Analysis

Adam Kelly (`ak2316@cam.ac.uk`)

June 2, 2021

At its heart, analysis is the study of ideas that depend on the notion of *limits*. The main concepts of analysis (such as convergence, continuity, differentiation and integration) will all depend quite fundamentally on a limiting process.

This article constitutes my notes for the 'Analysis I' course, held in Lent 2021 at Cambridge. These notes are *not a transcription of the lectures*, and differ significantly in quite a few areas. Still, all lectured material should be covered[1].

Contents

---

[1]A tiny bit of analysis is assumed, namely the content covered in the 'Numbers and Sets' course at Cambridge. Specifically the reader should be aware of least upper bounds/suprema along with the least upper bound axiom. If the reader is unfamiliar with this content, they are referred to Chapter 2 of my Numbers and Sets notes

1 Sequences and Convergence

One of the fundamental objects of study in analysis is *sequences*. In particular, in almost all areas of this course we will be concerned (either implicitly or explicitly) with the notion of convergence.

1.1 Limits & The Reals

We will define convergence as follows.

Definition 1.1 (Convergence)**.** A sequence $a_1, a_2, \dots \in \mathbb{R}$ is said to *converge* to the limit $a \in \mathbb{R}$ if given any $\epsilon > 0$, we can find an integer N such that $|a_n - a| < \epsilon$ for all $n \geq N$. We write $\lim_{n \to \infty} a_n = a$, or $a_n \to a$ as $n \to \infty$.

This definition is (notably) purely algebraic. We can sensibly define this notion of convergence for any ordered field (for example, $\mathbb{Q}$). What takes us from algebra to analysis is the fundamental property of the real numbers.

Axiom 1.2 (Fundamental Axiom of Analysis)**.** *If $a_1, a_2, \dots \in \mathbb{R}$ is an increasing sequence and there exists $A \in \mathbb{R}$ such that $a_i \leq A$ for all $i \in \mathbb{N}$, then there exists $a \in \mathbb{R}$ such that $a_n \to a$ as $n \to \infty$.*

In other words – an increasing sequence of real numbers that is bounded above converges. Also this clearly implies that a decreasing sequence of reals bounded below converges.

Limits obey the properties that you would naturally expect.

Proposition 1.3 (Uniqueness of Limits)**.** *If $a_n \to a$ and $a_n \to b$ as $n \to \infty$, then $a = b$.*

Proof. Assume that $a \neq b$. Given given any $\epsilon > 0$, we can find integers N_1 and N_2 such that

$$|a_n - a| \leq \epsilon, \quad \text{for all } n \geq N_1$$
$$|a_n - b| \leq \epsilon, \quad \text{for all } n \geq N_2$$

Then letting $\epsilon = |a - b|/3$ and taking $N = \max\{N_1, N_2\}$, we have by the triangle inequality

$$|a - b| \leq |a_n - a| + |a_n - b| \leq 2\epsilon = \frac{2}{3}|a - b|$$

for all $n \geq N$. Thus we must have $|a - b| = 0$, and $a = b$. $\square$

Proposition 1.4 (Convergence of Subsequences)**.** *If $a_n \to a$ as $n \to \infty$ and $n(1) < n(2) < \cdots$, then $a_{n(j)} \to a$ as $j \to \infty$.*

Proof. We note that $n(j) \geq j$. Now $a_n \to a$ implies that given some ϵ, we can find an N such that $|a_j - a| < \epsilon$ for all $j \geq N$. But then this implies that $|a_{n(j)} - a| < \epsilon$ for all $j \geq N$, since $j \geq N$ implies $n(j) \geq N$. So $a_{n(j)} \to a$ also. $\square$

Proposition 1.5 (Manipulating Limits)**.** *Let a_n and b_n be sequences. Then the following hold.*

(i) *If $a_n \to a$ and $b_n \to b$ then $a_n + b_n \to a + b$.*

(ii) *If $a_n \to a$ then $ca_n \to ac$ for a constant c.*

(iii) *If $a_n \to a$ and $b_n \to b$ then $a_n b_n \to ab$.*

(iv) *If $a_n \to a$ and $a, a_n \neq 0$ for all n, then $1/a_n \to 1/a$.*

Proof. We prove each individually.

(i) Given some $\epsilon > 0$, we can find integers N_a and N_b such that $|a_n - a| \leq \epsilon/2$ for all $n \geq N_a$ and $|b_n - b| \leq \epsilon/2$ for all $n \geq N_b$.

Then letting $N = \max\{N_a, N_b\}$, by the triangle inequality we have

$$|(a_n - a) + (b_n - b)| \leq |a_n - a| + |b_n - b| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

for all $n \geq N$. Thus $a_n + b_n \to a + b$.

(ii) Given $\epsilon > 0$, we can find some N such that $|a_n - a| \leq \epsilon/|c|$ for all $n \geq N$. Then $|ca_n - ca| \leq \epsilon$ for all $n \geq N$. So $ca_n \to ca$.

(iii) Given some $\epsilon > 0$, we can find integers N_a and N_b such that $|a_n - a| \leq \sqrt{\epsilon}$ for all $n \geq N_a$ and $|b_n - b| \leq \sqrt{\epsilon}$ for all $n \geq N_b$. Then again letting $N = \max\{N_a, N_b\}$, we have

$$|(a_n - a)(b_n - b)| \leq \epsilon,$$

for all $n \geq N$. Hence $(a_n - a)(b_n - b) \to 0$, so $a_n b_n - ab_n - ba_n + ab \to 0$, and using the previous two properties we have $a_n b_n \to ab$.

(iv) Since the sequence converges to $a \neq 0$, there must be some $r > 0$ such that $|a_n| > r$, for all n. Then given some $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that $|a_n - a| < |a|\epsilon r$ for all $n \geq N$. That is,

$$\left| \frac{1}{a_n} - \frac{1}{a} \right| = \left| \frac{a - a_n}{aa_n} \right| < \frac{|a|\epsilon r}{|a|r} = \epsilon,$$

for all $n \geq N$. Hence $1/a_n \to 1/a$. $\qquad \square$

Proposition 1.6 (Squeeze Theorem). *Let a_n, b_n and x_n be sequences such that $a_n \leq x_n \leq b_n$ for all n. Then if $a_n \to \ell$ and $b_n \to \ell$ as $n \to \infty$, we also have $x_n \to \ell$.*

Proof. Given some $\epsilon > 0$, we can find integers N_a and N_b such that $|a_n - \ell| < \epsilon$ for all $n \geq N_a$ and $|b_n - \ell| < \epsilon$ for all $n \geq N_b$. Then letting $N = \max\{N_a, N_b\}$, we have $\ell - \epsilon < a_n \leq x_n \leq b_n < \ell + \epsilon$ for all $n \geq N$. That is, $|x_n - \ell| < \epsilon$ for all $n \geq N$. Hence $x_n \to \ell$. $\qquad \square$

With these results in the our toolbox, we can then prove our first actual analysis result.

Lemma 1.7 (Axiom of Archimedes). $\frac{1}{n} \to 0$ *as $n \to \infty$.*

Proof. The sequence $\frac{1}{n}$ is a decreasing sequence bounded below, and thus has a limit a. Considering the sequence $\frac{1}{2n} = \frac{1}{2} \cdot \frac{1}{n}$, this tends to a limit $\frac{a}{2}$, but since $\frac{1}{2n}$ is a subsequence of $\frac{1}{n}$, it also tends to a limit a. Thus $\frac{a}{2} = a$, so $a = 0$ as required. $\qquad \square$

Now while this is an article about real analysis, we will frequently employ the complex numbers. Of course, to do that we need to be able to do some analysis with $\mathbb{C}$, and indeed the definition of a limit still makes sense in $\mathbb{C}$. All of the above properties also hold, apart from the squeeze theorem (we need to be careful because $\mathbb{C}$ cannot be ordered like $\mathbb{R}$!)

1.2 Bolzano–Weierstrass

An equivalent but important and quite useful form of the fundamental axiom is the 'Bolzano-Weierstrass theorem', that every bounded sequence has a convergent subsequence.

Theorem 1.8 (Bolzano-Weierstrass). *If $a_n \in \mathbb{R}$ and there exists K such that $|a_n| \leq K$ for all n, then we can find $n_1 < n_2 < n_3 < \cdots$ and $a \in \mathbb{R}$ such that $a_{n_j} \to a$ as $j \to \infty$.*

Proof. Consider a real sequence a_n. We say that an element a_i of the sequence is *good* if $a_j \geq a_i$ for all $j \geq i$. If the sequence has an infinite number of good elements, then taking them as a subsequence will be increasing. If there is a finite number of good elements, then letting N be the index of the last good element, we have that for any a_j with $j \geq N + 1$, there exists some a_k with $k > j$ such that $a_k < a_j$. Then repeating this, we can get a collection of decreasing elements. Thus every real sequence has a monotonic subsequence. Since the the sequence is bounded, we then have that this subsequence converges. $\square$

Remark. This theorem says nothing about the uniqueness of the subsequence's limit. For example, consider the sequence $x_n = (-1)^n$. Then $x_{2n+1} \to -1$ and $x_{2n} \to 1$.

The proof given above is quite clean but is not the only standard proof of the theorem. Another common method which involves the method of 'repeated bisection', which is colloquially known as 'lion hunting'. This method will be discussed shortly.

The Bolzano-Weierstrass theorem also holds in $\mathbb{C}$ (and the same proof method will show with a little induction that it holds in $\mathbb{R}^n$, but we won't dwell on that in this article).

Theorem 1.9 (Bolzano-Weierstrass in $\mathbb{C}$). *If $a_n \in \mathbb{C}$ is a bounded sequence, then it has a convergent subsequence.*

Proof. Begin by writing $a_n = x_n + iy_n$. Then since a_n is bounded, the sequences x_n and y_n are also bounded. We can apply Bolzano–Weierstrass to x_n to get a subsequence $x_{n(1)}, x_{n(2)}, \ldots$ with $x_{n(j)} \to x$ as $j \to \infty$, for some $x \in \mathbb{R}$. Then we can apply Bolzano-Weierstrass again to $y_{n(j)}$ to get a subsequence $y_{n(k(1))}, y_{n(k(2))}, \ldots$ with $y_{n(k(j))} \to y$ as $j \to \infty$ for some $y \in \mathbb{R}$.

Thus $x_{n(k(j))} \to x$ and $y_{n(k(j))} \to y$ and thus $a_{n(k(j))} \to x + iy$ as $j \to \infty$, so we have a convergent subsequence. $\square$

Before we end our discussion of the Bolzano–Weierstrass theorem, we will take a short digression on Lion Hunting (you need to know this proof though!)

