $(a_n)_{n\geq 1}$, there exists a subsequence $(a_{n_i})_{i\geq 1}$ such that

$$\limsup_{n\to\infty} a_n = \lim_{n\to\infty} a_{n_i}$$

Proof. Warning: this proof is fairly involved and reasonably long.

The cases where the limit equals ∞ or $-\infty$ are trivial. This proof will focus on the case when the limit is equal to some $t \in \mathbb{R}$. Recall the following definition from the chapter on sequences:

Definition 5.19. a_m is a peak if and only if $a_m > a_n \ \forall n > m$.

Case 1: there are infinitely many peaks. Suppose these peaks are at $n_1 < n_2 < \ldots$. By definition of peak,

$$a_{n_1} > a_{n_2} > \ldots$$

so we have

$$a_{n_k} = \sup\{a_n \mid n \geq n_k\}$$

Let $b_k = a_{n_k}$.

$$\begin{aligned}\lim_{k\to\infty} b_k &= \lim_{n_k\to\infty} \sup\{a_n \mid n \geq n_k\}\\ &= \lim_{n\to\infty} \sup\{a_m \mid m \geq n\}\\ &= \limsup_{n\to\infty} a_n, \text{ by definition}\end{aligned}$$

Case 2: there are finitely many peaks. Suppose the last peak is at N_0.

Proposition 5.20.

$$\forall n > N_0, b_n = \sup\{a_m \mid m > n\} = t$$

In other words, the sequence b_n is eventually constant after a certain point N_0.

Proof. Assume b_n is not eventually constant. Since b_n is non-increasing and it is not eventually constant, b_n is strictly decreasing. Then there exists $N > N_0$ such that

$$b_N > b_{N+1} \Leftrightarrow \sup\{b_m \mid m > N\} > \sup\{b_m \mid m > N+1\}$$

Since the LHS set has only one extra element b_{N+1} than the RHS set,

$$b_{N+1} > \sup\{b_m \mid m > N+1\}$$

so b_{N+1} is greater than all of $b_{N+1}, b_{N+2}, \ldots$, so b_{N+1} is a peak. However, we picked N such that $N > N_0$, so $N+1$ is past the last peak, contradiction. So b_n is eventually constant. Since $\lim_{n\to\infty} b_n = t$, b_n is eventually constant and equal to t. □

Case (a): there are infinitely many terms = t. Construct a subsequence from these terms, done.

Case (b): there are finitely many terms = t. Suppose the last term equal to t (if such term exists) is at N'. Then we pick N_1 such that $N_1 > \max(N', N_0)$, so N_1 is beyond both the last peak and the last term equal to t. It is easy to check that

$$\forall n > N_1,\ a_n < t$$

It cannot be the case that $a_n = t$ since n is beyond the last term equal to t, and it cannot be the case that $a_n > t$ since from the above proposition, we get

$$a_n > t = \sup\{a_m \mid m \geq n\}$$

which contradicts the definition of supremum. As every term is strictly less than t, the only hope for an infinite subsequence that converges to t is one that is increasing.

Proposition 5.21. There exists a subsequence a_{n_k} such that $\forall k$,

$$t - \frac{1}{k} < a_{n_k} < t$$

Proof. We construct this subsequence inductively. Clearly, we want to pick n_k such that $n_k > N_1$, as $a_{n_k} < t$ immediately follows from above.

For the base case, we have

$$a_{n_1} > t - 1$$

which is true as by the definition of supremum, there must exist $n_1 > N_1$ which satisfies the inequality. (Intuitively, since the lim sup is t, there must exist some term that is arbitrarily close to t.)

For the inductive case, suppose we have constructed a finite increasing subsequence with $a_{n_1} > t-1$, $a_{n_2} > t - \frac{1}{2}$, ..., $a_{n_k} > t - \frac{1}{k}$. We want to show two things:

- $n_{k+1} > t - \frac{1}{k+1}$
- $a_{n_k} < a_{n_{k+1}}$ and $n_k < n_{k+1}$, i.e. the sequence is increasing

For the first point, using the same argument as above, there exists some $n_{k+1} > N_1$ such that n_{k+1} is arbitrarily close to t. Since $\frac{1}{k+1}$ is finite and positive, the inequality is satisfied.

For the second point,

$$\max(a_{N_1+1}, a_{N_1+2}, ..., a_{n_k}) < a_{n_{k+1}}$$

using the same arbitrarily close argument. Since $a_{n_{k+1}}$ is greater, and therefore different from all the other terms after N_1, we have $n_{k+1} > n_k$. □

By the sandwich theorem, since the lower and upper bounds both approach t as $k \to \infty$, we have

$$\lim_{k\to\infty} a_{n_k} = t$$

as required. □

Proposition 5.22. A sequence a_n converges to $l \in \mathbb{R}$ if and only if

$$\limsup_{n\to\infty} a_n = \liminf_{n\to\infty} = l$$

Proof. If. Suppose

$$\limsup_{n\to\infty} a_n = \liminf_{n\to\infty} = l$$

Let

$$b_n = \sup\{a_m \mid m \geq n\}, \qquad c_n = \inf\{a_m \mid m \geq n\}$$

By definition of infimum and supremum, for all $n \in N$,

$$c_n \leq a_n \leq b_n$$

so by squeeze theorem, $a_n \to l$ as $n \to \infty$.

Only if. Conversely, suppose $a_n \to l$ as $n \to \infty$. Given some $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $n > N$,

$$l - \varepsilon < a_n < l + \varepsilon$$

(Intuitively, this means every term beyond N is within ε of the limit l.) It follows that

$$l - \varepsilon \leq c_n \leq b_n \leq l + \varepsilon$$

(since loosely speaking, the infimum c_n and the supremum b_n are both arbitrarily close to some term in the sequence a_n, which is within the finite and positive distance of ε from the limit.) So we have

$$|c_n - l| < \varepsilon \qquad\qquad |b_n - l| < \varepsilon$$

and both $c_n \to l$ and $b_n \to l$. So

$$\limsup_{n\to\infty} a_n = \liminf_{n\to\infty} a_n = l$$

as required. □

Test 5.23 (n^{th} root test). Suppose $\limsup_{n\to\infty} |a_n|^{\frac{1}{n}} = l$.

- $l > 1 \Rightarrow$ diverges.
- $l < 1 \Rightarrow$ converges.
- $l = 1 \Rightarrow$ inconclusive: may converge absolutely, converge, or diverge.

Proof. **$\mathbf{l > 1}$.** Since $\limsup_{n\to\infty} |a_n|^{\frac{1}{n}} = l$, there exists a subsequence a_{n_i} such that

$$\lim_{n\to\infty} |a_{n_i}|^{\frac{1}{n_i}} = l$$

Then given some $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $\forall n_i > N \Leftrightarrow i > N$,

$$\left| |a_{n_i}|^{\frac{1}{n_i}} - l \right| < \varepsilon$$

Suppose $\varepsilon = \frac{l-1}{2}$. Since $l > 1$, $\varepsilon > 0$.

$$\begin{aligned} \left| |a_{n_i}|^{\frac{1}{n_i}} - l \right| < \frac{l-1}{2} &\Leftrightarrow \frac{1-l}{2} < |a_{n_i}|^{\frac{1}{n_i}} - l < \frac{l-1}{2} \\ &\Leftrightarrow \frac{1+l}{2} < |a_{n_i}|^{\frac{1}{n_i}} < \frac{3l-1}{2} \end{aligned}$$

Let $c = \frac{1+l}{2}$. Then $c > 1$.

$$|a_{n_i}|^{\frac{1}{n_i}} > c \Leftrightarrow |a_{n_i}| > c^{n_i}$$

Since the sequence c^{n_i} converges to ∞, the sequence $|a_{n_i}|$ also converges to ∞, so the sequence $|a_n|$ also converges to ∞. Then $\lim\limits_{n\to\infty} |a_n| \neq 0$, so $\sum_{n=1}^{\infty} |a_n|$ diverges, and $\sum_{n=1}^{\infty} a_n$ diverges.

$\mathbf{l < 1}$. Consider the sequence b_n defined by

$$b_n = \sup\{|a_m|^{\frac{1}{m}} \mid m \geq n\}$$

By definition of limit superior, $\lim\limits_{n\to\infty} b_n = l$. Then given some $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $\forall n > N$,

$$\begin{aligned} |b_n - l| < \varepsilon &\Leftrightarrow l - \varepsilon < b_n < l + \varepsilon \\ &\Leftrightarrow l - \varepsilon < \sup\{|a_m|^{\frac{1}{m}} \mid m \geq n\} < l + \varepsilon \\ &\Rightarrow |a_n|^{\frac{1}{n}} < l + \varepsilon \end{aligned}$$

Let $\varepsilon = \frac{l-1}{2}$, and let $c = l + \varepsilon = \frac{l+1}{2}$. Since $l < 1$, $c < 1$.

$$|a_n|^{\frac{1}{n}} < c \Leftrightarrow |a_n| < c^n$$

Since $\sum_{n=1}^{\infty} c^n$ converges, $\sum_{n=1}^{\infty} |a_n|$ also converges, so $\sum_{n=1}^{\infty} a_n$ converges.

l = 1. (Finish this off when done with L'Hopital and limits.) $\square$

Example. Use the n^{th} root test to determine if the following series converges or diverges:

$$\sum_{n=1}^{\infty} (\frac{n}{n+1})^{n^2} 4^n$$

Solution. Let

$$a_n = (\frac{n}{n+1})^{n^2} 4^n$$

$$\begin{aligned}
\limsup_{n\to\infty} |a_n|^{\frac{1}{n}} &= \limsup_{n\to\infty} (\frac{n}{n+1})^n \cdot 4 \\
&= \limsup_{nto\infty} \frac{4}{(1+\frac{1}{n})^n} \\
&= \frac{4}{e} \\
&> 1
\end{aligned}$$

so the series diverges.

Chapter 6

Power series

6.1 What is a power series?

Power series is, in some sense, a generalisation of the geometric series

$$\sum_{n=1}^{\infty} x^n$$

by introducing additional parameters that allow us to control the behaviour of the series. A power series is in the form of

$$\sum_{n=0}^{\infty} a_n (x-c)^n$$

where the two extra parameters are

- a_n, a variable weight on each power of x
- c, a constant by which we shift the series

6.1.1 What do we want to know about power series?

For the geometric series

$$\sum_{n=1}^{\infty} x^n$$

we note that it converges if and only if $|x| < 1$. For which values of x does a given power series converge?

When $|x| < 1$, the geometric series above converges to the function

$$f(x) = \frac{1}{1-x}$$

i.e. the function can be "represented" by the geometric series. What are the necessary and sufficient conditions for a power series to represent a function $f : (a, b) \to \mathbb{R}$? In other words, what are the necessary and sufficient conditions for a power series to converge to that function on the interval (a, b)?

6.1.2 Polynomials

In general, we can express a polynomial of degree n as

$$p(x) = a_0 + a_1 \cdot x + \cdots + a_n \cdot x^n$$

where $a_n \neq 0$. We can define a power series where $c = 0$ and $a_m = 0\ \forall m > n$, and we match the rest of the coefficients. Since n is finite, $p(x)$ is finite as well, so the power series converges. Clearly, the power series converges to $p(x)\ \forall x \in \mathbb{R}$.

6.2 Radius of convergence

First, we prove that every power series has a radius of convergence.

Theorem 6.1. Let

$$\sum_{n=0}^{\infty} a_n (x-c)^n$$

be a power series. There exists $0 \le R \le \infty$ such that the series converges absolutely for $0 \le |x-c| < R$ and diverges for $|x-c| > R$.

Proof. WLOG, assume $c = 0$. We can directly replace x by $x - c$ in the following steps, so this assumption is just to save some typing (or writing).

Trivially, when $|x| = 0$, each term of the power series is equal to 0, so the whole power series converges to 0.

Otherwise, suppose the power series converges for some $x_0 > 0$. Then the corresponding sequence $a_n x_0^n$ must also converge (to 0), and by Proposition 1.17, it must be bounded. Then there exists M such that

$$|a_n x_0^n| \le M \quad \forall n \in \mathbb{N}$$

Take arbitrary x such that $0 < |x| < |x_0|$. Let

$$r = \left| \frac{x}{x_0} \right|$$

Clearly $0 < r < 1$. Then

$$\begin{aligned} |a_n x^n| &= |a_n x_0^n| \, r^n \\ &\le M r^n \end{aligned}$$

Since $|r| < 1$, the series $\sum_{n=0}^{\infty} M r^n$ converges. By direct comparison, the series

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

converges absolutely. Since x was arbitrary, the series converges for all x with $|x| < |x_0|$.

Now we define

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

Clearly 0 must be in the set, so $R \ge 0$.

- If $R = 0$, then the power series converges only for $x = 0$. Looking back at what we're trying to prove about R, we only care about what diverges but not what converges, because no x satisfies $0 \le |x| < 0$. Anyway, since the power series only converges for $x = 0$, it diverges for all x with $|x| > 0$.
- If $R > 0$, then the power series converges absolutely for all x such that $|x| < R$, since by definition of supremum, the power series must converge for some x_0 such that $|x| < |x_0| < R$ (if such x_0 doesn't exist, then there must exist an upper bound smaller than R). By construction of R, we also know that the power series must diverge for all x such that $|x| > R$.

- If $R = \infty$, then the power series converges for all x.

$\square$

Now we are ready to state the definition of the radius of convergence.

Definition 6.2. If the power series

$$\sum_{n=0}^{\infty} a_n (x-c)^n$$

converges for $|x-c| < R$ and diverges for $|x-c| > R$, then $0 \leq R \leq \infty$ is called the radius of convergence of the power series.

Remark. The definition says nothing about what happens when $|x-c| = R$. In fact, the power series may either converge absolutely, converge, or diverge at the boundaries.

There are two main ways of finding the radius of convergence.

Theorem 6.3. The radius of convergence R of a power series

$$\sum_{n=0}^{\infty} a_n (x-c)^n$$

is given by

$$R = \lim_{n\to\infty} \left| \frac{a_n}{a_{n+1}} \right| \qquad \text{or} \qquad \frac{1}{R} = \lim_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right|$$

Proof. Let

$$R = \lim_{n\to\infty} \left| \frac{a_n}{a_{n+1}} \right|$$

(We don't know if R is the radius of convergence yet.) We want to apply d'Alembert's ratio test here, so we evaluate:

$$\begin{aligned} \lim_{n\to\infty} \left| \frac{a_{n+1}(x-c)^{n+1}}{a_n (x-c)^n} \right| &= |x-c| \lim_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right| \\ &= \frac{|x-c|}{R} \end{aligned}$$

We know that the power series converges if

$$\frac{|x-c|}{R} < 1 \Leftrightarrow |x-c| < R$$

and diverges if

$$\frac{|x-c|}{R} > 1 \Leftrightarrow |x-c| > R$$

so R is the radius of convergence of the power series. $\square$

Theorem 6.4 (Cauchy-Hadamard theorem)**.** The radius of convergence R of a power series

$$\sum_{n=0}^{\infty} a_n (x-c)^n$$

is given by

$$R = \frac{1}{\limsup_{n\to\infty} |a_n|^{\frac{1}{n}}}$$

Proof. Let

$$R = \frac{1}{\limsup_{n\to\infty} |a_n|^{\frac{1}{n}}}$$

(We don't know if R is the radius of convergence yet.) We want to apply the n^{th} root test to the series:

$$\begin{aligned}\limsup_{n\to\infty} |a_n(x-c)^n|^{\frac{1}{n}} &= |x-c| \limsup_{n\to\infty} |a_n|^{\frac{1}{n}} \\ &= \frac{|x-c|}{R}\end{aligned}$$

We know that the power series converges if

$$\frac{|x-c|}{R} < 1 \Leftrightarrow |x-c| < R$$

and diverges if

$$\frac{|x-c|}{R} > 1 \Leftrightarrow |x-c| > R$$

so R is the radius of convergence of the power series. □

We will go through a few examples to see how we could find the radius of convergence using the above theorems, and why they are probably not the safest way if you don't know what you're doing.

Example. Determine the radius and interval of convergence of the power series

$$\frac{n}{4^n}(x+3)^n$$

Solution.

$$\begin{aligned}R &= \lim_{n\to\infty} \left| \frac{a_n}{a_{n+1}} \right| \\ &= \lim_{n\to\infty} \left| \frac{n}{4^n} \cdot \frac{4^{n+1}}{n+1} \right| \\ &= \lim_{n\to\infty} \left| \frac{n}{n+1} \cdot 4 \right| \\ &= 4\end{aligned}$$

so the radius of convergence is 4. The series converges if

$$|x+3| < 4 \Leftrightarrow -7 < x < 1$$

When $x = -7$, the series becomes

$$\sum_{n=0}^{\infty} \frac{n}{4^n}(-4)^n = \sum_{n=0}^{\infty} n(-1)^n$$

which diverges by the divergence test.

When $x = 1$, the series becomes

$$\sum_{n=0}^{\infty} \frac{n}{4^n} \cdot 4^n = \sum_{n=0}^{\infty} n$$

which diverges by the divergence test since the terms do not tend to 0.

Therefore, the interval of convergence is $(-7, 1)$.

Example. Determine the radius and interval of convergence of the power series

$$\frac{4^n}{n}(x+3)^n$$

Solution.

$$\begin{aligned} R &= \lim_{n\to\infty} \left| \frac{a_n}{a_{n+1}} \right| \\ &= \lim_{n\to\infty} \left| \frac{4^n}{n} \cdot \frac{n+1}{4^{n+1}} \right| \\ &= \lim_{n\to\infty} \left| \frac{n+1}{n} \cdot \frac{1}{4} \right| \\ &= \frac{1}{4} \end{aligned}$$

so the radius of convergence is $\frac{1}{4}$. The series converges if

$$|x+3| < \frac{1}{4} \Leftrightarrow -\frac{13}{4} < x < -\frac{11}{4}$$

When $x = -\frac{13}{4}$, the series becomes

$$\sum_{n=0}^{\infty} \frac{4^n}{n} \left(-\frac{1}{4}\right)^n = \sum_{n=0}^{\infty} \frac{(-1)^n}{n}$$

which converges by the alternating series test (this is known as the alternating harmonic series).

When $x = -\frac{11}{4}$, the series becomes

$$\sum_{n=0}^{\infty} \frac{4^n}{n} \left(\frac{1}{4}\right)^n = \sum_{n=0}^{\infty} \frac{1}{n}$$

which is the harmonic series, so it diverges (e.g. by the integral test).

Therefore, the interval of convergence is $\left[-\frac{13}{4}, -\frac{11}{4}\right)$.

Example. Determine the radius and interval of convergence of the power series

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

Solution.

$$\begin{aligned}
R &= \lim_{n\to\infty}\left|\frac{a_n}{a_{n+1}}\right| \\
&= \lim_{n\to\infty}\left|\frac{n!}{(n+1)!}\right| \\
&= \lim_{n\to\infty}\left|\frac{1}{n+1}\right| \\
&= 0
\end{aligned}$$

so the radius of convergence is 0. Then we know that the power series only converges for $x = 0$, so the interval of convergence is $\{0\}$.

Remark. Intervals are sets.

The last statement doesn't sound too rigorous, does it? How do we suddenly *know* that the series will converge for $x = 0$? Even though we can still plug the terms into the formulae and get them to work, there are better and more robust ways of finding the radius and interval of convergence. We will attempt this example using another approach.

Solution.

$$\begin{aligned}
\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| &= \lim_{n\to\infty}\left|\frac{(n+1)!\,x^{n+1}}{n!\,x^n}\right| \\
&= |x| \lim_{n\to\infty} n
\end{aligned}$$

The power series converges for $x = 0$, and the limit $\to \infty$ for $x \neq 0$. Therefore, the radius of convergence is 0, and the interval of convergence is $\{0\}$.

Note that a_n refers to the whole term of the series here, while in Theorem 6.3 and Theorem 6.4, a_n refers to only the coefficient in front of each term.

Example. Determine the radius and interval of convergence of the power series

$$\sum_{n=0}^{\infty}(2x)^{2n}$$

If we try to find the radius of convergence using Theorem 6.3 then we get

$$\begin{aligned}
R &= \lim_{n\to\infty}\left|\frac{a_n}{a_{n+1}}\right| \\
&= \lim_{n\to\infty}\left|\frac{2^{2n}}{2^{2n+2}}\right| \\
&= \frac{1}{4}
\end{aligned}$$

which is wrong, as we will see later. Recall how we stated Theorem 6.3: notice that we are finding the ratio of successive terms a_n and a_{n+1}, which correspond to the coefficients of x^n and x^{n+1}. However, for this power series, $a_n = 0$ for all odd n. How do we define the ratio then? That's why the approach above is more robust:

Solution.

$$\begin{aligned}\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| &= \lim_{n\to\infty}\left|\frac{(2x)^{2n+2}}{(2x)^{2n}}\right| \\ &= \left|(2x)^2\right| \\ &= 4x^2\end{aligned}$$

The power series converges if

$$4x^2 < 1 \Leftrightarrow x^2 < \frac{1}{4} \Leftrightarrow |x| < \frac{1}{2}$$

so the radius of convergence is $\frac{1}{2}$.

When $x = -\frac{1}{2}$, the series becomes

$$\sum_{n=0}^{\infty}(-1)^{2n} = \sum_{n=0}^{\infty}1$$

which clearly diverges.

When $x = \frac{1}{2}$, the series becomes

$$\sum_{n=0}^{\infty}1^{2n} = \sum_{n=0}^{\infty}1$$

which also diverges.

Therefore, the interval of convergence is $\left(-\dfrac{1}{2}, \dfrac{1}{2}\right)$.

Example. Determine the radius and interval of convergence of

$$\sum_{n=0}^{\infty}(-1)^n\frac{x^{2n+1}}{(2n+1)!}$$

Solution.

$$\begin{aligned}\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| &= \lim_{n\to\infty}\left|\frac{(-1)^{n+1}x^{2n+3}}{(2n+3)!}\cdot\frac{(2n+1)!}{(-1)^n x^{2n+1}}\right| \\ &= x^2\lim_{n\to\infty}\left|\frac{1}{(2n+3)(2n+2)}\right| \\ &= 0 \\ &< 1\end{aligned}$$

so the power series converges for all $x \in \mathbb{R}$. Therefore, the radius of convergence is ∞ and the interval of convergence is $\mathbb{R}$.

6.2.1 Differentiation

Theorem 6.5. Suppose the power series

$$\sum_{n=0}^{\infty}a_n(x-c)^n$$

has radius of convergence R. Then the power series

$$\sum_{n=0}^{\infty} n a_n (x-c)^{n-1}$$

also has radius of convergence R.

Proof. WLOG, assume $c = 0$ and suppose $|x| < R$. Take arbitrary ρ such that $|x| < \rho < R$. Let

$$r = \frac{|x|}{\rho}$$

then we have $0 < r < 1$ for $|x| \neq 0$. (Trivially, the power series must converge for $|x| = 0$.)

Intuition. We took some arbitrary $\rho < R$ so we know that the power series $\sum |a_n \rho^n|$ converges. Ultimately, we want to show that the differentiated power series must converge for all $|x| < \rho$. So we artifically introduce $|a_n \rho^n|$, and find some way to show that the rest of the expression also converges, then the whole series converges.

We rewrite *the absolute value of* each term in the differentiated power series such that we can extract information from the original power series. We take the absolute value of each term so we get $|x|$ somewhere, which allows us to substitute in r.

$$\begin{aligned} \left|n a_n x^{n-1}\right| &= \left|\frac{n a_n \rho^n x^{n-1}}{\rho^n}\right| \\ &= \frac{n}{\rho} \cdot \left(\frac{|x|}{\rho}\right)^{n-1} \cdot |a_n \rho^n| \\ &= \frac{n r^{n-1}}{\rho} |a_n \rho^n| \end{aligned} \tag{6.1}$$

The series $\sum n r^{n-1}$ converges by the ratio test:

$$\lim_{n\to\infty} \frac{(n+1)r^n}{n r^{n-1}} = r \lim_{n\to\infty} \frac{n+1}{n} = r < 1$$

and so it is bounded. Then there exists $M \in \mathbb{R}$ such that for all $n \in \mathbb{N}$,

$$\left|n r^{n-1}\right| \leq M$$

Substituting back into Equation (6.1), we get

$$\frac{n r^{n-1}}{\rho} |a_n \rho^n| \leq \frac{M}{\rho} |a_n \rho^n|$$

Since $\rho < R$, we know that the power series $\sum |a_n \rho^n|$ converges (from Theorem 6.1), so

$$\frac{M}{\rho} |a_n \rho^n|$$

also converges. By comparison test, $\sum \left|n a_n x^{n-1}\right|$ converges. Since $\rho < R$ was arbitrary, we can make ρ as close to R as we want, so the differentiated power series converges for all $|x| < R$.