Aside: How To Hunt Lions

The term 'lion hunting' comes most likely from the satirical paper 'A Contribution to the Mathematical Theory of Big Game Hunting', by a group of mathematicians under the pseudoname 'H. Pétard'. They present the following method of hunting a lion (that we know exists) in the Sahara Desert.

> Bisect the desert by a line running N-S. The lion is either in the E portion or in the W portion; let us suppose him to be in the W portion. Bisect this portion by a line running E-W. The lion is either in the N portion or in the S portion; let us suppose him to be in the N portion. We continue this process. indefinitely, constructing a sufficiently strong fence about the chosen portion at each step. The diameter of the chosen portions approaches zero, so that the lion is ultimately surrounded by a fence of arbitrarily small perimeter.

While this is tongue and cheek, it does give us a method of finding things by 'bisecting intervals', which we can apply to quite a few different theorems in Analysis.

Now it's time to prove Bolzano-Weierstrass by hunting lions.

Proof (Bolzano-Weierstrass, Lion Hunting Style). We are going to define two sequences a_n and b_n inductively as follows. Begin by setting $[a_1, b_1] = [-K, K]$, and let $c_1 = (a_1 + b_1)/2$ be the midpoint of this interval.

Then there is two possibilities:

1. $x_n \in [a_1, c_1]$ for infinitely many values of n.

2. $x_n \in [c_1, b_1]$ for infinitely many values of n.

Of course both of these can hold at the same time, but if the first one holds we set $a_2 = a_1$ and $b_2 = c_1$, and if it doesn't then we set $a_2 = c_1$ and $b_2 = b_1$.

Repeating this process we construct a_n and b_n such that $x_m \in [a_n, b_n]$ for infinitely many values of m[2]. Then we have $a_{n-1} \le a_n \le b_n \le b_{n-1}$, and also $b_n - a_n = (b_{n-1} - a_{n-1})/2$.

Now a_n is increasing and bounded above, and b_n is decreasing and bounded below and thus $a_n \to a \in [a_1, b_1]$ and $b_n \to b \in [a_1, b_1]$. Then we have $b - a = (b - a)/2$ using the above result, and thus $a = b$.

Since $x_m \in [a_n, b_n]$ for infinitely many values of m, we can construct a sequence n_j such that $n_{j+1} > n_j$ and $x_{n_{j+1}} \in [a_{j+1}, b_{j+1}]$. Then $a_j \le x_{n_j} \le b_j$, and thus $x_{n_j} \to a$. □

---

[2]This is lion hunting! You can kind of imagine that we are hunting for a number that a subsequence converges to, using the fact that there must be infinitely many terms near that number

1.3 Cauchy Sequences & The General Principle of Convergence

So far we have defined the notion of a sequence converging to some explicit limit. However, it is possible to determine if a sequence converges without considering this limit explicitly. This is done by considering how 'close' the terms in the sequence eventually get, as we shall see.

We begin by describing what it means for a sequence to be *Cauchy*.

Definition 1.10 (Cauchy Sequence). A sequence $a_1, a_2, \dots \in \mathbb{C}$ is said to be *Cauchy* if for every $\epsilon > 0$ there exists an integer N such that for all $n, m \geq N$ we have $|a_n - a_m| \leq \epsilon$.

We can almost immediately write down our first lemma.

Lemma 1.11 (Convergence Implies Cauchy). *If a sequence converges then it is Cauchy.*

Proof. Consider a convergent sequence $a_n \to a$. Given $\epsilon > 0$, we can find N such that $|a_n - a| \leq \epsilon/2$ for all $n \geq N$. Then by the triangle inequality we have
$$|a_n - a_m| \leq |a_n - a| + |a - a_m| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$
for all $m, n \geq N$. Thus the sequence is Cauchy. $\qquad\square$

The converse of this result is also true! This gives us a powerful result about convergence, which is quite widely applicable (particularly because we can avoid talking about the limit explicitly, as we said before).

Lemma 1.12 (Completeness). *If a sequence is Cauchy then it converges.*

Proof. Let $a_1, a_2, \dots$ be a Cauchy sequence. We will first show that the sequence is bounded. Because the sequence is Cauchy, can find an integer $N \in \mathbb{N}$ such that $n, m \geq N$ implies that $|a_n - a_m| \leq 1$. This implies that if $n \geq N$ we have
$$|x_n| \leq |x_n - x_N| + |x_N| \leq 1 + |x_N|.$$
Thus we have $|x_n| \leq \max_{1 \leq r \leq N} |x_r| + 1$ for all n, so the sequence is bounded.

Applying the Bolzano–Weierstrass theorem, we then have a subsequence $a_{n_1}, a_{n_2}, \dots$ with $a_{n_j} \to a$ as $j \to \infty$.

So given $\varepsilon > 0$, we can find M such that $|a_{n_j} - a| \leq \varepsilon/2$ for all $n_j \geq M$. We can also find M' such that $n, m \geq M'$ implies $|a_n - a_m| < \varepsilon/2$.

Then for all $n \geq M'$, choosing n_j such that $n_j \geq \max\{M, M'\}$, we have

$$|a_n - a| \leq |a_n - a_{n_j}| + |a_{n_j} - a| \leq \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

Thus the sequence converges (namely to a). $\qquad\square$

Stripping away the detail, the reason this result holds is since Cauchy sequences are bounded, we can get a convergent subsequence. Then since all of the terms get arbitrarily close, the whole sequence must converge along with the subsequence.

Combining these two lemmas gives us the 'general principle of convergence', a result also known as Cauchy's criterion.

Theorem 1.13 (General Principle of Convergence/Cauchy's Criteron). *A sequence converges if and only if it is a Cauchy sequence.*

1.4   Limits of Functions

So far we have developed the notion of a limit of a sequence, but we can sensibly come up with a definition for what limits are in the context of functions. We will need to do a little bit of setup first, however.

We first need to be able to distinguish points (element of $\mathbb{C}$) where its possible to converge to those points with a sequence of elements that are all in some set A.

Definition 1.14 (Limit Point). Let $A \subseteq \mathbb{C}$ and $a \in \mathbb{C}$. We say that a is a *limit point* of A if for any $\delta > 0$ there is some $z \in A$ such that $0 < |z - a| < \delta$.

We can then define a limit of a function by its behaviour as we approach a given limit point.

Definition 1.15 (Limit of a Function). Let $A \subseteq \mathbb{C}$ and let $a \in \mathbb{C}$ be a limit point of A. Then for $f : A \to \mathbb{C}$, we say that the *limit of f* as z approaches a is ℓ, written $\lim_{z \to a} f(z) = \ell$, if given $\epsilon > 0$ there is some $\delta > 0$ such that whenever $0 < |z - a| < \delta$ and $z \in A$ we have $|f(z) - \ell| < \epsilon$.

These definitions should match the informal notion of a limits that would have been given in a calculus course. The reader should note that there is no requirement that $f(a) = \ell$, or even that f be defined at a at all.

There is a natural relation between this notion of a limit point and a limit and our previous definition of limits of sequences.

Proposition 1.16 (Sequence Definition of Limit Points). *Let $A \subseteq \mathbb{C}$ and $a \in \mathbb{C}$. Then a is a limit point of A if and only if there is a sequence $z_n \in A$ such that $z_n \to a$ as $n \to \infty$ and $z_n \neq a$ for all n.*

Proof. If a is a limit point then for every positive integer n we can take $\delta = 1/n$ and obtain a $z_n \in A$ such that $0 < |z_n - a| < 1/n$. Then by the squeeze theorem $z_n \to a$ as $n \to \infty$, and also $z_n \neq a$ for all n.

Conversely, if there is such a sequence $z_n \in A$, then given $\delta > 0$ there is an N such that $z_N \in A$ and $0 < |z_N - a| < \delta$, so a is a limit point of A. $\square$

Proposition 1.17 (Sequence Definition of Limits). *Let $A \subseteq \mathbb{C}$ and let $a \in \mathbb{C}$ be a limit point of A. Then for $f : A \to \mathbb{C}$ we have $\lim_{z \to a} f(z) = \ell$ if and only if $\lim_{n \to \infty} f(z_n) = \ell$ for every sequence $z_n \in A$ satisfying $z_n \neq a$ and $z_n \to a$ has $f(z_n) \to \ell$.*

Proof. If $\lim_{z \to a} f(z) = \ell$, consider a sequence $z_n \in A$ where $z_n \neq a$ and $z_n \to a$. Given $\epsilon > 0$, there exists $\delta > 0$ such that $0 < |z - a| < \delta$ and $z \in A$ implies $|f(z) - \ell| < \epsilon$.

Also there exists N such that $n \geq N$ implies $0 < |z_n - a| < \delta$. Thus for $n \geq N$ we have $|f(z_n) - \ell| < \epsilon$, so $\lim_{n \to \infty} f(z_n) = \ell$.

Conversely, if $\lim_{z \to a} f(z)$ was not ℓ, then there would be some ϵ such that for every δ there is a $z \in A$ such that $0 < |z - a| < \delta$ but $|f(z) - \ell| \geq \epsilon$. Then taking $\delta = 1/n$ for each positive integer n we could construct a sequence $z_n \in A$ with $z_n \neq 0$ where $0 < |z - a| < 1/n$ but $|f(z_n) - \ell| \geq \ell$ for all n. But then $z_n \to a$ and $f(z_n) \not\to \ell$, which is a contradiction. $\square$

In practice, which of these definitions is more useful will depend on context, however the sequences definition does allow us to use the results we developed at the start of this section.

Proposition 1.18 (Properties of Limits of Functions). *Let $A \subseteq \mathbb{C}$ and let a be a limit point of A. Then if $f, g : A \to \mathbb{C}$ are functions the following hold.*

(i) *If $\lim_{z \to a} f(z) = \ell$ and $\lim_{z \to a} g(z) = m$, then $\lim_{z \to a} f(z) + g(z) = \ell + m$.*

(ii) *If $\lim_{z \to a} f(z) = \ell$ then $\lim_{z \to a} cf(z) = c\ell$.*

(iii) *If $\lim_{z \to a} f(z) = \ell$ and $\lim_{z \to a} g(z) = m$ then $\lim_{z \to a} f(z)g(z) = \ell m$.*

(iv) *If $\lim_{z \to a} f(z) = \ell$ and $\ell \neq 0$ then $\lim_{z \to a} 1/f(z) = 1/\ell$.*

Proof. These all follow directly from the sequence definition of limits along with Theorem 1.5. $\square$

2 Infinite Series

The notion of convergence lets us talk quite sensibly about what it would mean to add up an infinite number of things, which is quite exciting.