Now suppose $|x| > R$.

Lemma 6.6. $|a_n x^n|$ is unbounded.

Proof. Suppose $|a_n x^n|$ is bounded. Then there exists $M \in \mathbb{R}$ such that

$$|a_n x^n| \leq M \quad \forall n \in \mathbb{N}$$

Now take arbitrary r such that $|x| > r > R$. Consider

$$\begin{aligned} |a_n r^n| &= |a_n x^n| \cdot \left(\frac{r}{|x|}\right)^n \\ &\leq M \cdot \left(\frac{r}{|x|}\right)^n \end{aligned}$$

We know that the series

$$\sum_{n=0}^{\infty} M \cdot \left(\frac{r}{|x|}\right)^n$$

is a geometric series with common ratio < 1, so it converges. By comparison, $\sum |a_n r^n|$ converges as well. But then it should diverge since $r > R$, so we have a contradiction. Therefore, $|a_n x^n|$ is unbounded. □

Since $|a_n x^n|$ is unbounded by Lemma 6.6, it follows that $\left|n a_n x^{n-1}\right|$ is unbounded as well, since $n \geq 1$. Then the sequence $n a_n x^{n-1}$ does not converge to zero, so the series diverges. □

We'll skip over the proof for the following lemma and take it for granted:

Lemma 6.7. Suppose the power series

$$\sum_{n=0}^{\infty} a_n (x-c)^n$$

has radius of convergence R and converges to $f(x)$ for all $|x-c| < R$. Then f is differentiable in $|x-c| < R$ and

$$f'(x) = \sum_{n=0}^{\infty} n a_n (x-c)^{n-1} \quad \text{for } |x-c| < R$$

If we apply Lemma 6.7 recursively, then we get f is infinitely differentiable in $|x-c| < R$.

6.2.2 Smooth and analytic functions

Definition 6.8 (Smooth functions)**.** A function $f : (a,b) \to \mathbb{R}$ is smooth, or infinitely differentiable, on (a,b) if there exist continuous derivatives of all orders on (a,b), i.e. $f^{(n)}$ exists for all $n \geq 1$.

Its derivatives are defined recursively by

$$f^{(1)} = f', \ f^{(n+1)} = (f^{(n)})'$$

Definition 6.9 (Analytic functions)**.** A function $f : (a,b) \to \mathbb{R}$ is analytic if for every $c \in (a,b)$, there exists a power series with non-zero radius of convergence which converges to f for every x in some neighbourhood of c.

Basically, a function f is not analytic if its corresponding power series only converges at a certain point, i.e. has radius of convergence $= 0$. The power series must converge to f for all the (infinitely many) points that are close enough to the centre of the power series. We will illustrate a well-known function which is smooth but not analytic, but first we will need to prove a lemma:

Lemma 6.10. For all $n \in \mathbb{N}$,

$$\lim_{x\to\infty} \frac{x^n}{e^x} = 0$$

Proof. The Maclaurin series of e^x is

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} > \frac{x^k}{k!}$$

for all $k \in \mathbb{N}$. Put $k = n+1$, we get

$$0 < \frac{x^n}{e^x} < \frac{x^n}{\frac{x^{n+1}}{(n+1)!}} = \frac{(n+1)!}{x}$$

The upper bound converges to 0 since it is just the sequence $\frac{1}{x}$ multiplied by a constant factor. By the sandwich theorem, we have the desired result. □

Now here's the smooth but not analytic function:

Proposition 6.11. Define $\phi(x) : \mathbb{R} \to \mathbb{R}$ by

$$\phi(x) = \begin{cases} e^{\frac{1}{x}}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

Then ϕ has derivatives of all orders at 0 and

$$\phi^{(n)}(0) = 0 \quad \forall n \in \mathbb{N}$$

Proof. We can show (e.g. by induction) that $\phi^{(n)}(x)$ has the form

$$\phi^{(n)}(x) = \begin{cases} p_n\left(\frac{1}{x}\right) e^{\frac{1}{x}}, & x > 0 \\ 0, & x < 0 \end{cases}$$

Then we show by induction that $\phi^{(n)}(0) = 0$.

For the base case, we consider both the left and right derivative at 0. Clearly, $\phi'(0^-) = 0$ since $\phi(0) = 0$ and $\phi(h) = 0$ for all $h < 0$. Now we evaluate the right derivative:

$$\begin{aligned}
\phi'(0^+) &= \lim_{h\to 0^+} \frac{\phi(h) - \phi(0)}{h} \\
&= \lim_{h\to 0^+} \frac{e^{-\frac{1}{h}} - 0}{h} \\
&= \lim_{h\to 0^+} \frac{e^{-\frac{1}{h}}}{h} \\
&= \lim_{x\to\infty} xe^{-x} \\
&= \lim_{x\to\infty} \frac{x}{e^x} \\
&= 0
\end{aligned}$$

Since $\phi'(0^-) = \phi'(0^+) = 0$, we have $\phi'(0) = 0$.

Suppose $\phi^{(k)}(0) = 0$ for some $k \in \mathbb{N}$. Since $\phi^{(k)}(h) = 0$ for all $h < 0$, it follows that $\phi^{(k+1)}(0^-) = 0$. We consider the right derivative:

$$\begin{aligned}
\phi^{(k+1)}(0) &= \lim_{h \to 0^+} \frac{\phi^{(k)}(h) - \phi^{(k)}(0)}{h} \\
&= \lim_{h \to 0^+} \frac{p_k\left(\frac{1}{h}\right) e^{-\frac{1}{h}} - 0}{h} \\
&= \lim_{x \to \infty} \frac{x p_k(x)}{e^x} \\
&= 0
\end{aligned}$$

so we have $\phi^{(k+1)}(0^-) = \phi^{(k+1)}(0^+) = 0$. □

Corollary 6.12. ϕ is smooth but not analytic at 0.

Proof. From Proposition 6.11, we know that ϕ is smooth at 0. However, we note that for the Taylor series of ϕ

$$\sum_{n=0}^{\infty} \frac{\phi^{(n)}(0)}{n!} x^n$$

the coefficient of every power of x is 0, so every term is 0. Then the series converges (sums) to 0 for all x, so the series does not converge to ϕ in any neighbourhood of 0.

Intuition. Even though $\phi(x) = 0$ for all $x < 0$, any neighbourhood of 0 must contain some $x > 0$ as well: as neighbourhoods are open intervals, and in order for any open interval to contain 0, its upper bound must > 0. But then $\phi(x) > 0$ for all $x > 0$, which does not agree with the Taylor series. So there does not exist a neighbourhood of 0 such that the series converges to $\phi(x)$ for every x in the neighbourhood.

□

6.2.3 Maclaurin and Taylor series

Definition 6.13 (Taylor series)**.** The Taylor series of a function f is defined by

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

Definition 6.14 (Maclaurin series)**.** The Maclaurin series of a function is the Taylor series centred at 0.

We'll jump straight into examples.

Example. Find the Taylor series of

$$f(x) = \frac{1}{x}$$

centred around $x = 2$, and its radius of convergence.

Solution.

$$\begin{aligned} f(x) = x^{-1} &\Rightarrow f(2) = \frac{1}{2} \\ f'(x) = -x^{-2} &\Rightarrow f'(2) = -\frac{1}{4} \\ f''(x) = 2x^{-3} &\Rightarrow f''(2) = 2 \cdot \frac{1}{8} \end{aligned}$$

In general,

$$\begin{aligned} f^{(n)}(x) &= (-1)^n \cdot n!\, x^{-(n+1)} \quad \text{for all } n \geq 0 \\ \Rightarrow f^{(n)}(2) &= (-1)^n \cdot n! \cdot \frac{1}{2^{n+1}} \end{aligned}$$

So we have

$$\begin{aligned} f(x) &= \sum_{n=0}^{\infty} \frac{f^{(n)}(2)}{n!}(x-2)^n \\ &= \sum_{n=0}^{\infty} \frac{(-1)^n \cdot n! \cdot \frac{1}{2^{n+1}}}{n!}(x-2)^n \\ &= \sum_{n=0}^{\infty} \frac{(-1)^n}{2^{n+1}}(x-2)^n \end{aligned}$$

The series converges if

$$\begin{aligned} \lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| &= \lim_{n\to\infty}\left|\frac{(-1)^{n+1}(x-2)^{n+1}}{2^{n+2}} \cdot \frac{2^{n+1}}{(-1)^n(x-2)^n}\right| \\ &= \frac{|x-2|}{2} \\ &< 1 \\ &\Leftrightarrow |x-2| < 2 \end{aligned}$$

so the radius of convergence is 2.

Example.

(a) Find the Maclaurin series of e^x and its radius of convergence.

(b) Hence, find the Maclaurin series of e^{-5x^2} and its radius of convergence.

Solution.

(a) Let $f(x) = e^x$. We know that $f^{(n)}(x) = e^x$ for all $n \geq 0$, so

$$f^{(n)}(0) = e^0 = 1$$

Then the Maclaurin series is simply

$$
\begin{aligned}
e^x &= \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!}\, x^n \\
&= \sum_{n=0}^{\infty} \frac{x^n}{n!}
\end{aligned}
$$

The series converges if

$$
\begin{aligned}
\lim_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right| &= \lim_{n\to\infty} \left| \frac{x^{n+1}}{(n+1)!} \cdot \frac{n!}{x^n} \right| \\
&= |x| \lim_{n\to\infty} \frac{1}{n+1} \\
&= 0 \\
&< 1
\end{aligned}
$$

so the series converges for all $x \in \mathbb{R}$, and the radius of convergence is ∞.

(b) If there wasn't *"hence"* before this part, it would be perfectly valid to find the Maclaurin series from scratch. However, we must use the result from (a). This isn't too hard. We substitute $-5x^2$ into the series above:

$$
\begin{aligned}
e^{-5x^2} &= \sum_{n=0}^{\infty} \frac{(-5x^2)^n}{n!} \\
&= \sum_{n=0}^{\infty} (-1)^n\, 5^n\, \frac{x^{2n}}{n!}
\end{aligned}
$$

The series converges for all $x \in \mathbb{R}$ (since $-5x^2 \in \mathbb{R}$ for all x), so the radius of convergence is ∞.

Example. Find the Maclaurin series for

$$
f(x) = x^4 e^{-3x^2}
$$

and its radius of convergence.

Solution. Don't even try differentiating the expression.

We know the Maclaurin series for e^x:

$$
e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}
$$

so we multiply x^4 outside the summation and substitute in $-3x^2$ to get

$$\begin{aligned} f(x) &= x^4 \sum_{n=0}^{\infty} \frac{(-3x^2)^n}{n!} \\ &= x^4 \sum_{n=0}^{\infty} (-1)^n \, 3^n \, \frac{x^{2n}}{n!} \\ &= \sum_{n=0}^{\infty} (-1)^n \, 3^n \, \frac{x^{2n+4}}{n!} \end{aligned}$$

The series converges if (we drop the $(-1)^n$ because its absolute value is 1)

$$\begin{aligned} \lim_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right| &= \lim_{n\to\infty} \left| \frac{3^{n+1} x^{2n+6}}{(n+1)!} \cdot \frac{n!}{3^n x^{2n+4}} \right| \\ &= x^2 \lim_{n\to\infty} \frac{3}{n+1} \\ &= 0 \\ &< 1 \end{aligned}$$

so the series converges for all $x \in \mathbb{R}$, so the radius of convergence is ∞.

Example. Find the Taylor series for

$$f(x) = \ln x$$

centred around $x = 2$, and its radius of convergence.

Solution.

$$\begin{aligned} f(x) = \ln x &\Rightarrow f(2) = \ln 2 \\ f'(x) = x^{-1} &\Rightarrow f'(2) = \frac{1}{2} \\ f''(x) = -x^{-2} &\Rightarrow f''(2) = -\frac{1}{4} \\ f'''(x) = 2x^{-3} &\Rightarrow f'''(2) = 2 \cdot \frac{1}{8} \end{aligned}$$

In general,

$$\begin{aligned} f^{(n)}(x) &= (-1)^{n-1} \cdot (n-1)! \, x^{-n} \quad \text{for all } n \geq 1 \\ \Rightarrow f^{(n)}(2) &= (-1)^{n-1} \cdot (n-1)! \cdot \frac{1}{2^n} \end{aligned}$$

Note that the general form of the derivative does not hold for $n = 0$. So we extract that out of the series to get

$$\begin{aligned}\ln x &= \ln 2 + \sum_{n=1}^{\infty} \frac{(-1)^{n-1}(n-1)!}{2^n\, n!}(x-2)^n \\ &= \ln 2 + \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n2^n}(x-2)^n\end{aligned}$$

The series converges if

$$\begin{aligned}\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| &= \lim_{n\to\infty}\left|\frac{(x-2)^{n+1}}{(n+1)2^{n+1}}\cdot\frac{n2^n}{(x-2)^n}\right| \\ &= \frac{|x-2|}{2}\lim_{n\to\infty}\frac{n}{n+1} \\ &= \frac{|x-2|}{2} \\ &< 1 \\ &\Leftrightarrow |x-2| < 2\end{aligned}$$

so the radius of convergence is 2.

Example. Find the Maclaurin series for

$$\int \frac{\sin x}{x}\,\mathrm{d}x$$

Solution. The indefinite integral has no closed form. However, we know the Maclaurin series for $\sin x$:

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

and we know that the integral of a power series is equal to the term-by-term integrated series. So we first find the Maclaurin series for

$$\frac{\sin x}{x}$$

and then integrate the series term by term.

$$\begin{aligned}\frac{\sin x}{x} &= \frac{1}{x}\sum_{n=0}^{\infty} \frac{(-1)^{2n+1}}{(2n+1)!}\, x^{2n+1} \\ &= \sum_{n=0}^{\infty} \frac{(-1)^{2n+1}}{(2n+1)!}\, x^{2n}\end{aligned}$$

Then

$$\begin{aligned}\int \frac{\sin x}{x}\,\mathrm{d}x &= \int \sum_{n=0}^{\infty} \frac{(-1)^{2n+1}}{(2n+1)!}\, x^{2n}\,\mathrm{d}x \\ &= C + \sum_{n=0}^{\infty} \int \frac{(-1)^{2n+1}}{(2n+1)!}\, x^{2n}\,\mathrm{d}x \\ &= C + \sum_{n=0}^{\infty} \frac{(-1)^{2n+1}}{(2n+1)(2n+1)!}\, x^{2n+1}\end{aligned}$$

Example. Find the Taylor series of

$$x^2 + 8x + 1$$

centred around $x = 10$, and its radius of convergence.

Solution.

$$\begin{aligned} f(x) = x^2 + 8x + 1 &\Rightarrow f(10) = 181 \\ f'(x) = 2x + 8 &\Rightarrow f'(10) = 28 \\ f''(x) = 2 &\Rightarrow f''(10) = 2 \\ f^{(n)}(x) = 0 & \quad \text{for all } n \geq 3 \end{aligned}$$

So we have

$$\begin{aligned} f(x) &= \sum_{n=0}^{\infty} \frac{f^{(n)}(10)}{n!} (x-10)^n \\ &= 181 + 28(x-10) + \frac{2}{2!}(x-10)^2 \\ &= 181 + 28(x-10) + (x-10)^2 \end{aligned}$$

Since the series is finite, it converges for all $x \in \mathbb{R}$, so the radius of convergence is ∞.

Chapter 7

Numerical methods

7.1 Motivation

There are numerous reasons why we may want to solve equations. For instance, we may want to find the roots of a function f, i.e. the values of x that satisfy $f(x) = 0$. In the case of quadratic equations in the general form $ax^2 + bx + c = 0$, the roots are given by the quadratic formula

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

However, it is not always this easy to find the roots of any arbitrary function, or solve any arbitrary equation. In fact, we may not always be able to find analytic solutions to equations. Our next best bet would be approximating the values of the solutions. The tools that allow us to approximate *numerical* values of solutions are called *numerical methods*.

7.2 Newton-Raphson method

The Newton-Raphson method allows us to approximate the roots of a given function.

7.2.1 How it works

Consider a differentiable (and hence continuous) function f. We pick an initial guess x_0. Then, we calculate the x-intercept of the tangent line to the curve $y = f(x)$ at x_0 to obtain x_1. We keep iterating this process until we are satisfied. This could mean either:

- We have performed sufficiently many iterations, or
- The result is close enough to the root, i.e. $f(x)$ is close enough to 0 (since f is continuous).

Definition 7.1 (Newton-Raphson)**.** Given an approximation x_n, the next approximation x_{n+1} is given by

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

7.2.2 Derivation

Given some approximation of the root x_n, we define the next approximation x_{n+1} to be the x-intercept of the tangent line passing through $f(x_n)$. Recall that the equation of this tangent line is

$$y = f(x_n) + f'(x_n)(x - x_n) \tag{7.1}$$

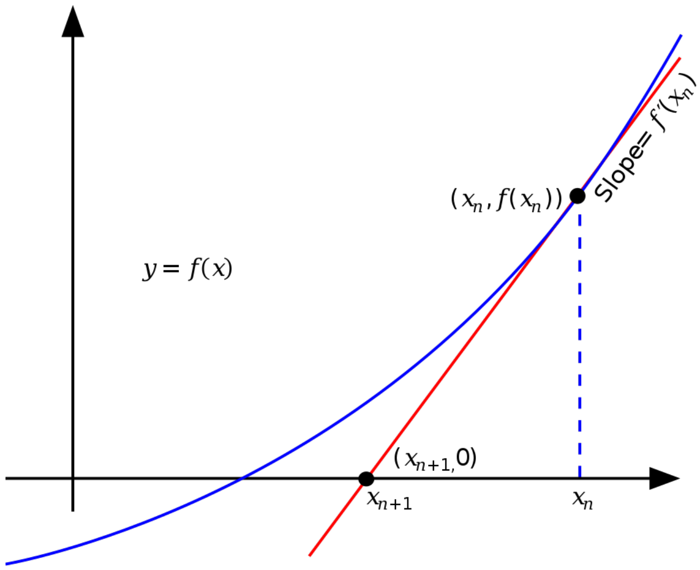


Figure 7.1: Newton-Raphson method

Intuition. The slope of any line is defined to be *rise over run*, so the difference in y-coordinate (*rise*) is equal to the slope times the *run* $(x - x_n)$. This difference is relative to the y-coordinate we started with, which is $f(x_n)$.

Alternatively, we can derive this from the definition of slope:

$$f'(x_n) = \frac{y - f(x_n)}{x - x_n}$$

Substituting $(x, y) = (x_{n+1}, 0)$ into Equation (7.1), we get

$$0 = f(x_n) + f'(x_n)(x_{n+1} - x_n) \tag{7.2}$$

and a little bit of rearranging gives us

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

Since we have $f'(x_n)$ in the denominator, we must have that $f'(x_n) \neq 0$.

7.2.3 Order of convergence

We might want to know how fast the Newton-Raphson method can give us a *good enough* approximation to a root of the function. An approximation x is *good enough* if it is within some predefined value ε of the root r, so $|x - r| \leq \varepsilon$. We define $\varepsilon_n = |x_n - r|$. (In other literature, there are no absolute value bars.)

Let f be a function with a continuous second derivative. Then by Taylor's theorem, we can express $f(x)$ as a 1st order Taylor polynomial centred at x_n:

$$f(x) = f(x_n) + f'(x_n)(x - x_n) + \frac{1}{2}f''(c)(x - x_n)^2$$

for some c between x and x_n. We set $x = r$ to get

$$f(r) = f(x_n) + f'(x_n)(r - x_n) + \frac{1}{2}f''(c)(r - x_n)^2$$

for some c between x_n and r. Since r is a root, by definition, $f(r) = 0$. So

$$0 = f(x_n) + f'(x_n)(r - x_n) + \frac{1}{2}f''(c)(r - x_n)^2 \tag{7.3}$$

Subtracting Equation (7.2) from Equation (7.3) gives us

$$0 = f'(x_n)(r - x_{n+1}) + \frac{1}{2}f''(c)(r - x_n)^2$$

Rearrange the equation a little bit to get

$$x_{n+1} - r = \frac{f''(c)}{2f'(x_n)}(x_n - r)^2$$

Taking the absolute value of both sides, we get

$$\begin{aligned} |x_{n+1} - r| &= \frac{|f''(c)|}{2|f'(x_n)|}|x_n - r|^2 \\ \varepsilon_{n+1} &= \frac{|f''(c)|}{2|f'(x_n)|}\varepsilon_n^2 \end{aligned}$$

From this result, we say that the Newton-Raphson method has **quadratic convergence**. However, certain conditions have to be satisfied:

Theorem 7.2. The Newton-Raphson method has at least quadratic convergence if the following conditions are satisfied:

1. $f'(x) \neq 0$ for every $x \in I := [r - \varepsilon_0, r + \varepsilon_0]$
2. f'' is continuous on I—this is the assumption for Taylor's theorem.
3. $M\varepsilon_0 < 1$, where M is defined to be

$$M = \frac{1}{2}\left(\sup_{x \in I}|f''(x)|\right)\left(\sup_{x \in I}\frac{1}{|f'(x)|}\right)$$

Remark. A few notes on the third condition:

- The supremum is attained because f is continuous (since f' exists), so it is the global maximum of f restricted to I.
- We must take the suprema separately, as opposed to something like

$$\sup_{x \in I}\frac{|f''(x)|}{2|f'(x)|}$$

  since the x in the numerator is some unknown value c between x_n and r, as determined by Taylor's theorem, while the x in the denominator is our guess $x_n \neq c$.
- We specify $M\varepsilon_0 < 1$ because we want ε_n for all $n \geq 1$ to be no larger than ε_0, so that all of our subsequent approximations will be within I (hence I is defined as such):

$$\varepsilon_1 = \frac{|f''(c)|}{2|f'(x)|}\varepsilon_0^2 < 1 \times \varepsilon_0 = \varepsilon_0$$

It follows by induction that

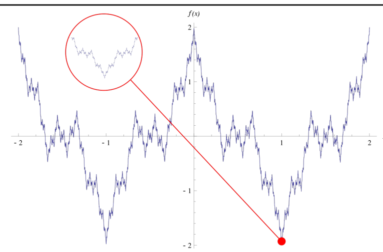
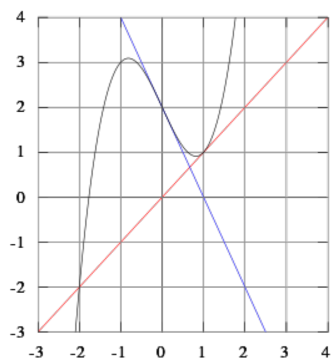
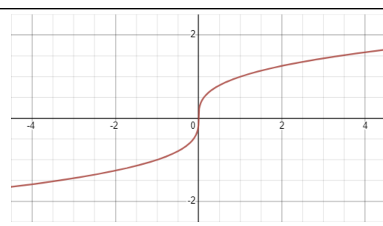
$$\varepsilon_n \leq M^{2n-1}\varepsilon_0^{2n} = \frac{1}{M}(M\varepsilon_0)^{2n}$$

By the squeeze theorem, ε_n converges to 0 (i.e. the approximations get arbitrarily close to the root) if $M\varepsilon_0 < 1$.

7.2.4 Limitations

Some of the limitations of the Newton-Raphson method are outlined in Table 7.1.

Table 7.1: Limitations of the Newton-Raphson method

Limitation	**Example**	**Workaround**
The function is not differentiable.	 The Weierstrass function is continuous but not differentiable.	Use the secant method instead.
Tangent lines keep oscillating.	 When applied to the function $y = x^3 - 2x + 2$, Newton's method oscillates between 0 and 1.	Pick a different starting point, perhaps using the bisection method.
Diverges.	 Consider the function $f(x) = \sqrt[3]{x}$. $$x_{n+1} = x_n - \frac{x_n^{1/3}}{\frac{1}{3}x_n^{-2/3}} = -2x_n$$ which diverges for all $x_n \neq 0$. But then $f'(0)$ is undefined.	RIP.

7.3 Quasi-Newton methods: secant method

In the Newton-Raphson method, we computed the first derivative at every approximation to find the next approximation. However, computing derivatives may be expensive. Instead, we may use an approximation for the derivative. Numerical methods that approximate the derivative instead of computing its precise value are called *quasi-Newton methods*. One such method is the secant method.