2.1 Convergent & Divergent Series

The definition of convergence for an infinite series is rather natural, and comes from considering the sequence of partial sums.

Definition 2.1 (Convergence of an Infinite Series)**.** For a sequence $a_j \in \mathbb{C}$, we say that the series $\sum_{j=1}^{\infty} a_j$ *converges* to S if the sequence of partial sums converges to S. That is, $\sum_{j=1}^{n} a_j \to S$ as $n \to \infty$. Otherwise we say that it *diverges*.

Remark (Notation)**.** If a series converges, then we will typically write $\sum_{j=1}^{\infty} a_j = S$, but some care is needed as if the series does not converge then this is nonsense. We will also write S_n to denote the partial sum of a series, if it is clear from context what series is being referred to.

In this chapter we will be primarily concerned with finding ways to show whether a given series converges. But first, a few preliminaries.

Proposition 2.2 (Adding Series)**.** *If two series converge, say $\sum_{j=1}^{\infty} a_j = a$ and $\sum_{j=1}^{\infty} b_j = b$, then $\sum_{j=1}^{\infty} (\lambda a_j + \mu b_j)$ also converges, and we have*

$$\sum_{j=1}^{\infty} (\lambda a_j + \mu b_j) = \lambda a + \mu b.$$

Proof. Considering the sequence of partial sums we have

$$\sum_{j=1}^{n} (\lambda a_j + \mu b_j) = \lambda \sum_{j=1}^{n} a_j + \mu \sum_{j=1}^{n} b_j \longrightarrow \lambda a + \mu b$$

as $n \to \infty$, as required. $\qquad\square$

It should also be intuitively clear that the first few terms of an infinite series will not affect its convergence.

Proposition 2.3 (Ignoring Initial Terms)**.** *Suppose there exists N such that $a_j = b_j$ for all $j \geq N$. Then either $\sum_{j=1}^{\infty} a_j$ and $\sum_{j=1}^{\infty} b_j$ both converge or they both diverge.*

Proof. For $n \geq N$, we have $\sum_{j=1}^{n} a_n - \sum_{j=1}^{n} b_n = \sum_{j=1}^{N} a_n - \sum_{j=1}^{N} b_n$, which is a constant. Thus the sequences of partial sums either both converge or diverge. $\qquad\square$

We can also apply the general principle of convergence to infinite series. This gives us another useful way of proving that a series converges, particularly when we don't know what the series converges to which is often the case.

Theorem 2.4 (General Principle of Convergence/Cauchy's Criterion for Series). *The infinite series $\sum_{j=1}^{\infty} a_j$ converges if and only if for every $\epsilon > 0$ there is an integer N such that for all $n, m \geq N$ we have $|\sum_{j=n}^{m} a_j| \leq \epsilon$.*

Proof. This follows directly from the definition of the sequence of partial sums being Cauchy. $\square$

2.2 Convergence Tests

In this section we will develop some 'tests' of convergence, so that given some infinite series we will have some tools to try and determine relatively quickly if it converges and diverges.

We begin with the observation that if the terms of the series do not tend to 0, then the series must diverge.

Proposition 2.5 (Limit of Terms in a Convergent Series). *Suppose that $\sum_{j=1}^{\infty} a_j$ converges. Then $a_j \to 0$ as $j \to \infty$.*

Proof. Let $S_n = \sum_{j=1}^{n} a_j$. Then $S_n \to S$ for some S, and thus we must have $S_{n+1} \to S$. Adding these sequences we get $S_{n+1} - S_n \to 0$, that is, $a_{n+1} \to 0$ as $n \to \infty$ as required. $\square$

Somewhat unfortunately, while necessary this is not a sufficient condition for convergence. The most commonly cited counterexample to this is the Harmonic series.

Example 2.6 (Divergence of the Harmonic Series). Consider the infinite series with $a_k = 1/k$. Then $a_k \to 0$, but $\sum_{k=1}^{\infty} a_k$ diverges.

To see this, let $H_n = \sum_{k=1}^{n} a_k$ be the partial sum of the series. Then we have

$$H_{2n} = H_n + \frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{n+n}$$
$$\geq H_n + \underbrace{\frac{1}{2n} + \frac{1}{2n} + \cdots + \frac{1}{2n}}_{n \text{ times}} = H_n + \frac{1}{2}.$$

So if the series converged, we would have $H_n \to a$ and $H_{2n} \to a$, giving $a \geq a + 1/2$, which is a contradiction.

The result about the limit of terms in a series can be helpful though! For example, let's consider a series you are likely familiar with.

Example 2.7 (Convergence/Divergence of the Geomeric Series)**.** Consider the infinite series $\sum_{j=1}^{\infty} x^{j-1}$. Writing it's partial sum as $S_n = \sum_{j=1}^{n} x^{j-1}$, A little bit of algebra gives us that

$$S_n = \begin{cases} \frac{1-x^n}{1-x} & \text{if } x \neq 1, \\ n & \text{if } x = 1. \end{cases}$$

So if $|x| < 1$, then $x^n \to 0$ and $S_n \to \frac{1}{1-x}$. Otherwise, if $|x| \geq 1$ then the series cannot converge since the terms do not tend to 0.

It's best to think of Theorem 2.5 as giving the sort of a 'bare minimum' property that a series has to have in order to converge. Of course, it's also quite helpful for sanity checks!

2.2.1 Series of Non-Negative Terms

We just saw a necessary condition for convergence, now let's have a look at some *sufficient* conditions for convergence. For the time being we will restrict ourselves to considering infinite series where all of the terms are *real and non-negative.*

The first result we have is a direct consequence of our fundamental axiom.

Theorem 2.8 (Bounded Partial Sums)**.** *If a_n is a non-negative sequence and the partial sums S_n are bounded above, then $\sum_{n=1}^{\infty} a_n$ converges.*

Proof. Since all of the terms are non-negative, S_n is a bounded and increasing sequence, and thus converges by our fundamental axiom. $\square$

In a similar vein is the comparison test, which allows us to show convergence by comparing a series with another series whose convergence we know.

Theorem 2.9 (The Comparison Test)**.** *Suppose that $0 \leq a_j \leq b_j$ for all j. Then if $\sum_{j=1}^{\infty} b_j$ converges, so does $\sum_{j=1}^{\infty} a_j$.*

Proof. Since $B_n = \sum_{j=1}^{n} b_j$ is an increasing sequence whose limit is say B, we have $B_n \leq B$ for all n. Thus $\sum_{j=1}^{n} a_n \leq B_n \leq B$, so we have a series of non-negative terms whose partial sums are bounded. Thus $\sum_{j=1}^{\infty} a_n$ converges. $\square$

Example 2.10 (Using the Comparison Test)**.** We will prove that $\sum_{n=1}^{\infty} \frac{1}{n^2}$ converges using the comparison test.

Since $\frac{1}{n^2} < \frac{1}{n(n-1)}$ for $n \geq 2$, and $\sum_{n=2}^{\infty} \frac{1}{n(n-1)}$ converges as $\sum_{n=2}^{N} \frac{1}{n(n-1)} = 1 - \frac{1}{N} \to 1$ as $N \to \infty$, we get that $\sum_{n=1}^{\infty} \frac{1}{n^2}$ converges by comparison.

Using the comparison test we can derive two more useful tests, both of which come from comparing a series with the geometric series. You'll notice that the proofs for both results are relatively similar!

Theorem 2.11 (The Root Test). *Let a_n be a non-negative sequence and suppose $\sqrt[n]{a_n} \to a$ as $n \to \infty$. Then if $a < 1$, $\sum_{n=1}^{\infty} a_n$ converges, and if $a > 1$, $\sum_{n=1}^{\infty} a_n$ diverges.*

Proof. If $a < 1$, we can choose b such that $a < b < 1$, and there exists an integer N such that for all $n \geq N$ we have $\sqrt[n]{a_n} < b$, that is, $a_n < b^n$. But then $\sum_{n=N}^{\infty} b^n$ converges since $b < 1$. Thus by comparison $\sum_{n=1}^{\infty} a_n$ converges too.

If $a > 1$, then for $n \geq N$ $\sqrt[n]{a_n} > 1$ implies that $a_n > 1$, but then the terms in the series do not tend to zero and thus $\sum_{n=1}^{\infty} a_n$ diverges. $\square$

Theorem 2.12 (The Ratio Test). *Let a_n be a non-negative sequence and suppose that $\frac{a_{n+1}}{a_n} \to \ell$ as $n \to \infty$. Then if $\ell < 1$, $\sum_{n=1}^{\infty} a_n$ converges, and if $\ell > 1$, $\sum_{n=1}^{\infty} a_n$ diverges.*

Proof. If $\ell < 1$, we can choose b such that $\ell < b < 1$, and there exists an integer N such that $\frac{a_{n+1}}{a_n} < b$ for all $n \geq N$. Therefore

$$a_n = \frac{a_n}{a_{n-1}} \cdot \frac{a_{n-1}}{a_{n-2}} \cdots \frac{a_{N+1}}{a_N} \cdot a_N$$
$$< \left(a_N b^{-N}\right) b^n,$$

for $n > N$. But then $\sum_{n=N+1}^{\infty} b^n$ converges as $b < 1$, and since $a_N b^{-N}$ is a constant, $\sum_{n=1}^{\infty} a_n$ converges by comparison.

If $\ell > 1$, then $a_{n+1} > a_n$ and the terms in the series do not tend to zero and thus $\sum_{n=1}^{\infty} a_n$ diverges. $\square$

Remark (A Deadly Sin). In both the root and ratio test, if we find that either $a = 1$ or $\ell = 1$, we *cannot draw any conclusions about the convergence of the series.* To see this, consider $\sum \frac{1}{n}$ and $\sum \frac{1}{n^2}$.

Some examples of using the root and ratio tests are shown below, but we won't do too many since there's a most likely a few already on your example sheets (and there's not really much to the basic technique).

Example 2.13 (Using the Ratio Test). We will prove that $\sum_{n=1}^{\infty} \frac{n}{2^n}$ converges using the ratio test.

Let $a_n = n/2^n$. Then we have $\frac{a_{n+1}}{a_n} = \frac{n+1}{2^{n+1}} \cdot \frac{2^n}{n} = \frac{n+1}{2n}$, and thus $\frac{a_{n+1}}{a_n} \to \frac{1}{2} < 1$ as $n \to \infty$. Then by the ratio test $\sum_{n=1}^{\infty} \frac{n}{2^n}$ converges.