7.3.1 How it works

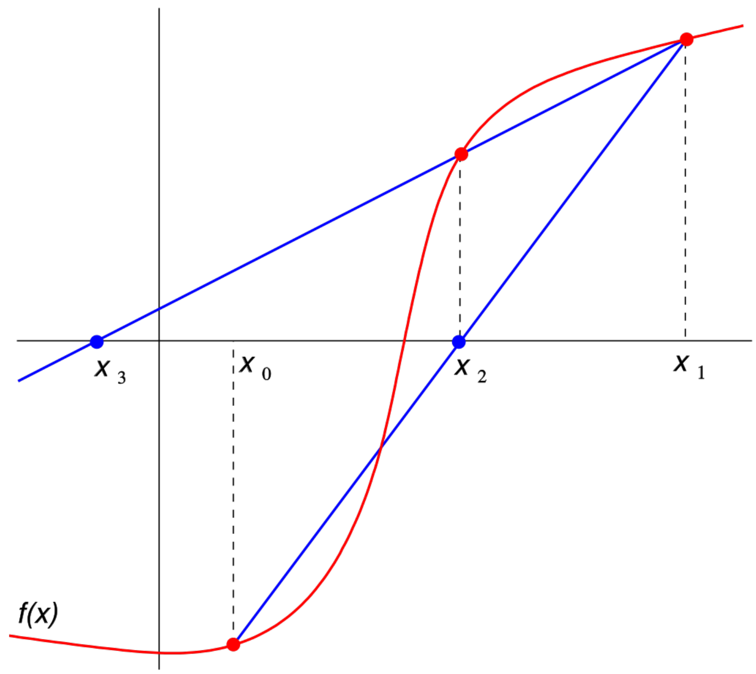


Figure 7.2: Secant method

Suppose we want to find the roots of a function f. We pick two initial guesses x_0 and x_1, and draw the secant passing through the two points $(x_0, f(x_0))$ and $(x_1, f(x_1))$. Our next approximation x_2 is the x-intercept of this secant line. We keep drawing the secant lines using our two most recent guesses until satisfied.

Definition 7.3 (Secant method)**.** Given the two most recent approximations x_n and x_{n-1}, the next approximation x_{n+1} is given by

$$x_{n+1} = x_n - f(x_n)\frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})}$$

Intuition. We approximate the tangent line by the secant line, so

$$f'(x_n) \approx \frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}}$$

Substituting this into Definition 7.1 and replacing the approximation with equality gives us the required form.

7.3.2 Derivation

The equation of the secant line between $(x_n, f(x_n))$ and $(x_{n-1}, f(x_{n-1}))$ is given by

$$y = f(x_n) + (x - x_n)\frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}} \tag{7.4}$$

Substituting $(x, y) = (x_{n+1}, 0)$ into Equation (7.4) gives us

$$0 = f(x_n) + (x_{n+1} - x_n)\frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}}$$

and a little bit of rearranging turns this into

$$x_{n+1} = x_n - f(x_n)\frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})}$$

7.3.3 Order of convergence

The secant method bears the advantage of lower computational costs, as it does not require computing derivatives. With this advantage comes a disadvantage: the secant method converges to a root slower than the Newton-Raphson method. In fact, the order of convergence is the golden ratio

$$\frac{1 + \sqrt{5}}{2}$$

but we will not go into the derivation here.

7.4 Initial value problem

An initial value problem is a differential equation

$$\frac{\mathrm{d}y}{\mathrm{d}t} = f(t, y)$$

given the initial conditions (t_0, y_0). Solutions to initial value problems are, of course, functions, but sometimes we may only want to know the value of a solution function y evaluated at a certain point t_n. The following methods allow us to generate a sequence of approximations at $t_1, t_2, \ldots$.

7.4.1 Euler's method

In Euler's method, we approximate the solution curve by tangent lines. Given t_0 and y_0, suppose we want to approximate the solution at t_1. The tangent line passing through (t_0, y_0) has the equation

$$y = y_0 + f(t_0, y_0)(t - t_0)$$

since $f(t_0, y_0)$ is given to be the slope of the tangent line at t_0. Setting $(t, y) = (t_1, y_1)$ gives us

$$y_1 = y_0 + f(t_0, y_0)(t_1 - t_0)$$

In general,

> **Definition 7.4** (Euler's method)**.** Given t_n and the approximation of the solution at this point y_n, the next approximation at t_{n+1} is given by
>
> $$y_{n+1} = y_n + f(t_n, y_n)(t_{n+1} - t_n)$$

Sometimes, we may want to use a uniform step size, i.e. for every n, $t_{n+1} - t_n = h$ for some h. Then we can rewrite Euler's method as

$$y_{n+1} = y_n + hf(t_n, y_n)$$

Notice we can also neatly express t_n in terms of h and t_0:

$$t_n = t_0 + hn$$

7.4.2 Modified Euler's method (Heun's method)

The improved Euler's method, also known as Heun's method, is the simplest example of a *predictor-corrector method*. Heun's method *corrects* some of the error in Euler's method by taking the average of the slopes at (t_n, y_n) and (t_{n+1}, y^*_{n+1}), where y^*_{n+1} denotes the approximation at x_{n+1} using Euler's method:

$$y^*_{n+1} = y_n + hf(t_n, y_n)$$

We state this mathematically.

Definition 7.5 (Heun's method)**.** Given t_n and the approximation of the solution at this point y_n, the next approximation at t_{n+1} is given by

$$y_{n+1} = y_n + \frac{1}{2}(f(x_n, y_n) + f(x_{n+1}, y^*_{n+1}))h$$

7.4.3 Runge-Kutta method

The Runge-Kutta method is another example of a predictor-corrector method. It takes the weighted average of four slopes:

Definition 7.6 (Runge-Kutta method)**.** Given t_n and the approximation of the solution at this point y_n, the next approximation at t_{n+1} is given by

$$y_{n+1} = y_n + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)h$$

where

- $k_1 = f(t_n, y_n)$—the slope at t_n
- $k_2 = f(t_n + \frac{h}{2}, y_n + \frac{h}{2}k_1)$—the slope halfway between t_n and t_{n+1}
- $k_3 = f(t_n + \frac{h}{2}, y_n + \frac{h}{2}k_2)$—the slope halfway between t_n and t_{n+1}, this time using k_2
- $k_4 = f(t_n + h, y_n + hk_3)$—the slope at t_{n+1}

Chapter 8

Metric spaces

This chapter draws heavily from the first two lectures of Introduction to Metric Spaces taught by Paige Dote at MIT in 2022.

8.1 Definition

In the previous chapters, we used absolute value bars extensively to denote the distance between two real numbers. There are a few properties that make them intuitive and natural when quantifying distance. For all $x, y \in \mathbb{R}$,

- $|x-y| \geq 0$. Distances should be non-negative! Also, $|x-y| = 0 \Leftrightarrow x = y$.
- $|x-y| = |x-y|$. The distance between x and y should be equal to the distance between y and x.
- Triangle inequality: $|x-z| \leq |x-y| + |y-z|$. This comes naturally from the definition of the absolute value bars, and we seemingly take this for granted when proving things in $\mathbb{R}$.

Metric spaces generalise this concept of distance to arbitrary sets.

Definition 8.1 (Metric space). A metric space is a set X with a metric $d : X \times X \to \mathbb{R}$ such that for all $x, y, z \in X$, the metric d satisfies the following properties:

- Positive definite: 1) $d(x,y) \geq 0$ and 2) $d(x,y) = 0 \Leftrightarrow x = y$
- Symmetric: $d(x,y) = d(y,x)$
- Triangle inequality: $d(x,z) \leq d(x,y) + d(y,z)$

Positive definiteness restricts the codomain of d to $[0,\infty)$, so we can write $d : X \times X \to [0,\infty)$ instead.

8.2 Common metrics in $\mathbb{R}^n$

We will now look at some common metrics.

Definition 8.2 (Supremum metric). The supremum metric $d_\infty : \mathbb{R}^n \times \mathbb{R}^n \to [0,\infty)$ is defined by

$$d_\infty(x,y) = \max_{1 \leq i \leq n} |x_i - y_i|$$

We check that it is a metric.

- Positive definite:

- Clearly, for all i, $|x_i - y_i| \geq 0$, so $d_\infty(x, y) \geq 0$.
- $d_\infty(x, y) = 0 \Leftrightarrow |x_i - y_i| = 0$ for all $i \Leftrightarrow x = y$.

- Symmetry: $d_\infty(x, y) = \max|x_i - y_i| = \max|y_i - x_i| = d_\infty(y, x)$. This follows immediately from the symmetry of absolute value bars.
- Triangle inequality:

$$\begin{aligned} d_\infty(x, z) &= \max|x_i - z_i| && \text{(attained at some } i_0\text{)} \\ &\leq \max(|x_i - y_i| + |y_i - z_i|) && \text{holds for } i_0\text{, so holds for all } i \\ &\leq \max|x_i - y_i| + \max|y_i - z_i| && \text{may not be attained at same } i \\ &= d_\infty(x, y) + d_\infty(y, z). \end{aligned}$$

This follows from the triangle inequality for absolute value bars.

Definition 8.3 (ℓ^1 metric)**.** The ℓ^1 metric $d_1 : \mathbb{R}^n \times \mathbb{R}^n \to [0, \infty)$ is defined by

$$d_1(x, y) = \sum_{i=1}^{n} |x_i - y_i|$$

We check that it is a metric.

- Positive definite:
  - Clearly, for all i, $|x_i - y_i| \geq 0$, so their sum also ≥ 0, and $d_1(x, y) \geq 0$.
  - $d_1(x, y) = 0 \Leftrightarrow \sum|x_i - y_i| = 0 \Leftrightarrow x_i = y_i$ for all $i \Leftrightarrow x = y$.
- Symmetry:

$$d_1(x, y) = \sum|x_i - y_i| = \sum|y_i - x_i| = d_1(y, x).$$

- Triangle inequality:

$$\begin{aligned} d_1(x, z) &= \sum|x_i - z_i| \\ &\leq \sum|x_i - y_i| + |y_i - z_i| \\ &= \sum|x_i - y_i| + \sum|y_i - z_i| && \text{by linearity of summation} \\ &= d_1(x, y) + d_1(y, z) \end{aligned}$$

Now, we will consider the set of continuous functions $C^0([a, b])$, where every element is a function of type $[a, b] \to \mathbb{R}$ which is continuous on $[a, b]$.

Example. Show that $d : C^0([0, 1]) \times C^0([0, 1]) \to [0, \infty)$ defined by

$$d(f, g) = \sup_{x \in [0,1]} |f(x) - g(x)|$$

is a metric.

Solution. Positive definiteness and symmetry are straightforward. For triangle inequality, take arbitrary $f, h \in C^0([0, 1])$. Since f and h are both continuous on a closed interval $[0, 1]$, the function $|f(x) - h(x)|$ is also continuous on $[0, 1]$. By the Extreme Value Theorem, there exists x_0 such that

$$|f(x_0) - h(x_0)| = \sup_{x \in [0,1]} |f(x) - h(x)|$$

Then we have

$$
\begin{aligned}
d(f,h) &= \sup_{x\in[0,1]} |f(x)-h(x)| \\
&= |f(x_0)-h(x_0)| \\
&\le |f(x_0)-g(x_0)| + |g(x_0)-h(x_0)| && \text{by } \triangle \\
&\le \sup_{x\in[0,1]} |f(x)-g(x)| + \sup_{x\in[0,1]} |g(x)-h(x)| && \text{by definition of supremum} \\
&= d(f,g)+d(g,h)
\end{aligned}
$$

Next, we consider the set of continuously differentiable functions $C^1([a,b])$.

Example. Show that $d : C^1([0,1]) \times C^1([0,1]) \to [0,\infty)$ defined by

$$d(f,g) = \sup_{x\in[0,1]} |f(x)-g(x)| + \sup_{x\in[0,1]} |f'(x)-g'(x)|$$

is a metric.

Solution. For triangle inequality, from the above example, we know that

$$\sup_{x\in[0,1]} |f(x)-h(x)| \le \sup_{x\in[0,1]} |f(x)-g(x)| + \sup_{x\in[0,1]} |g(x)-h(x)|$$

since $f,g,h \in C^0([0,1])$ as well (they are all continuous on $[0,1]$). Also, their derivatives $f',g',h' \in C^0([0,1])$ by definition, so

$$\sup_{x\in[0,1]} |f'(x)-h'(x)| \le \sup_{x\in[0,1]} |f'(x)-g'(x)| + \sup_{x\in[0,1]} |g'(x)-h'(x)|$$

We add the two inequalities and we are done.