Example 2.14 (Using the Root Test). We will prove that $\sum_{n=1}^{\infty} \left(\frac{n+1}{3n+5}\right)^n$ converges using the root test.

Let $a_n = \left(\frac{n+1}{3n+5}\right)^n$. Then $\sqrt[n]{a_n} = \frac{n+1}{3n+5}$, and thus $\sqrt[n]{a_n} \to \frac{1}{3}$ as $n \to \infty$. Then by the root test $\sum_{n=1}^{\infty} \left(\frac{n+1}{3n+5}\right)^n$ converges.

In the case that we have our sequence of non-negative terms is decreasing, we can determine the convergence by looking at a related series involving a relatively small subsequence of terms.

Theorem 2.15 (Cauchy's Condensation Test). *If a_n is a decreasing sequence of positive numbers, then $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=0}^{\infty} 2^n a_{2^n}$ either both converge or both diverge.*

Proof. It suffices to show that the partial sums of both series are either both bounded or both unbounded. Let $S_n = \sum_{j=1}^{n} a_j$ and $T_n = \sum_{j=0}^{n} 2^j a_{2^j}$ be the partial sums of each series.

If $n < 2^k$ we have

$$S_n \leq a_1 + (a_2 + a_3) + \cdots + (a_{2^k} + \cdots + a_{2^{k+1}-1})$$
$$\leq a_1 + 2a_2 + \cdots + 2^k a_{2^k} = T_k,$$

thus $S_n \leq T_k$. So if T_n is bounded so is S_n. Also if $n > 2^k$, we have

$$S_n \geq a_1 + a_2 + (a_3 + a_4) + \cdots + (a_{2^{k-1}+1} + \cdots + a_{2^k})$$
$$\geq \frac{1}{2}a_1 + a_2 + 2a_4 + \cdots + 2^{k-1}a_{2^k} = \frac{1}{2}T_k,$$

so $2S_n \geq T_k$. So if S_n is bounded so is T_k. Thus S_n and T_n are either both bounded or both unbounded. $\qquad\square$

You may notice that this proof somewhat resembles our proof that the harmonic series diverges, and indeed you can show the following stronger result.

Theorem 2.16 (Convergence of $\sum n^{-\alpha}$). *$\sum_{n=1}^{\infty} \frac{1}{n^\alpha}$ converges if and only if $\alpha > 1$.*

Proof. If $\alpha \leq 0$ then the terms in the series do not tend to zero, so the series diverges. So for $\alpha > 0$ we apply Cauchy's condensation test. This series converges if and only if the series

$$\sum_{k=0}^{\infty} 2^k \cdot \frac{1}{2^{\alpha k}} = \sum_{k=0}^{\infty} 2^{(1-\alpha)k}$$

converges. By comparison with the geometric series, this will converge if and only if $2^{1-\alpha} < 1$, that is, $\alpha > 1$. $\qquad\square$

In a way Cauchy's condensation test is slightly different to the others convergence tests since it just gives us a different series that may be easier to show convergence or divergence for.

Still, it is natural to reach for Cauchy's condensation test when the root or ratio tests fail. For example, all of the series below cannot be shown to converge or diverge using the root or ratio tests:

$$\sum_{n=1}^{\infty} \frac{1}{n^\alpha}, \quad \sum_{n=2}^{\infty} \frac{1}{n \log n}, \quad \sum_{n=2}^{\infty} \frac{1}{(\log n)^2}, \quad \sum_{n=2}^{\infty} \frac{1}{n(\log n)^2}, \quad \text{and} \quad \sum_{n=2}^{\infty} \frac{\log n}{n^2}.$$

If you try, you will find that these tests are inconclusive. However, all can be shown to converge/diverge using Cauchy's condensation test.[3]

2.2.2 Alternating Series

We will now look at a convergence test for series whose terms alternate from positive to negative.

Theorem 2.17 (Alternating Series Test). *If a_n is a decreasing sequence of positive numbers with $a_n \to 0$ as $n \to \infty$ then $\sum_{j=1}^{\infty}(-1)^{j+1}a_j$ converges.*

Proof. Let $S_n = \sum_{j=1}^{n}(-1)^{j+1}a_j$ denote the partial sum of the series. Then $S_{2n} = S_{2n-2} + (a_{2n-1} - a_{2n}) \geq S_{2n-2}$. Also

$$S_{2n} = a_1 - (a_2 - a_3) - \cdots - (a_{2n-2} - a_{2n-1}) - a_{2n} \leq a_1.$$

Therefore S_{2n} is increasing and bounded above, and thus converges. Now let $S_{2n} \to S$ as $n \to \infty$. Then $S_{2n+1} = S_{2n} + a_{2n+1} \to S + 0 = S$ as $n \to S$. Thus $S_n \to S$, and $\sum_{j=1}^{\infty}(-1)^{j+1}a_j$ converges. $\square$

Example 2.18. We will prove that $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}$ converges using the alternating series test.[4]

We note that $\frac{1}{n}$ is a decreasing sequence of positive numbers with $\frac{1}{n} \to 0$. Thus by the alternating series test $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}$ converges.

2.3 Absolute Convergence

To finish our discussion on infinite series (for now!) we will introduce the stronger notion of *absolute convergence*. The main motivation for introducing this will become clear as we develop some of the properties of absolute convergence, but informally knowing that a series converges absolutely will allow us to be a little less careful when working with the series.

Definition 2.19 (Absolute Convergence). For some sequence $a_n \in \mathbb{C}$, we say that the series $\sum_{n=1}^{\infty} a_n$ converges *absolutely* if the series $\sum_{n=1}^{\infty} |a_n|$ converges.

---

[3] You could also use the integration test (which we have not discussed), but that's most likely going to be slower as you have to integrate things.

[4] Later on we will see that this is equal to $\log 2$.

We can see that absolute convergence is a strictly stronger property than 'regular' convergence. Indeed absolute convergence implies convergence, but not the other way round.

Theorem 2.20 (Absolute Convergence Implies Convergence). *If $\sum_{j=1}^{\infty} a_j$ converges absolutely then it converges.*

Proof. By the triangle inequality we have

$$\left| \sum_{j=n}^{m} a_j \right| \leq \sum_{j=n}^{m} |a_j| \, ,$$

and thus by Cauchy's criterion $\sum_{j=1}^{\infty} a_j$ converges. $\qquad\square$

To see that the converse isn't true (and that indeed it is a strictly stronger notion), consider the series $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}$, which converges but does not converge absolutely. We sometimes say such series converge *conditionally*.

When trying to determine whether a series converges absolutely, since $|a_n| \geq 0$ we are free to apply all of the results that we developed in the previous subsection. We will see this in the next example.

Example 2.21 (Showing Absolute Convergence). Consider the infinite series $\sum_{n=1}^{\infty} z^n/2^n$. We will determine for which values of z this infinite series converges absolutely.

By the root test, we know that $\sum_{n=1}^{\infty} |z^n/2^n|$ converges if $|z| < 2$ and diverges if $|z| > 2$. Also for $|z| = 2$ we can see that the terms of the series do not go to zero, so the series diverges.

Thus $\sum_{n=1}^{\infty} z^n/2^n$ converges absolutely for $|z| < 2$ and diverges for $|z| \geq 2$.

Now it was claimed at the start of this subsection that absolute convergence allows us to be a 'little less careful'. Allow me to elaborate on this.

In general, *you must be very careful when manipulating infinite series*. To see why, consider this informal example.

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \cdots = \log 2$$
$$1 + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} + \frac{1}{7} - \frac{1}{4} + \cdots = \frac{3}{2} \log 2$$

The two series above both converge, and they both have all of the same terms. The only difference is that the terms are in a different order, and this change has completely altered the value of the series.

What's nice about absolutely convergent series is that *we don't have to worry about this.*

Definition 2.22 (Rearrangement). A *rearrangement* of the series $\sum_{n=1}^{\infty} a_n$ is another series $\sum_{n=1}^{\infty} a_{\sigma(n)}$, where $\sigma : \mathbb{N} \to \mathbb{N}$ is a bijection.

Theorem 2.23 (Rearranging Absolutely Convergent Series). *If $\sum_{j=1}^{\infty} a_j$ is absolutely convergent then any rearrangement of the series will converge to the same value.*

Proof. Let $\sum_{j=1}^{\infty} a_{\sigma(j)}$ be a rearrangement of the series. Given $\epsilon > 0$, we wish to show that there exists an integer N such that $|\sum_{j=1}^{\infty} a_j - \sum_{j=1}^{n} a_{\sigma(j)}| < \epsilon$ for all $n \geq N$.

By Cauchy's criterion there exists an integer m such that $\sum_{j=m+1}^{\infty} |a_j| < \epsilon$. We then choose N such that $\{a_1, a_2, \ldots, a_m\} \subseteq \{a_{\sigma(1)}, a_{\sigma(2)}, \ldots, a_{\sigma(N)}\}$. This can be done by setting $N = \max_{1 \leq i \leq m} \sigma^{-1}(i)$. Then if $n \geq N$ we have

$$\sum_{j=1}^{\infty} a_j - \sum_{j=1}^{n} a_{\sigma(j)} = \sum_{j \in A_n} a_j,$$

where $A_n \subseteq \{m+1, m+2, \ldots\}$. Then applying the triangle inequality we have

$$\left| \sum_{j=1}^{\infty} a_j - \sum_{j=1}^{n} a_{\sigma(j)} \right| = \left| \sum_{j \in A_n} a_j \right| \leq \left| \sum_{j=m+1}^{\infty} a_j \right| \leq \sum_{j=m+1}^{\infty} |a_j| < \epsilon,$$

as required. $\square$

To see just how not-true this is for series that do not converge absolutely, we just need to read the statement of Riemann's Rearrangement Theorem.

Theorem 2.24 (Riemann's Rearrangement Theorem). *Let $a_j \in \mathbb{R}$. Suppose that $\sum_{j=1}^{\infty} a_j$ is convergent but not absolutely convergent. Then given any $x \in \mathbb{R}$ there is a rearrangement of this series such that $\sum_{j=1}^{\infty} a_{\sigma(j)} = x$.*

Proof. Left as an exercise. $\square$

3 Continuity

The next idea we will develop is that of 'continuity'.

3.1 Continuity of Functions

Continuity is all about the behaviour of a function around a point. For a function to be be continuous at some x, we need $f(y)$ to be close to $f(x)$ whenever y is sufficiently close to x.

Definition 3.1 (Continuity). Let $A \subseteq \mathbb{C}$ and $f : A \to \mathbb{C}$. We say that f is *continuous at* $a \in A$ if given any $\epsilon > 0$ we can find a $\delta > 0$ such that $|f(x) - f(y)| < \epsilon$ for all $y \in A$ such that $|x - y| < \delta$.

We say that f is *continuous* if it is continuous at every $a \in A$.

You may notice that this definition uses exactly the definition of the limit of a function mentioned in subsection 1.4. This immediately gives us two more equivalent definitions of continuity.

Proposition 3.2 (Limit Definition of Continuity). *Let $f : A \to \mathbb{C}$. Then f is continuous at $a \in A$ if and only if $\lim_{z \to a} f(z) = f(a)$.*

Proposition 3.3 (Sequence Definition of Continuity). *Let $f : A \to \mathbb{C}$. Then f is continuous at $a \in A$ if and only if for every sequence $a_n \in A$ with $a_n \to a$ we have $f(a_n) \to f(a)$.*

Each of the definitions have various advantages or disadvantages. Still it is worth noting that the sequential/limit continuity definitions can be used to quite easily show certain properties of continuity that would otherwise be quite fiddly with the ϵ-δ definition. Using sequential continuity is also a natural way to show that a function *is not* continuous at a point. Let's take a look at some examples.

Example 3.4 (The Constant Function is Continuous). If $f(x) = c$ then f is continuous. To see this, we can take any value of δ in the definition.

Example 3.5 ($f(x) = x$ is Continuous). If $f(x) = x$ then f is continuous. To see this, we can take $\delta = \epsilon$ in the definition.

Example 3.6 (Step Functions are Discontinuous). Consider the function $f : \mathbb{R} \to \mathbb{R}$ with

$$f(x) = \begin{cases} -1 & \text{if } x < 1, \\ 1 & \text{if } x \geq 1. \end{cases}$$
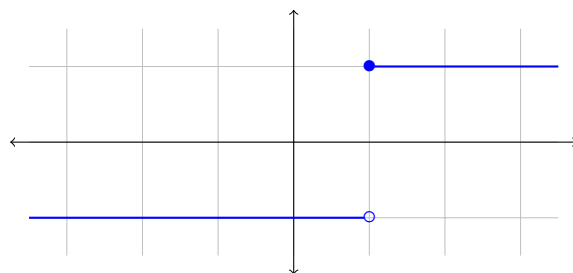
This function is not continuous at 1. To see this, note that $1 - 1/n \to 1$ but $f(1 - 1/n) \to -1 \neq f(1)$ as $n \to \infty$.

Example 3.7 (The Domain Matters). Consider the function $f : \mathbb{Q} \to \mathbb{R}$ with

$$
\begin{cases}
1 & \text{if } x^2 > 2, \\
0 & \text{if } x^2 < 2.
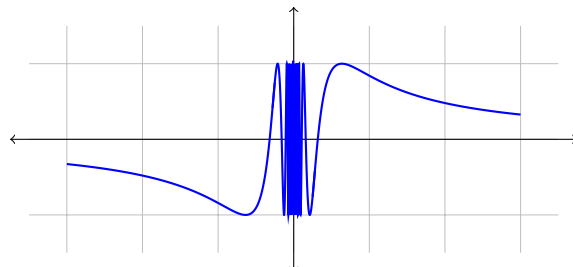\end{cases}
$$

This function is continuous, since for every $a \in \mathbb{Q}$ there is an interval about a for which f is constant, so f is continuous at a. If f was defined on $\mathbb{R}$ instead of $\mathbb{Q}$, then it would be discontinuous only at $\pm\sqrt{2}$. But these points are not in our domain, so we don't need to worry about them.

Example 3.8 (Continuity of $\sin 1/x$). Consider the function[5] $f : \mathbb{R} \to \mathbb{R}$ with $f(x) = \sin(1/x)$ if $x \neq 0$ and $f(0) = 0$.

This function is not continuous at $x = 0$. To see this, consider the sequence $a_n = \frac{1}{2\pi n + \pi/2}$. Then $a_n \to 0$, but $f(a_n) = 1 \to 1 \neq f(0)$ as $n \to \infty$.

Example 3.9 (A Functional Equation with Continuity). We wish to find all continuous functions $f : \mathbb{R} \to \mathbb{R}$ such that $f(x) + f(2x) = 0$.

Letting $x = 0$ we get $f(0) = 0$. Then rearranging we have $f(x) = -f(x/2)$, and repeatedly using this gives

$$
f(x) = -f\left(\frac{x}{2}\right) = f\left(\frac{x}{4}\right) = -f\left(\frac{x}{8}\right) = \cdots = (-1)^n f\left(\frac{x}{2^n}\right).
$$

Then since f is continuous we have $f(x) = \lim_{n \to \infty} (-1)^n f(x/2^n) = f(0) = 0$. Thus $f(x) = 0$ is the only such function, which clearly works.

---

[5]I know that we haven't defined what sin is, but we will use it in examples because they are instructive. If this bothers you, feel free to come back after our discussion of power series.

When attempting to determine if a function is continuous, one should keep in mind the following properties of continuity (all of which follow directly from our basic properties of limits).

Proposition 3.10 (Sums, Products and Reciporicals of Continuous Functions). *Let $f, g : A \to \mathbb{C}$ be functions. Then the following hold.*

(i) *If f and g are continuous at a, then so is their sum $f + g$.*

(ii) *If f and g are continuous at a, then so is their product fg.*

(iii) *If f is continuous at a and $f(a) \neq 0$ then so is $1/f$.*

Proof. We prove each individually using the limit definition of continuity.

(i) If $\lim_{z \to a} f(z) = f(a)$ and $\lim_{z \to a} g(z) = g(a)$ then $\lim_{z \to a} (f+g)(z) = (f+g)(a)$.

(ii) If $\lim_{z \to a} f(z) = f(a)$ and $\lim_{z \to a} g(z) = g(a)$ then $\lim_{z \to a} (fg)(z) = (fg)(a)$.

(iii) If $\lim_{z \to a} f(z) = f(a)$ and $f(a) \neq 0$, then $\lim_{z \to a} 1/f(z) = 1/f(a)$. $\qquad\square$

Along with the fact that $f(x) = x$ and $f(x) = c$ are continuous, these properties imply that all polynomials are continuous.

Proposition 3.11 (The Composition of Continuous Functions is Continuous). *Let $f : A \to B$ and $g : B \to \mathbb{C}$ with $A, B \subseteq \mathbb{C}$. Then if f is continuous at $a \in A$ and g is continuous at $f(a)$, then $g \circ f$ is also continuous at a.*

Proof. If $a_n \to a$ then by the continuity of f we have $f(a_n) \to f(a)$, and by the continuity of g we have $g(f(a_n)) \to g(f(a))$. So $g \circ f$ is continuous at a. $\qquad\square$

3.2 The Intermediate Value Theorem

Continuous functions are nice because they have a few nice properties. The first of these that we will prove is the *intermediate value theorem*, a central theorem in elementary analysis. The idea of the intermediate value theorem is that if a continuous function starts at one value and ends at a different value, then it must take on all of the values in between.
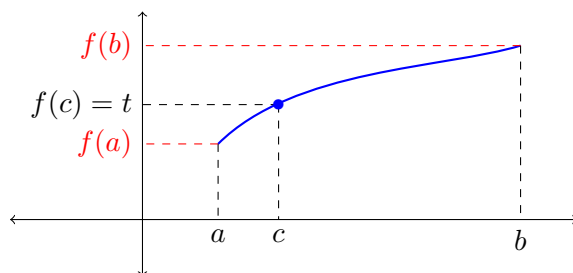
In the proof below we will employ suprema[6], but we will give another proof afterwards that does not use this (at the expense of being slightly longer).

Theorem 3.12 (The Intermediate Value Theorem)**.** *Let $f : [a, b] \to \mathbb{R}$ be a continuous function. Then for every t between $f(a)$ and $f(b)$ there is an $c \in [a, b]$ with $f(c) = t$.*

Proof. Suppose without loss of generality[7] that $f(a) < t < f(b)$. Consider the set $S = \{x \in [a, b] : f(x) < t\}$. This set is bounded and since $a \in S$ it is non-empty. So we can let $c = \sup S$, and note that $a \leq c \leq b$.

Let $n \geq 1$ be an integer. Then since c is the supremum there must exist some $x_n \in S$ such that $c - \frac{1}{n} < x_n \leq c$. By the squeeze theorem we have $x_n \to c$ as $n \to \infty$. Also f is continuous so $f(x_n) \to f(c)$. We also constructed x_n to be in S, giving $f(x_n) < t$ for all n. This implies that $f(c) \leq t$.

We know that $c \in [a, b]$ but $c \neq b$, so there is some integer N such that for $n \geq N$ we have $c + 1/n \leq b$. Using that c is the supremum, we have $c + 1/n \notin S$ for all $n \geq N$, that is, $f(c + 1/n) \geq t$ for all $n \geq N$. Then by continuity we have $f(c) \geq t$.

But then $f(c) \geq t$ and $f(c) \leq t$, so we must have $f(c) = t$. $\qquad\square$

Before we look at some examples of using the intermediate value theorem, there's a few things worth noting about the intermediate value theorem.

First of all, we say absolutely nothing about uniqueness in the theorem. It is very much possible for a function to take on an intermediate value multiple times. Second of all, when applying the intermediate value theorem (commonly abbreviated to IVT) a good general problem solving technique is trying to apply IVT to other related functions such as $g(x) = f(x) - x$ or $g(x) = f(x) - t$ and things like that. You will see this 'trick' show up in the examples below, and also again when we get to Rolle's theorem and the mean value theorem.

Example 3.13 (Fixed Points of Continuous Maps from $[0, 1]$ to Itself)**.** We will prove that if $f : [0, 1] \to [0, 1]$ is a continuous function then there is some $c \in [0, 1]$ such that $f(c) = c$.

Consider the function $g(x) = f(x) - x$. Then g is continuous with $g(0) = f(0)$ and $g(1) = f(1) - 1$. Since $0 \leq f(x) \leq 1$, we must have $g(0) \geq 0$ and $g(1) \leq 0$. Thus by the intermediate value theorem there exists some $c \in [0, 1]$ such that $g(c) = 0$. But then we have $f(c) - c = 0$, that is, $f(c) = c$ as required.

---

[6]If you are unfamiliar with this term and/or the least upper bound principle, feel free to have a look at Chapter 2 of my 'Numbers and Sets' course notes.