Example (PSET 1 Q2)**.** Is $d : C^1([0,1]) \times C^1([0,1]) \to [0,\infty)$ defined by

$$d(f,g) = \sup_{x\in[0,1]} |f'(x)-g'(x)|$$

a metric on $C^1([0,1])$? If so, prove it. If not, show what properties of a metric d satisfies, and explain which properties of a metric d fails.

Solution.

- Positive definite:
  - Clearly, $d(f,g) \ge 0$ for all $f,g \in C^1([0,1])$.
  - If $f=g$, then $f'(x)-g'(x)=0$ for all $x\in[0,1]$, so $d(f,g)=0$. However, if $d(f,g)=0$, we get $f'(x)=g'(x)$ for all $x\in[0,1]$, but this does not imply $f=g$. For instance, take $f(x)=x$ and $g(x)=x+1$.

  Therefore, d is not positive definite.
- Symmetry: follows from absolute value bars.
- Triangle inequality: similar to a previous example with $C^0([0,1])$.

Example. Let (X, d) be a metric space. Show that $\tilde{d} : X \times X \to [0, \infty)$ defined by

$$\tilde{d}(x, y) = \frac{d(x, y)}{1 + d(x, y)}$$

is a metric, and show that X is bounded in the metric $\tilde{d}$.

Solution.

- Positive definite:
  - Since $d(x, y) \geq 0$ and $1 + d(x, y) \geq 0$, $\tilde{d}(x, y) \geq 0$.
  - $\tilde{d}(x, y) = 0 \Leftrightarrow d(x, y) = 0 \Leftrightarrow x = y$.
- Symmetry:
$$\tilde{d}(x, y) = \frac{d(x, y)}{1 + d(x, y)} = \frac{d(y, x)}{1 + d(y, x)} = \tilde{d}(y, x)$$
- Triangle inequality:
$$\begin{aligned}
\tilde{d}(x, z) &= \frac{d(x, z)}{1 + d(x, z)} \\
&= 1 - \frac{1}{1 + d(x, z)} \\
&\leq 1 - \frac{1}{1 + d(x, y) + d(y, z)} \\
&= \frac{d(x, y) + d(y, z)}{1 + d(x, y) + d(y, z)} \\
&= \frac{d(x, y)}{1 + d(x, y) + d(y, z)} + \frac{d(y, z)}{1 + d(x, y) + d(y, z)} \\
&\leq \frac{d(x, y)}{1 + d(x, y)} + \frac{d(y, z)}{1 + d(y, z)} \\
&= \tilde{d}(x, y) + \tilde{d}(y, z)
\end{aligned}$$

Fix any $x \in X$. For every $y \in X$, $\tilde{d}(x, y) < 1$, so X is bounded in $\tilde{d}$.

Example. Show that the map $\frac{\mathrm{d}}{\mathrm{d}x} : C^1([a, b]) \to C^0([a, b])$ is continuous with respect to the uniform distance.

Solution. Take arbitrary $f, g \in C^1([a, b])$. We want to show that

$$\forall \varepsilon > 0,\ \exists \delta > 0 : 0 < d_{C^1}(f, g) < \delta \Rightarrow d_{C^0}(\frac{\mathrm{d}}{\mathrm{d}x} f, \frac{\mathrm{d}}{\mathrm{d}x} g) < \varepsilon$$

By definition,

$$d_{C^1}(f, g) = \sup_{x \in [a,b]} |f(x) - g(x)| + \sup_{x \in [a,b]} |f'(x) - g'(x)|$$

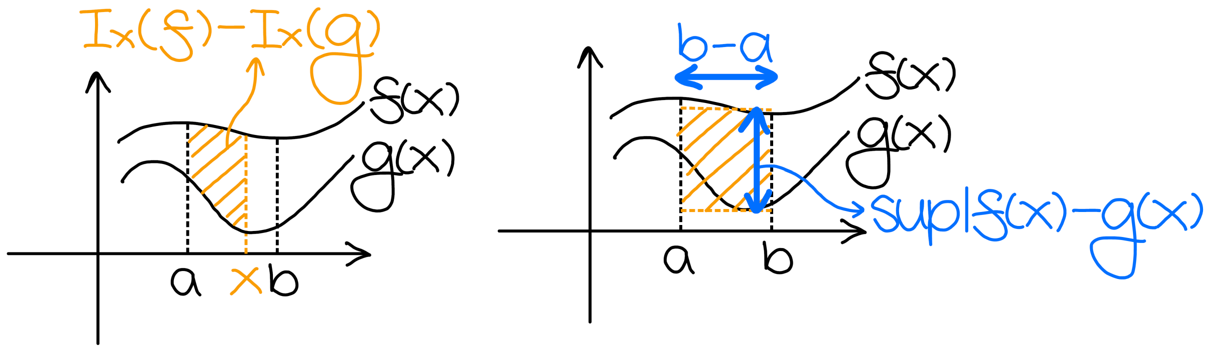


Figure 8.1: Integration is continuous

and

$$d_{C^0}(f', g') = \sup_{x \in [a,b]} |f'(x) - g'(x)|$$

Observe that $d_{C^0}(f', g') \leq d_{C^1}(f, g)$, so set $\delta = \varepsilon$.

Example (PSET 1 Q5). Show that the map $I_t : C^0([a, b]) \to C^1([a, b])$ is continuous, where

$$I_x(f) = \int_a^x f(t)\, \mathrm{d}t$$

Assume we are using the uniform distances for $C^0([a, b])$ and $C^1([a, b])$ as metrics.

Solution. Take arbitrary $f, g \in C^0([a, b])$. We want to show that

$$\forall \varepsilon > 0,\ \exists \delta > 0 : 0 < d_{C^0}(f, g) < \delta \Rightarrow d_{C^1}(I_x(f), I_x(g)) < \varepsilon$$

As usual, we start from the consequent:

$$\begin{aligned}
d_{C^1}(I_x(f), I_x(g)) &= \sup_{x \in [a,b]} |I_x(f) - I_x(g)| + \sup_{x \in [a,b]} |f(x) - g(x)| \\
&= \sup_{x \in [a,b]} \left| \int_a^x f(t) \mathrm{d}t - \int_a^x g(t) \mathrm{d}t \right| + \sup_{x \in [a,b]} |f(x) - g(x)| \\
&= \sup_{x \in [a,b]} \left| \int_a^x f(t) - g(t) \mathrm{d}t \right| + \sup_{x \in [a,b]} |f(x) - g(x)| && \text{by linearity of integration} \\
&\leq \sup_{x \in [a,b]} \int_a^x |f(t) - g(t)| \mathrm{d}t + \sup_{x \in [a,b]} |f(x) - g(x)| && \text{by analogue version of } \triangle \\
&\leq (b - a) \sup_{x \in [a,b]} |f(x) - g(x)| + \sup_{x \in [a,b]} |f(x) - g(x)| && \text{refer to Figure 8.1} \\
&= (1 + b - a) \sup_{x \in [a,b]} |f(x) - g(x)| \\
&< \varepsilon
\end{aligned}$$

So we pick $\delta = \dfrac{\varepsilon}{1 + b - a}$.

8.3 Sequences in metric spaces

As metric spaces, in some sense, generalise $\mathbb{R}^n$, we can extend some of our previous definitions to metric spaces.

Definition 8.4 (Convergence). A sequence $\{x_n\}$ in a metric space (X, d) converges to x if

$$\forall \varepsilon > 0,\ \exists N \in \mathbb{N} : \forall n \in \mathbb{N},\ n > N \Rightarrow d(x_n, x) < \varepsilon$$

We replaced the absolute value bars with d.

Definition 8.5 (Cauchy sequence). A sequence $\{x_n\}$ in a metric space (X, d) converges to x if

$$\forall \varepsilon > 0,\ \exists N \in \mathbb{N} : \forall n, m \in \mathbb{N},\ n, m > N \Rightarrow d(x_n, x_m) < \varepsilon$$

Definition 8.6 (Continuous functions). Let (X, d_X) and (Y, d_Y) be metric spaces. A function $f : X \to Y$ is continuous at $x \in X$ if

$$\forall \varepsilon > 0,\ \exists \delta > 0 : \forall y \in X,\ 0 < d_X(x, y) < \delta \Rightarrow d_Y(f(x), f(y)) < \varepsilon$$

This one is a bit different.

Definition 8.7 (Bounded). A sequence $\{x_n\}$ in a metric space (X, d) is bounded if there exists $p \in X$ and $M \in \mathbb{R}$ such that for all $n \in \mathbb{N}$,

$$d(x_n, p) \leq M$$

A set $A \subseteq X$ is bounded if there exists $p \in X$ and $M \in \mathbb{R}$ such that for all $x \in X$,

$$d(x, p) \leq M$$

With these definitions, we are equipped with the basic tools to deal with general metric spaces in meaningful ways, as we did with the set of real numbers $\mathbb{R}$ in previous chapters. Some of the examples will look a great deal similar to the analogous ones in the reals.

Proposition 8.8. Let (X, d) be a metric space and let x_n be a convergent sequence in X such that $x_n \to x$. Show that this limit is unique.

Proof. Suppose $x_n \to y \in X$. Since $x_n \to x$, we have

$$\forall \varepsilon > 0,\ \exists N_1 \in \mathbb{N} : n > N_1 \Rightarrow d(x_n, x) < \frac{\varepsilon}{2}$$

and since $x_n \to y$, we have

$$\forall \varepsilon > 0,\ \exists N_2 \in \mathbb{N} : n > N_2 \Rightarrow d(x_n, y) < \frac{\varepsilon}{2}$$

Fix some $\varepsilon > 0$. For all $n > \max\{N_1, N_2\}$,

$$\begin{aligned} d(x,y) &\leq d(x,x_n) + d(x_n,y) && \text{by } \triangle \\ &= d(x_n,x) + d(x_n,y) && \text{by symmetry} \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\ &= \varepsilon \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, it follows that $d(x,y) = 0$, and by positive definiteness, $x = y$. □

Proposition 8.9. Let $x_n \to x$. Then for all $y \in X$, $d(x_n, y) \to d(x, y)$.

Proof. We want to show that

$$\forall \varepsilon > 0,\ \exists N \in \mathbb{N} : n > N \Rightarrow |d(x_n,y) - d(x,y)| < \varepsilon$$

Note that we are using the absolute value bars instead of d since $d(x,y) \in \mathbb{R}$ but not X.

Since we are given that $x_n \to x$, we have

$$\forall \varepsilon > 0,\ \exists N \in \mathbb{N} : n > N \Rightarrow d(x_n,x) < \varepsilon$$

Fix some $\varepsilon > 0$. Then for all $n > N$,

$$\begin{aligned} |d(x_n,y) - d(x,y)| &\leq d(x_n,x) && \text{by reverse } \triangle \\ &< \varepsilon \end{aligned}$$

□

Proposition 8.10. Let $x_n \to x$ and $y_n \to y$. Then $d(x_n, y_n) \to d(x, y)$.

Proof. We want to show that

$$\forall \varepsilon > 0,\ \exists N \in \mathbb{N} : n > N \Rightarrow |d(x_n,y_n) - d(x,y)| < \varepsilon$$

Given $x_n \to x$, we have

$$\forall \varepsilon > 0,\ \exists N_1 \in \mathbb{N} : n > N_1 \Rightarrow d(x_n,x) < \frac{\varepsilon}{2}$$

and given $y_n \to y$, we have

$$\forall \varepsilon > 0,\ \exists N_2 \in \mathbb{N} : n > N_2 \Rightarrow d(y_n,y) < \frac{\varepsilon}{2}$$

Fix $\varepsilon > 0$. Then for all $n > \max\{N_1, N_2\}$,

$$\begin{aligned} |d(x_n,y_n) - d(x,y)| &\leq |d(x_n,y_n) - d(x_n,y)| + |d(x_n,y) - d(x,y)| && \text{by } \triangle \\ &\leq d(y_n,y) + d(x_n,x) && \text{by reverse } \triangle \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\ &= \varepsilon \end{aligned}$$

□

Proposition 8.11 (PSET 2, Q1)**.** Let x_n and y_n be Cauchy sequences in X. Then $d(x_n, y_n)$ converges.