[7]If $t = f(a)$ or $t = f(b)$ then we are done. Also the case $f(a) > t > f(b)$ follows similarity.

Example 3.14 (Existence nth Roots)**.** We will show that for any positive integer n and $y > 0$ there exists a real number x such that $x^n = y$.

Consider the function $f(x) = x^n$. This function is continuous on $[0, 1+y]$, and since $f(0) = 0$ and $f(1+y) = (1+y)^n$, we have $f(0) < y < f(1+y)$. Thus by the intermediate value theorem there is some $x \in [0, 1+y]$ such that $f(x) = y$, that is, $x^N = y$.

Now before we look on let's have a look at an alternative proof for the intermediate value theorem. This proof uses the same idea as the previous proof (try to see why) but avoids the use of suprema.

Aside: Hunting For ~~Lions~~ Intermediate Values

It was mentioned before that the 'lion hunting' technique that we used to prove the Bolzano–Weierstrass theorem could also be used to prove other theorems in Analysis. One such example is the intermediate value theorem.

The steps in the proof are quite similar to that of the proof of the Bolzano-Weierstrass theorem, the only difference is what we are 'hunting for', which in this case is our intermediate value, and how we finish off the proof.

Proof (Intermediate Value Theorem, Lion Hunting Style). Again suppose without loss of generality that $f(a) < t < f(b)$. We are going to define two sequences a_n and b_n inductively as follows. Begin by setting $a_1 = a$ and $b_1 = b$, and let $c = (a_1 + b_1)/2$ be the midpoint a_1 and b_1.

Then there is two possibilities:

1. $f(c_1) \geq t$.

2. $f(c_1) < t$.

If the first one holds, we set $a_2 = a_1$ and $b_2 = c_1$, and if it doesn't then we set $a_2 = c_1$ and $b_2 = b_1$. Repeating this process, we construct a_n and b_n such that $f(a_n) \leq t \leq f(b_n)$, $a_{n-1} \leq a_n \leq b_n \leq b_{n-1}$, and also $b_n - a_n = (b_{n-1} - a_{n-1})/2$.

Now a_n is increasing and bounded above, and b_n is decreasing and bounded below and thus $a_n \to c \in [a, b]$, and $b_n \to c' \in [a, b]$. Then we have $c - c' = (c - c')/2$ using the above result, and thus $c = c'$.

Since f is continuous at c and $a_n \to c$, we know that $f(a_n) \to c$ as $n \to \infty$. Since $f(a_n) \leq t$, we also know that $f(c) \leq t$. Similarity we also have $f(b_n) \geq t$, so $f(c) \geq t$. But then $t \leq f(c) \leq t$, and thus $f(c) = t$, and we are done. $\square$

To get a feel for the construction, try drawing a rough sketch of the process. If you consider the case of a function that takes on the desired

intermediate value multiple times, you should also see the that this proof does not give the same value for c as the previous proof.

A natural question upon learning of the intermediate value theorem is if it is equivalent to definition of continuity. The answer to this question is no – there are functions that have the 'intermediate value property', without being continuous. One example is $\sin(1/x)$ which we showed is discontinuous at $x = 0$ in Theorem 3.8. Another more elaborate example is 'Conway's Base-13 Function', which has the intermediate value property but is discontinuous everywhere. You can read more about this function on Wikipedia.

3.3 The Boundedness Theorem

The last result about continuous functions that we will prove is the *boundedness theorem*, another central theorem in elementary analysis. The statement of this result is straightforward: a continuous function on a closed bounded interval is bounded and attains its bounds.

First we will prove that a continuous function defined on some closed interval is indeed bounded.

Lemma 3.15 (Continuous Function on Closed Interval is Bounded)**.** *If $f : [a,b] \to \mathbb{R}$ is continuous then there exists K such that $|f(x)| \leq K$ for all $x \in [a,b]$.*

Proof. If f was unbounded, then given any integer $n \geq 1$ there exists $x_n \in [a,b]$ such that $|f(x_n)| > n$. By Bolzano–Weierstrass, x_n has a convergent subsequence $x_{n_j} \to x$. Also since $a \leq x \leq b$, we must have $x \in [a,b]$. By continuity of f, $f(x_{n_j}) \to f(x)$, but $|f(x_{n_j})| > n_j$, and thus does not tend to a limit. Thus we have a contradiction. $\qquad\square$

Now we can prove that such a continuous function attains its bounds, giving us the boundedness theorem.

Theorem 3.16 (Boundedness Theorem)**.** *If $f : [a,b] \to \mathbb{R}$ is continuous then there exists $x_1, x_2 \in [a,b]$ such that $f(x_1) \leq f(x) \leq f(x_2)$ for all $x \in [a,b]$.*

Proof. By our previous lemma, we know that f is bounded. Now we define the set

$$A = \{f(x) : x \in [a,b]\}.$$

This set is bounded and non-empty and thus has a supremum $M = \sup A$. Then for each positive integer n, $M - 1/n$ cannot be an upper bound for A. This implies that there is some $x_n \in [a,b]$ such that $M - 1/n < f(x_n) \leq M$.

By Bolzano–Weierstrass, x_n has a convergent subsequence $x_{n_j} \to x$. Since $a \leq x_n \leq b$, we know $a \leq x \leq b$. By the continuity of f, $f(x_{n_j}) \to$

$f(x)$, but $f(x_{n_j}) \to M$ by construction. So $f(x) = M$. So $x_2 = x$. The minimum follows analogously. $\qquad \square$

Alternate Proof. As before, let $M = \sup A$ and suppose that there was no x_2 such that $f(x_2) = M$. Then let $g(x) = \frac{1}{M - f(x)}$ for $x \in [a,b]$. This function is well defined and continuous. Now g must be bounded by the previous lemma, so $g(x) \le K$ for all $x \in [a,b]$, for some K. This means that $f(x) \le M - 1/k$ on $[a,b]$. This is contradiction since we set M as the supremum. $\qquad \square$

4 Differentiability

You are likely familiar with differentiability (and particularly the computation of derivatives) from calculus. While this knowledge should certainly not be disregarded, we are going to go from the beginning, doing everything with a little more care than it got in calculus. Throughout this section we will deal with functions $f : A \subseteq \mathbb{R} \to \mathbb{R}$, though the basic definitions will apply to $\mathbb{C}$.[8].

Here's our basic definition:

Definition 4.1 (Differentiability). Let $A \subseteq \mathbb{R}$ and $f : A \to \mathbb{R}$. Then if x is a limit point of A, we say that f is *differentiable at x* with derivative $f'(x)$ if

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

We say that f is *differentiable* on A if f is differentiable at every $x \in A$.

Let's pause for a moment. The core idea of differentiation is we want to *approximate a function around some point using a linear map*[9]. Indeed, our definition of the derivative is directly equivalent to the following: f is differentiable at x if

$$f(x + h) = f(x) + f'(x)h + \varepsilon(h),$$

where $\lim_{h \to 0} \varepsilon(h)/h = 0$. That is, it is differentiable if we can approximate the function with some linear map where the error decreases faster than linearly.

4.1 Properties of Differentiable Functions

With that spiel over, we can move on to some properties of differentiable functions. The first is that differentiability implies continuity.

Proposition 4.2 (Differentiability Implies Continuity). *If f is differentiable at x, then f is continuous at x.*

Proof. Using the limit definition of continuity we have

$$\lim_{h \to 0} [f(x + h) - f(x)] = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h} \cdot (h) = f'(x) \cdot 0 = 0,$$

thus $\lim_{h \to 0} f(x + h) = f(x)$, so f is continuous at x. $\qquad \square$

---

[8]We will use differentiation over $\mathbb{C}$ in our discussion of power series though. One should also note that the other basic rules such as the sum, product and chain rules will also hold over $\mathbb{C}$, but being differentiable over $\mathbb{C}$ is a stronger condition than being differentiable over $\mathbb{R}$ or $\mathbb{R}^2$. For example, consider $f(z) = \overline{z}$, which is not complex differentiable.

[9]Indeed it is which this idea, not really that of 'the tangent to a curve' that is used to generalise the derivative to functions of multiple variables. Of course they are closely related.

Now we can prove some basic rules for computing derivatives.

Proposition 4.3 (Sum, Produt & Quotient Rules)**.** *Suppose that* $f : \mathbb{R} \to \mathbb{R}$ *and* $g : \mathbb{R} \to \mathbb{R}$ *are differentiable at* x *with derivatives* $f'(x)$ *and* $g'(x)$*. Then the following hold.*

(i) $f + g$ *is differentiable at* x*, with* $(f + g)'(x) = f'(x) + g'(x)$*,*

(ii) fg *is differentiable at* x*, with* $(fg)'(x) = f'(x)g(x) + f(x)g'(x)$*,*

(iii) *If* $g'(x) \neq 0$ *then* f/g *is differentiable at* x*, with* $(f/g)'(x) = \frac{g(x)f'(x) - g'(x)f(x)}{g^2(x)}$*.*

Proof. We prove each individually.

(i) This follows from the sum of limits.

(ii) We have by the continuity of f

$$\lim_{h \to 0} \frac{f(x + h)g(x + h) - f(x)g(x)}{h}$$
$$= \lim_{h \to 0} \left[f(x + h)\frac{g(x + h) - g(x)}{h} + g(x)\frac{f(x + h) - f(x)}{h} \right],$$

and by the continuity of f and g at x we have $(fg)'(x) = f(x)g'(x) + f'(x)g(x)$.

(iii) Similarly we have

$$\lim_{h \to 0} \frac{f(x + h)/g(x + h) - f(x)/g(x)}{h}$$
$$= \lim_{h \to 0} \frac{1}{g(x)g(x + h)} \left[g(x)\frac{f(x + h) - f(x)}{h} - f(x)\frac{g(x + h) - g(x)}{h} \right],$$

and by the continuity of f and g we have $(f/g)'(x) = \frac{g(x)f'(x) - g'(x)f(x)}{g^2(x)}$. $\qquad \square$

We will now prove the chain rule, which tells us how to compute the derivative of the composition of functions. Unfortunately this proof is quite 'tricky', and trying to do something like we did in the proof above will not work. Instead, we need to return to our equivalent definition of differentiability, with $f(x + h) = f(x) + f'(x)h + \varepsilon(h)$, where $\lim_{h \to 0} \varepsilon(h)/h = 0$.