Remark. We **cannot** assume that x_n and y_n converges—this is only true if X is *Cauchy complete*.

Proof. Since x_n is Cauchy,

$$\forall \varepsilon > 0,\ \exists N_1 \in \mathbb{N} : n, m > N_1 \Rightarrow d(x_n, x_m) < \frac{\varepsilon}{2}$$

and since y_n is Cauchy,

$$\forall \varepsilon > 0,\ \exists N_2 \in \mathbb{N} : n, m > N_2 \Rightarrow d(y_n, y_m) < \frac{\varepsilon}{2}$$

(At this point, it shouldn't be hard to envision that the proof will involve the triangle inequality, which involves adding, hence the $\varepsilon/2$.) We want to show that $d(x_n, y_n)$ converges, but we don't know what the limit could look like. It should probably be something like $d(x, y)$, but then x_n and y_n do not necessarily converge in X! However, remember that $d(x_n, y_n)$ is an element of $\mathbb{R}$ (but not X), and that $\mathbb{R}$ is Cauchy complete. We will try to show that $d(x_n, y_n)$ is Cauchy, then it follows that it is convergent.

Fix $\varepsilon > 0$. For all $n, m > \max\{N_1, N_2\}$,

$$\begin{aligned}
|d(x_n, y_n) - d(x_m, y_m)| &\leq |d(x_n, y_n) - d(x_n, y_m)| + |d(x_n, y_m) - d(x_m, y_m)| && \text{by } \triangle \\
&\leq d(y_n, y_m) + d(x_n, x_m) && \text{by reverse } \triangle \\
&\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\
&= \varepsilon
\end{aligned}$$

□

Proposition 8.12. Every convergent sequence in a metric space is bounded.

Proof. Given a sequence x_n that converges to x in some metric space (X, d), we have

$$\forall \varepsilon > 0,\ \exists N \in \mathbb{N} : n > N \Rightarrow d(x_n, x) < \varepsilon$$

Let $\varepsilon = 1$. Then $d(x_n, x) < 1$ with $n > N$. But then we know nothing about the terms before (and at) N, so we let

$$M = \max\{1, d(x_1, x), d(x_2, x), \ldots, d(x_N, x)\}$$

Note that M is finite since it is the maximum of finitely many terms. For $n \leq N$, $d(x_n, x) \leq M$ by construction. For $n > N$, $d(x_n, x) < 1 \leq M$. This covers all $n \in \mathbb{N}$. □

Proposition 8.13. Every convergent sequence in a metric space is Cauchy.

Proof. We want to show that

$$\forall \varepsilon > 0,\ \exists N \in \mathbb{N} : n, m > N \Rightarrow d(x_n, x_m) < \varepsilon$$

Given a sequence x_n that converges to x in some metric space (X, d), we have

$$\forall \varepsilon > 0,\ \exists N \in \mathbb{N} : n > N \Rightarrow d(x_n, x) < \varepsilon$$

As n is a dummy variable, we can replace it with m. It seems like we will be using the triangle inequality, which will involve adding two terms, so we replace ε with $\varepsilon/2$ instead:

$$\forall \varepsilon > 0,\ \exists N \in \mathbb{N} : n > N \Rightarrow d(x_n, x) < \frac{\varepsilon}{2}$$

Fix $\varepsilon > 0$. For $n, m > N$,

$$\begin{aligned} d(x_n, x_m) &\leq d(x_n, x) + d(x_m, x) && \text{by } \triangle \text{ and symmetry} \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\ &= \varepsilon \end{aligned}$$

□

When we considered sequences in $\mathbb{R}$, we also proved that all Cauchy sequences are convergent. However, this is not true for any metric space in general. A metric space in which all Cauchy sequences are convergent is called **Cauchy complete**.

Proposition 8.14. Every subsequence of a convergent sequence is convergent.

Proof. Given a sequence x_n converges to x, we have

$$\forall \varepsilon > 0,\ \exists N \in \mathbb{N} : n > N \Rightarrow d(x_n, x) < \varepsilon$$

We show that x_{n_k} is convergent by showing that it also converges to x:

$$\forall \varepsilon > 0,\ \exists N \in \mathbb{N} : k > N \Rightarrow d(x_{n_k}, x) < \varepsilon$$

Crucially, note that $k > N$ instead of $n > N$. This is because n_k is basically a function of k, and the n is just to make it clear that it is a subsequence of x_n. We can show that $n_k \geq k$, for instance, by induction. So for every $k > N$, $n_k > N$, and $d(x_{n_k}, x) < \varepsilon$.

(We prove $n_k \geq k$ by induction is as follows. For the base case, $k = 1$ and $n_k \geq 1$ by convention, so $n_k \geq k$. Now suppose that $n_r \geq r$ for some r. Noting that n_r is a strictly increasing sequence, we have $n_{r+1} > n_r \geq r$, so $n_{r+1} \geq r + 1$.) □

8.4 Open sets

We introduce a new definition:

Definition 8.15 (Open sets)**.** Let (X, d) be a metric space. A set $A \subseteq X$ is open if for every $x \in A$, there exists $\varepsilon > 0$ such that

$$B(x, \varepsilon) := \{y \in X \mid d(x, y) < \varepsilon\} \subseteq A$$

We say that $B(x, \varepsilon)$ is a ball of radius ε centred at x.

Theorem 8.16 (Topological properties of open sets)**.** Let X be a metric space, and let $\{A_i\}_{i \in \Lambda}$ be open sets in X.

1. $\varnothing$ and X are open sets in X.
2. $\bigcup_{i \in I} A_i$ is open in X. (The arbitrary union of open sets is open.)
3. $\bigcap_{i=1}^{n} A_i$ is open in X. (The finite intersection of open sets is open.)

Remark. The symbol Λ (uppercase lambda) represents an **index set**. For example, for $\{A_1, A_2, \ldots, A_n\}$, the index set Λ might be $\{1, 2, \ldots, n\}$, so we can rewrite the collection as $\{A_i \mid i \in \Lambda\}$. However, we don't know if the index set is finite, or even countable, so we use Λ instead to encapsulate all of these possibilities. On the other hand, $I \subseteq \Lambda$.

Proof. 1. $\varnothing$ is open as it is vacuously true that for every element in $\varnothing$, ... X is open because a ball centred at any arbitrary point in X with arbitrary radius must be a subset of X by construction: we are taking elements from X to form the ball! So for every $x \in X$, there exists (in fact, for every) $\varepsilon > 0$ such that $B(x, \varepsilon) \subseteq X$.

2. Take an arbitrary element $x \in \bigcup A_i$. Then $x \in A_k$ for some k. Since A_k is open, there exists $\varepsilon > 0$ such that $B(x, \varepsilon) \subseteq A_k$. But then $A_k \subseteq \bigcup A_i$, so $B(x, \varepsilon) \subseteq \bigcup A_i$.

3. Take an arbitrary element $x \in \bigcap_{i=1}^n A_i$. Then $x \in A_1$, $x \in A_2$, ..., $x \in A_n$. So there exists $\varepsilon_i > 0$ such that $B(x, \varepsilon_i) \subseteq A_i$ for every $1 \leq i \leq n$. $B(x, \varepsilon_i) \subseteq A_i$ does not guarantee $B(x, \varepsilon_i) \subseteq \bigcap_{i=1}^n A_i$, so we pick $\varepsilon = \min\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\}$. Then $B(x, \varepsilon) \subseteq A_1, A_2, \ldots, A_n$, and $B(x, \varepsilon) \subseteq \bigcap_{i=1}^n$. It is very reasonable to ask why we cannot take the *arbitrary* intersection of open sets, and must limit ourselves to a *finite* intersection. We can only guarantee that $\varepsilon > 0$ if it were a minimum of a finite number of terms; otherwise, ε_i may form a sequence that converges to 0. □

If you are doing Computing and have been taught by Paul Bilokon for this module, you might have heard of the term **open balls**. The balls are as we previously defined; we now show that they are open.

Proposition 8.17. Let (X, d) be a metric space. For any $x \in X$ and any $\varepsilon > 0$, $B(x, \varepsilon)$ is open in X.

Proof. We want to show that for every $y \in B(x, \varepsilon)$, there exists $\delta > 0$ such that

$$B(y, \delta) \subseteq B(x, \varepsilon)$$

Take $y \in B(x, \varepsilon)$ and let $\delta = \varepsilon - d(x, y)$. We verify that $\delta > 0$. Indeed: since $y \in B(x, \varepsilon)$, by definition, $d(x, y) < \varepsilon$. For every $z \in B(y, \delta)$,

$$\begin{aligned} d(x, z) &\leq d(x, y) + d(y, z) \\ &< d(x, y) + \delta \\ &= d(x, y) + (\varepsilon - d(x, y)) \\ &= \varepsilon \end{aligned}$$

so $z \in B(x, \varepsilon)$. □

Figure 8.2 encapsulates the whole proof.

Proposition 8.18. Any open set can be written as a union of open balls.

Proof. Take any open set $A \subseteq X$. Since A is open in X, for every $x \in A$, there exists $\varepsilon_x > 0$ such that

$$B(x, \varepsilon_x) \subseteq A$$

Since this holds for every x,

$$\bigcup_{x \in A} B(x, \varepsilon_x) \subseteq A$$

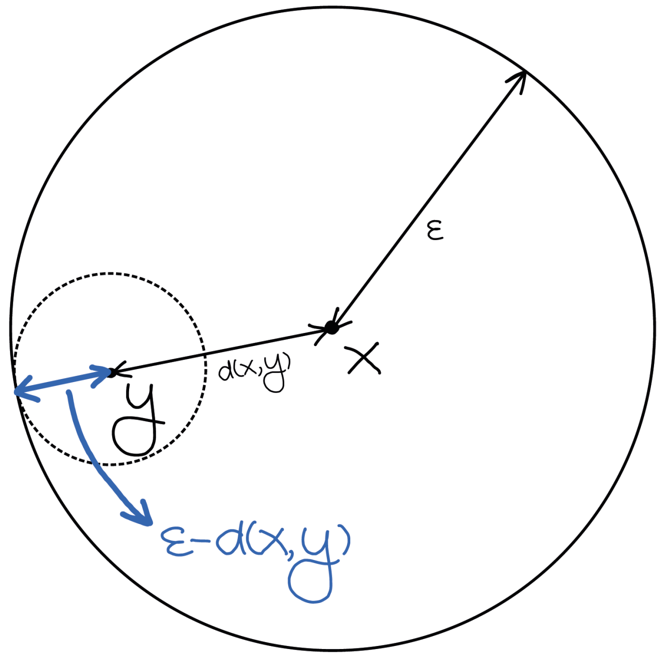


Figure 8.2: Open balls

For the other direction, take $x \in A$. Then

$$x \in B(x, \varepsilon_x) \subseteq \bigcup_{x \in A} B(x, \varepsilon_x)$$

Since this is true for every $x \in A$, $A \subseteq \bigcup_{x \in A} B(x, \varepsilon_x)$. Therefore, $A = \bigcup_{x \in A} B(x, \varepsilon_x)$. □

8.5 Closed sets

This is one definition of closed sets:

Definition 8.19 (Closed set). Let $A \subseteq X$. We say that A is closed if $X \backslash A := A^c$ is open in X, where A^c is the complement of A.

The complement of A is represented here as A^c, whereas in discrete maths (COMP40018), it is represented by $\overline{A}$. They are merely notational differences, but since the former is used in the notes on which this chapter is mostly based, I will stick with A^c for my own sanity.

Theorem 8.20. Let X be a metric space, and let $\{A_i\}_{i \in \Lambda}$ be closed sets in X.

1. $\varnothing$ and X are closed sets in X.
2. $\bigcap_{i \in I} A_i$ is closed in X. (The arbitrary intersection of closed sets is closed.)
3. $\bigcup_{i=1}^{n} A_i$ is closed in X. (The finite union of closed sets is closed.)

To prove the second and third item, we will need De Morgan's Laws for sets:

Theorem 8.21 (De Morgan's). Consider the sets $\{A_i\}_{i\in\Lambda}$ Then

$$\left(\bigcup_{i\in\Lambda} A_i\right)^c = \bigcap_{i\in\Lambda} A_i^c \qquad \text{and} \qquad \left(\bigcap_{i\in\Lambda} A_i\right)^c = \bigcup_{i\in\Lambda} A_i^c$$

Proof. 1. The complement of $\varnothing$ is $\varnothing^c = X$, which is open, so $\varnothing$ is closed. Similarly, $X^c = \varnothing$, so X is closed. Note that $\varnothing$ and X are open and closed at the same time, so these two properties are not mutually exclusive.

2. We want to show that

$$\left(\bigcap_{i\in I} A_i\right)^c = \bigcup_{i\in\Lambda} A_i^c$$

is open. Since A_i is closed for every $i \in \Lambda$, A_i^c is open by definition. From Theorem 8.16, we know that the arbitrary union of open sets is open.

3. We want to show that

$$\left(\bigcup_{i=1}^{n} A_i\right)^c = \bigcap_{i=1}^{n} A_i^c$$

is open. Since A_i is closed for every $1 \le i \le n$, A_i^c is open by definition. From Theorem 8.16, we know that the finite intersection of open sets is open. □

Proposition 8.22. Let (X, d) be a metric space, and $x \in X$. Then $\{x\}$ is a closed set in X.

Proof. We want to show that $Y := X \backslash \{x\}$ is open, and to show that Y is open, for every $y \in Y$, there must exist $\varepsilon > 0$ such that

$$B(y, \varepsilon) \subseteq Y$$

Take $y \in Y$, and let $\varepsilon = d(x, y)$. (To be extra intentional, we can let $\varepsilon = d(x, y)/2$ instead.) We know that $d(x, y) \neq 0$ since $x \notin Y$, and so $x \neq y \in Y$; then $d(x, y) > 0$ follows by positive definiteness. Since $d(x, y)$ is not strictly less than ε, $x \notin B(y, \varepsilon)$, so it follows that $B(y, \varepsilon) \subseteq Y$. □

We can prove that any finite set is closed in a similar manner. For any finite subset $A = \{x_1, x_2, \ldots, x_n\} \subseteq X$, we simply let $\varepsilon = \min_{1\le i\le n} d(x_i, y) = \min\{d(x_1, y), d(x_2, y), \ldots, d(x_n, y)\}$ instead.

The following is an alternative definition of closed sets, which we will show is equivalent to our previous definition.

Proposition 8.23 (PSET 2, Q2). Let (X, d) be a metric space. A subset $A \subseteq X$ is closed if and only if every convergent sequence in A converges in A.

Proof. Suppose $A \subseteq X$ is closed. Then A^c is open. Take any convergent sequence x_n in A (so $x_n \in A$ for every $n \in \mathbb{N}$), and suppose towards a contradiction that it converges to some $x \notin A$, i.e. $x \in A^c$. Since A^c is open, there exists $\varepsilon > 0$ such that $B(x, \varepsilon) \subseteq A^c$. However, for this ε, since $x_n \to x$, there exists $N \in \mathbb{N}$ such that for all $n > N$, $d(x_n, x) < \varepsilon$. So $x_n \in B(x, \varepsilon) \subseteq A^c$, which contradicts that x_n is a sequence in A.

Now suppose every convergent sequence in A converges in A. Suppose, towards a contradiction, that A^c is not open. Then there exists $y \in A^c$ such that for every $\varepsilon > 0$, $B(y, \varepsilon) \nsubseteq A^c$. Since $B(y, \varepsilon) \nsubseteq A^c$, this means there exists $x \in B(y, \varepsilon)$ such that $x \notin A^c$, so $x \in A$. We build a sequence x_n by taking $x_n \in B(y, 1/n)$ such that $x_n \in A$. Since $1/n \to 0$, it follows that $d(x_n, y) \to 0$, so $x_n \to y$. However, x_n

is a convergent sequence in A and $y \notin A$, contradiction. □

8.6 Neighbourhoods

There are several different definitions of neighbourhoods. We will adopt the definition on Wikipedia here:

Definition 8.24 (Neighbourhood). Let (X, d) be a metric space. A set U is a neighbourhood of a point $x \in X$ if there exists $\varepsilon > 0$ such that

$$B(x, \varepsilon) \coloneqq \{y \in X \mid d(x, y) < \varepsilon\} \subseteq U$$

Naturally, from this definition, $x \in B(x, \varepsilon) \subseteq U$, but other definitions (especially those concerning more general spaces than metric spaces) will specify that $x \in U$ as a requirement. An *open neighbourhood* is a neighbourhood which is also an open set.

Proposition 8.25. Let $\{x_n\}$ be a sequence in the metric space (X, d). Then, x_n converges to $x \in X$ if and only if for every neighbourhood of x, all but finitely many terms in x_n are in the neighbourhood of x.

Proof. Suppose $x_n \to x$ and take some arbitrary neighbourhood U of x. Then there exists $\varepsilon > 0$ such that $B(x, \varepsilon) \subseteq U$. For this ε, there exists $N \in \mathbb{N}$ such that $d(x_n, x) < \varepsilon$ for every $n > N$, and therefore, $x_n \in B(x, \varepsilon)$. There can only be finitely many terms (including and up to x_N, but not necessarily all of them) that are not in $B(x, \varepsilon)$, so all but finitely many terms are in $B(x, \varepsilon) \subseteq U$.

Now take an arbitrary $\varepsilon > 0$. The ball $B(x, \varepsilon)$ is a neighbourhood of x. Suppose all but finitely many terms in x_n are in $B(x, \varepsilon)$. Let x_N denote the last term that is not in the ball. (If no such term exists, then we are done.) Then for every $n > N$, $x_n \in B(x, \varepsilon)$. □

8.7 Continuous functions

We will show how continuous functions tie up the previous sections.

Proposition 8.26. Let (X, d_X) and (Y, d_Y) be metric spaces. Then $f : X \to Y$ is continuous at $c \in X$ if and only if for every sequence $\{x_n\}$ in X converging to c, $f(x_n) \to f(c)$.

Proof. Suppose f is continuous at c, and take any sequence $\{x_n\}$ in X that converges to c. We want to show that

$$\forall \varepsilon > 0,\ \exists N \in \mathbb{N} : n > N \Rightarrow d_Y(f(x_n), f(c)) < \varepsilon$$

Given that $x_n \to c$, we have

$$\forall \varepsilon > 0,\ \exists N \in \mathbb{N} : n > N \Rightarrow d_X(x_n, c) < \varepsilon$$

and given f continuous at c, we have for every $x \in X$,

$$\forall \varepsilon > 0,\ \exists \delta > 0 : d_X(x, c) < \delta \Rightarrow d_Y(x, c) < \varepsilon$$

Fix $\varepsilon > 0$. For this $\delta > 0$, by $x_n \to c$, we have

$$\exists N \in \mathbb{N} : n > N \Rightarrow d_X(x_n, c) < \delta$$

Therefore, by continuity of f, we have $d_Y(x_n, c) < \varepsilon$ for every $n > N$, as required.

Now suppose f is not continuous at c. We want to find a sequence $\{x_n\}$ such that $x_n \to c$ but $f(x_n) \not\to f(c)$. Given that f is not continuous at c,

$$\exists \varepsilon > 0 : \forall \delta > 0,\ \exists x \in X : d_X(x, c) < \delta \wedge d_Y(f(x), f(c)) \geq \varepsilon$$

Whenever we see $\forall \delta > 0$, think of $1/n$. For this $\varepsilon > 0$, we can construct a sequence x_n satisfying

$$d_X(x_n, c) < \frac{1}{n} \qquad \text{and} \qquad d_Y(f(x_n), f(c)) \geq \varepsilon > 0$$

By squeeze theorem, $d_X(x_n, c) \to 0$, so $x_n \to c$. However, $d_Y(f(x_n), f(c)) \not\to 0$, so $f(x_n) \not\to f(c)$. □

Proposition 8.27. Let (X, d_X) and (Y, d_Y) be metric spaces. Then $f : X \to Y$ is continuous at $c \in X$ if and only if for every open neighbourhood U of $f(c)$ in Y, the set $f^{-1}(U)$ contains an open neighbourhood of c in X.

Proof. Suppose f is continuous at c, and let U be an open neighbourhood of $f(c)$. We want to find some $\varepsilon > 0$ such that

$$B(c, \varepsilon) \subseteq f^{-1}(U)$$

Given f is continuous at c, for every $x \in X$,

$$\forall \varepsilon > 0,\ \exists \delta > 0 : d_X(x, c) < \delta \Rightarrow d_Y(f(x), f(c)) < \varepsilon$$

and given U is open in Y, there exists $\varepsilon > 0$ such that

$$B(f(c), \varepsilon) \coloneqq \{y \in Y \mid d(y, f(c)) < \varepsilon\} \subseteq U$$

For this ε, by continuity of f at c, there exists $\delta > 0$ such that for every $x \in X$,

$$d_X(x, c) < \delta \Rightarrow d_Y(f(x), f(c)) < \varepsilon$$

In other words, for every $x \in X$,

$$x \in B(c, \delta) \Rightarrow f(x) \in B(f(c), \varepsilon)$$

Since $x \in f^{-1}(f(x))$ (f^{-1} returns the preimage of its argument, which is a set in general), it follows that

$$B(c, \delta) \subseteq f^{-1}(B(f(c), \varepsilon)) \subseteq f^{-1}(U)$$

Now suppose for every open neighbourhood U of $f(c)$, $f^{-1}(U)$ contains an open neighbourhood of c. We want to show that for every $x \in X$,

$$\forall \varepsilon > 0,\ \exists \delta > 0 : d_X(x, c) < \delta \Rightarrow d_Y(f(x), f(c)) < \varepsilon$$

Fix $\varepsilon > 0$, and consider $U = B(f(c), \varepsilon)$. It is clearly an open neighbourhood of $f(c)$. Then $f^{-1}(U)$ contains an open neighbourhood V of c. By our definition of neighbourhoods, there exists $\delta > 0$ such that

$$B(c, \delta) \subseteq V \subseteq f^{-1}(U)$$

(An irrelevant but noteworthy point: we don't know if $f^{-1}(U)$ is open, but it is given that V is open.) In other words, for every $x \in X$,

$$x \in B(c, \delta) \Rightarrow x \in f^{-1}(U)$$

and therefore,
$$x \in B(c, \delta) \Rightarrow f(x) \in U = B(f(c), \varepsilon)$$
□

Proposition 8.28 (PSET 2, Q7)**.** Let (X, d_X) and (Y, d_Y) be metric spaces. Then $f : X \to Y$ is continuous if and only if for every set U open in Y, $f^{-1}(U)$ is open in X.

Proof. Suppose f is continuous. Then for every $c \in X$ we have
$$\forall \varepsilon > 0,\ \exists \delta > 0 : d_X(x, c) < \delta \Rightarrow d_Y(f(x), f(c)) < \varepsilon$$
Take any set U open in Y. We want to show that for every $x \in f^{-1}(U)$, there exists $\varepsilon > 0$ such that
$$B(x, \varepsilon) \subseteq f^{-1}(U)$$
Since U is open, for every $y \in Y$, there exists $\varepsilon > 0$ such that
$$B(y, \varepsilon) \subseteq U$$
For every $x \in f^{-1}(U)$, $f(x) \in U$. So there exists $\varepsilon > 0$ such that
$$B(f(x), \varepsilon) \subseteq U$$
Fix $x_0 \in X$. Then there exists $\varepsilon > 0$ such that $B(f(x_0), \varepsilon) \subseteq U$. Since f is continuous, there exists $\delta > 0$ such that for every $x \in X$,
$$x \in B(x_0, \delta) \Rightarrow f(x) \in B(f(x_0), \varepsilon) \subseteq U$$
so
$$x \in B(x_0, \delta) \Rightarrow x \in f^{-1}(U)$$
and therefore,
$$B(x_0, \delta) \subseteq f^{-1}(U)$$
This holds for every $x_0 \in X$, so $f^{-1}(U)$ is open, as required.

Now suppose for every set U open in Y, $f^{-1}(U)$ is open in X. We want to show that for every $c \in X$ and every $\varepsilon > 0$,
$$\exists \delta > 0 : d_X(x, c) < \delta \Rightarrow d_Y(f(x), f(c)) < \varepsilon$$
Fix $c \in X$ and $\varepsilon > 0$. Let $U = B(f(c), \varepsilon)$. It is clearly open in Y. ($f(c)$ exists since f is a function.) So $f^{-1}(U)$ is open in X. In particular, since $f(c) \in U$, it follows that $c \in f^{-1}(U)$. The rest follows exactly as the sufficient direction of the previous proof. Since $\varepsilon > 0$ is arbitrary, f is continuous at c. Since $c \in X$ is arbitrary, f is continuous. □