Proposition 4.4 (Chain Rule)**.** *Suppose* $f : \mathbb{R} \to \mathbb{R}$ *is differentiable at* x *and* $g : \mathbb{R} \to \mathbb{R}$ *is differentiability at* $f(x)$*. Then* $g \circ f$ *is differentiable at* x *with* $(g \circ f)'(x) = g'(f(x)) \cdot f'(x)$*.*

Proof. We have

$$f(x + h) = f(x) + f'(x)h + \varepsilon_f(h)$$
$$g(f(x) + h) = g(f(x)) + g'(f(x))h + \varepsilon_g(h),$$

where $\varepsilon_f(h)/h, \varepsilon_g(h)/h \to 0$ as $h \to 0$. From this we obtain

$$g(f(x + h)) = g(f(x) + f'(x)h + \varepsilon_f(h))$$
$$= g(f(x)) + g'(f(x))(f'(x)h + \varepsilon_f(h)) + \varepsilon_g(f'(x)h + \varepsilon_f(h)).$$

Rearranging we get

$$\lim_{h \to 0} \frac{g(f(x + h)) - g(f(x))}{h}$$
$$= \lim_{h \to 0} \frac{g'(f(x))(f'(x)h + \varepsilon_f(h)) + \varepsilon_g(f'(x)h + \varepsilon_f(h))}{h}$$
$$= g'(f(x))f'(x) + \lim_{h \to 0} g'(f(x))\frac{\varepsilon_f(h)}{h} + \lim_{h \to 0} \frac{\varepsilon_g(f'(x)h + \varepsilon_f(h))}{h}.$$

Then as $h \to 0$ we have $\varepsilon_f(h)/h \to 0$ and $f'(x)h + \varepsilon_f(h) \to 0$ so $\varepsilon_g(f'(x)h + \varepsilon_f(h))/h \to 0$, and thus we have $(g \circ f)'(x) = g'(f(x))f'(x)$. $\square$

Example 4.5 (Using the Chain Rule). Let $f(x) = \sin(x^2)$. Then since $(\sin x)' = \cos x$ we have $f'(x) = 2x\cos(x^2)$.

Example 4.6 (A Continuous Everywhere But Non-Differentiable Function). Consider the function $f(x) = x\sin(1/x)$ when $x \neq 0$, and $f(0) = 0$. This function is continuous everywhere[10], but we will show that it is not differentiable at $x = 0$.

At $x \neq 0$ this is differentiable by the chain rule, but at $x = 0$ we would need the following limit to exist:

$$\lim_{h \to 0} \frac{(0 + h)\sin\left(\frac{1}{x+h}\right) - 0\sin\left(\frac{1}{x}\right)}{h} = \lim_{h \to 0} \sin\left(\frac{1}{h}\right),$$

but this limit does not exist (a similar construction to Theorem 3.8 will show this), and thus f is not differentiable at $x = 0$.

Now in the example above we have a continuous everywhere function that's not differentiable at 0. A related example is one of a function that is differentiable everywhere (and hence continuous), but whose derivative is discontinuous.

---

[10]Feel free to check this!

Example 4.7 (A Function with Discontinuous Derivative)**.** Consider the function $f(x) = x^2 \sin(1/x)$ when $x \neq 0$ and $f(0) = 0$.
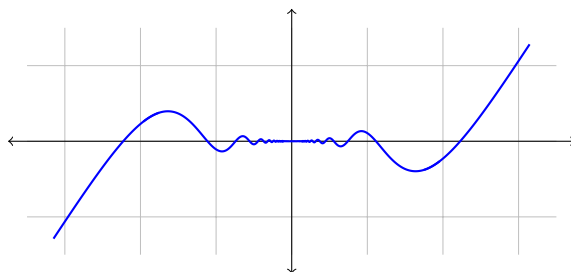
For $x \neq 0$ we have that f is differentiable by the chain rule, and the derivative is given by

$$f'(x) = x \sin\left(\frac{1}{x}\right) + \cos\left(\frac{1}{x}\right).$$

For $x = 0$, we compute the derivative directly. We have

$$f'(0) = \lim_{h \to 0} \frac{f(0+h) - f(0)}{h} = \lim_{h \to 0} \frac{h^2 \sin\left(\frac{1}{h}\right)}{h} = 0.$$

Thus we can see that f' is not continuous, since $f'(x) \nrightarrow 0$ as $x \to 0$ (as $\cos 1/x$ is not continuous).

There is however a limit as to how discontinuous a derivative can be. In particular the derivative of a differentiable function must have the intermediate value property.

Theorem 4.8 (Darboux's Theorem)**.** *If $f : \mathbb{R} \to \mathbb{R}$ is differentiable then f' has the intermediate value property. That is to say, if $a < b$ and $f'(a) < z < f'(b)$ then there exists $c \in (a, b)$ with $f'(c) = z$.*

Proof. Given $a < b$ and $z \in \mathbb{R}$ such that $f'(a) < z < f'(b)$, we wish to show there exists $c \in (a, b)$ such that $f(c) = z$. We can rewrite the condition as $f'(a) - z < 0 < f'(b) - z$. Now define $g(x) = f(x) - zx$, and note that we have $g'(a) < 0 < g'(b)$.

We want to find $c \in (a, b)$ such that $g'(c) = 0$. Now, g is continuous (since f is continuous), and thus is bounded on $[a, b]$, and the minimum is attained say at some point k. We can't have $k = a$, as that would imply that $g'(a) \geq 0$, and we also can't have $k = b$ as that would imply $g'(b) \leq 0$. Thus $k \in (a, b)$, and we must then have $g'(k) = 0$, and we are done. $\qquad \square$

4.2   Rolle's Theorem & The Mean Value Theorem

The sign of a function's derivative at a point can tell us quite a bit about the behaviour of that function near that point. In particular, knowing that the derivative at a point vanishes can be quite useful;.

Definition 4.9 (Local Maxima & Minima). Let $f : \mathbb{R} \to \mathbb{R}$. If there exists δ such that $|x - x_0| < \delta$ implies that $f(x_0) \leq f(x)$, then we call x a *local maximum*. Similarly if this implies that $f(x_0) \geq f(x)$, we call x a *local minimum*.

Our important result is as follows:

Lemma 4.10 (Derivative of Maxima). *Let $f : \mathbb{R} \to \mathbb{R}$. If x is a local maximum or minimum of f and f is differentiable at x then $f'(x) = 0$.*

Proof. Without loss of generality, assume that x is a local maximum. Then there exists δ such that $|h| < \delta_1$ implies that $f(x + h) \leq f(x)$.

Then if $0 < h < \delta$ we have

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

and thus we have $f'(x) \geq 0$.

Similarly if $-\delta < h < 0$ we have

$$\frac{f(x + h) - f(x)}{h} \leq 0,$$

and thus we have $f'(x) \leq 0$. Hence $f'(x) = 0$. $\square$

When we combine this result with the boundedness theorem from our discussion about continuity, we end up with *Rolle's theorem*.

Theorem 4.11 (Rolle's Theorem). *Let $f : [a, b] \to \mathbb{R}$ be a continuous function which is differentiable on (a, b). Then if $f(a) = f(b)$ then there exists $c \in (a, b)$ such that $f'(c) = 0$.*

Proof. Since f is differentiable, it is also continuous and thus by the boundedness theorem f is bounded on $[a, b]$ and these bounds are achieved. Thus we can let $M = \max_{x \in [a,b]} f(x)$ and $m = \min_{x \in [a,b]} f(x)$.

If $M = m$, then f must be constant and $f'(x) = 0$ for all $x \in (a, b)$, so we are done. Otherwise $M > f(a)$ or $m < f(a)$. If $M > f(a)$ then since the bounds are attained there exists $c \in (a, b)$ such that $f(c) = M$. But them M is a maximum, so by our previous lemma we have $f'(c) = 0$. A similar argument works if $m < f(a)$. $\square$

A direct consequence of Rolle's theorem[11] is another classic theorem of analysis, the *mean value theorem*, which is frequently abbreviated to MVT.

Theorem 4.12 (The Mean Value Theorem). *Let $f : [a, b] \to \mathbb{R}$ be a continuous function which is differentiable on (a, b). Then there exists $c \in (a, b)$ such that $f(b) - f(a) = f'(c)(b - a)$.*

---

[11]It's possible to establish this theorem without Rolle's theorem, and then Rolle's theorem pops out as a special case. The proof is (more or less) just what we did for Rolle's theorem laid out explicitly in the proof of this theorem.

Proof. Let $k = \frac{f(a)-f(b)}{a-b}$, and define $\phi(x) = f(x) - kx$. Note that $\phi(a) = \phi(b)$. Then by Rolle's theorem there exists $c \in (a, b)$ such that $\phi'(c) = 0$, that is $f'(c) = k$. $\qquad\square$

The mean value theorem says something notable: the size of the derivative controls the size of the function, or (in rougher terms) it puts a restriction on know 'badly behaved' the function can be. Also, if we appeal to geometric intuition (as we have tried not to, it's an easy way to go wrong) we can see that the mean value theorem says, as R. P. Burn wrote in *Numbers and functions*, "for the graph of a differentiable function, there is always a tangent parallel to the chord". Of course, we will quickly move on from geometrical thinking.[12]

Now let's think about applying the mean value theorem to two different functions. Suppose that $f, g : [a, b] \to \mathbb{R}$ is continuous and differentiable on (a, b), and $g(a) \neq g(b)$. Then the mean value theorem gives us $s, t \in (a, b)$ such that
$$\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{(b-a)f'(s)}{(b-a)g'(t)} = \frac{f'(s)}{g'(t)}.$$
A stronger version of the mean value theorem says that we can take $s = t$.

Theorem 4.13 (Cauchy's Mean Value Theorem). *Let $f, g : [a, b] \to \mathbb{R}$ be continuous functions which are differentiable on (a, b). Then there exists $c \in (a, b)$ such that*
$$(f(b) - f(a))g'(c) = f'(c)(g(b) - g(a)).$$

Proof. We define the function
$$\phi(x) = \begin{vmatrix} 1 & 1 & 1 \\ f(a) & f(x) & f(b) \\ g(a) & g(x) & g(b) \end{vmatrix}$$
$$= [f(x)g(b) - f(b)g(x)] - [f(a)g(b) - f(b)g(a)] + [f(a)g(x) - f(x)g(a)].$$

Then ϕ is continuous on $[a, b]$ and differentiable on (a, b)[13]. Also $\phi(a) = \phi(b) = 0$. Then by Rolle's theorem, there exists $c \in (a, b)$ such that $\phi'(c) = 0$.

Differentiating ϕ we then have
$$\phi'(x) = f'(x)[g(b) - g(a)] + g'(x)[f(a) - f(b)],$$
and thus $\phi'(c) = 0$ gives the desired result. $\qquad\square$

---

[12]I guess here is a good place to mention how we can reason geometrically in analysis. Drawing pictures and thinking geometrically is a great way to understand how things work, to come up with counterexamples and much more. Still it is important to remember that *basically nothing in this course is provable by appealing to geometric intuition.* Instead this type of thinking should just inform the 'analysis' side of us how to approach things.

[13]If the use of determinants bothers you, then feel free to just look at the expansion. If also remember how integrals work, you can obtain ϕ from working backwards from what we need to arrive at.

Cauchy's mean value theorem has many applications. For example, we can use the mean value theorem to establish L'Hôpital's rule for evaluating limits.

Example 4.14 (L'Hôpitel's Rule). We wish to evaluate $\lim_{x \to 0} \dfrac{e^x - 1}{\sin x}$.

By Cauchy's mean value theorem, there exist $c \in (0, x)$ such that $\frac{e^x - e^0}{\sin x - \sin 0} = \frac{e^c}{\cos c}$. Then $c \to 0$ as $x \to 0$, and thus we get

$$\lim_{x \to 0} \frac{e^x - 1}{\sin x} = \lim_{c \to 0} \frac{e^c}{\cos c} = 1.$$

In more generality:

Theorem 4.15 (L'Hôpital's Rule). *Suppose $f, g : [a, b] \to \mathbb{R}$ are continuous and differentiable on (a, b). Suppose that $f(a) = g(a) = 0$, that $g'(x)$ does not vanish near a and $f'(x)/g'(x) \to \ell$ as $x \to a$. Then $f(x)/g(x) \to \ell$ as $x \to a$.*

Proof. Since $g'(x)$ does not vanish near a, we can suppose that $g'(x) \neq 0$ for $x \in (a, b)$, as otherwise we could just consider the subinterval (a, b'), defined so that this is the case.

By Cauchy's mean value theorem we have

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

where $c \in (a, x)$. Then $c \to a$ as $x \to a$, and hence $f(x)/g(x) \to \ell$ as $x \to a$. □

The mean value theorem can also help us extend Theorem 4.10. Knowing the sign of a derivative over some interval can tell us immediately if it is constant, increasing or decreasing.

Proposition 4.16 (Sign of the Derivative). *Let $f : [a, b] \to \mathbb{R}$ be differentiable on (a, b). Then the following hold.*

(i) *If $f'(x) \geq 0$ for all $x \in (a, b)$, then f is monotonically increasing.*

(ii) *If $f'(x) = 0$ for all $x \in (a, b)$, then f is constant.*

(iii) *If $f'(x) \leq 0$ for all $x \in (a, b)$, then f is monotonically decreasing.*

Proof. By the mean value theorem we have $f(x_2) - f(x_1) = (x_2 - x_1)f'(x)$ where $a < x_1 < x < x_2 < b$. Then all of the results follow by considering the sign of $f'(x)$. □

4.3 Inverses of Functions

Before we move to slightly more exciting material, there's a few things about inverses of functions that we need to get out of the way. The proofs below are straightforward but require some care (by which I mean you have to avoid messing up which letter means what).

First of all, if we know that a continuous function is strictly increasing[14] then we can deduce that the function has an inverse that is continuous and strictly increasing also. A similar result holds for strictly decreasing functions.

Theorem 4.17 (Continuous Inverse Theorem). *Let $f : [a,b] \to \mathbb{R}$ be a continuous and strictly increasing function. Then if $c = f(a)$ and $d = f(b)$, $f : [a,b] \to [c,d]$ is bijective and the inverse $f^{-1} : [c,d] \to [a,b]$ is continuous and strictly increasing.*

Proof. As f is strictly increasing, it is an injection. It is also a surjection since if $c < t < d$ then there exists $x \in (a,b)$ such that $f(x) = t$. Thus f is a bijection, and has an inverse f^{-1}.

Since f is strictly increasing, we can write this as $x_1 < x_2 \iff f(x_1) < f(x_2)$, so letting $f(x_1) = y_1$ and $f(x_2) = y_2$ we have $f^{-1}(y_1) < f^{-1}(y_2) \iff y_1 < y_2$. Thus f^{-1} is also strictly increasing.

Now we will show f^{-1} is continuous at some $y \in [c,d]$. Given $\epsilon > 0$, let $x = f^{-1}(y)$. If $x \neq a, b$, we can find δ such that

$$f(x - \epsilon) \leq f(x) - \delta < f(x) < f(x) + \delta \leq f(x + \epsilon).$$

But then $|z - y| < \delta$ implies that $x - \epsilon < f^{-1}(z) < x + \epsilon$, so $|f^{-1}(z) - f^{-1}(y)| < \epsilon$. Thus f^{-1} is continuous on (a,b)

Otherwise, if $x = a$ we have $f(x) < f(x+\epsilon)$. Then $|z - y| < f(x+\epsilon) - f(x)$ implies that $|f^{-1}(z) - f^{-1}(y)| < \epsilon$, so f^{-1} is continuous at c. A similar argument shows that it is continuous at d. $\square$

Now this *is* the differentiability section, so we can also describe what the derivative of a differentiable function's inverse is. This result is known as the one variable *inverse function theorem*.

Theorem 4.18 (Inverse Function Theorem). *Let $f : [a,b] \to \mathbb{R}$ be continuous, strictly increasing, and differentiable on (a,b). Then let $f(a) = c$ and $f(b) = d$. Then $f : [a,b] \to [c,d]$ is a bijection, and f^{-1} is differentiable on (c,d), with*

$$(f^{-1})'(x) = \frac{1}{f'(f^{-1}(x))}.$$

---

[14]If a continuous function was not strictly increasing or strictly decreasing then we couldn't have a unique inverse.

Proof. By our previous theorem, f is a bijection and f^{-1} is a continuous and strictly increasing function. Let $y = f(x)$, with $x \in (a,b)$. We wish to show that $g'(y) = 1/f'(x)$.

Given $k \neq 0$, let h be given by $y+k = f(x+h)$, that is, $f^{-1}(y+k) = x+h$, where $h \neq 0$. Then

$$\frac{f^{-1}(y+k) - f^{-1}(y)}{k} = \frac{h}{f(x+h) - f(x)}.$$

Then since $f'(x) > 0$ and f^{-1} is continuous (so $h \to 0$ as $k \to 0$) we have $k \to 0$

$$\lim_{k \to 0} \frac{f^{-1}(y+k) - f^{-1}(y)}{k} = \lim_{h \to 0} \frac{x+h-x}{f(x+h) - f(x)} = \frac{1}{f'(x)},$$

as required. $\qquad\square$

4.4 Taylor's Theorem

We are now going to think about functions where we can take derivatives a number of times (getting higher-order derivatives). We will begin by looking at how we can get an analogue of Rolle's theorem for higher-order derivatives. I will apologise in advance that this discussion is a bit long, so feel free to just skip to Taylor's theorem.

The following generalisation will come directly from Rolle's theorem.

Theorem 4.19 (Higher-Order Rolle's Theorem)**.** *Let f be continuous and $n-1$-times differentiable on $[a,b]$ and also n-times differentiable on (a,b). Then if $f(a) = f'(a) = f''(a) = \cdots = f^{(n-1)}(a) = 0$ and $f(b) = 0$ then there is a $c \in (a,b)$ such that $f^{(n)}(c) = 0$.*

Proof. Since $f(a) = f(b)$, by Rolle's theorem there exists $c_1 \in (a,b)$ such that $f'(c_1) = 0$. Then since $f'(a) = f'(c_1) = 0$, we can use Rolle's theorem again to obtain $c_2 \in (a, c_1)$ such that $f''(c_2) = 0$. Continuing on like this we obtain a $c_n \in (a,b)$ such that $f^{(n)}(c_n) = 0$, as required. $\qquad\square$

We can also attempt to get some sort of 'higher-order' mean value theorem. Recall that our proof of the mean value theorem was more or less as follows:

> *Define the function $\phi(x) = f(x) - kx$, where we choose k such that the conditions of Rolle's theorem are satisfied. Then apply Rolle's theorem to ϕ.*

We can try and do the same with a more elaborate construction. We will do this in three steps.

1. Construct a polynomial $P(x)$ such that $P(a) = f(a)$, $P'(a) = f'(a)$, and so on until $P^{(n-1)}(a) = f^{(n-1)}(a)$.

2. Construct another polynomial[15] $E(x)$ such that $E^{(r)}(a) = 0$ for $r = 0, 1, \ldots, n-1$, but with $E(b) = f(b) - P(b)$.

3. Take $\phi(x) = f(x) - P(x) - E(x)$.

If we construct $\phi(x)$ in this way, then we will have $\phi(a) = \phi'(a) = \cdots = \phi^{(n-1)}(a) = 0$, and also $\phi(b) = 0$, so it will satisfy our higher-order Rolle's theorem.

Aside: Constructing Our Polynomials

While it is likely you have seen a construction for P before, I'd like to take a minute to spell it out in detail because it has some nice ideas.

To construct this polynomial, what we really need is some construction of a term which affects the value of the polynomial at a point a if and only if we are considering a certain derivative. We can do this as follows:

$$Q_k(x) = \frac{(x-a)^k}{k!}.$$

We can then (by taking derivatives) find that the value of this term is

$$Q_k^{(j)}(a) = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{if } j \neq k \end{cases}$$

With this we can immediately write down an explicit construction for $P(x)$.

$$P(x) = \sum_{k=0}^{n-1} f^{(k)}(a) Q_k(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(a)}{k!} (x-a)^k$$

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

This polynomial is known as the *Taylor polynomial* of degree $n-1$ for f about a.

We can use the exact same method to construct $E(x)$. We need to not affect the first $n-1$ derivatives at $x = a$, so we will need to add some multiple of $Q_n(x)$. However, this time we care about the value of the undifferentiated term at $x = b$:

$$Q_n(b) = \frac{(b-a)^n}{n!}.$$

---

[15]We can think of this as the 'error correcting term', because it fixes the discrepancy between $f(b)$ and $P(b)$ while not ruining the previous construction.

Since we want $E(b) = f(b) - P(b)$, if we are going to have $E(x) = \lambda Q_n(x)$ we are going to need

$$\lambda Q_n(b) = \lambda \frac{(b-a)^n}{n!} = f(b) - P(b)$$
$$\implies \lambda = \frac{[f(b) - P(b)] \cdot n!}{(b-a)^n}.$$

And thus we get

$$E(x) = \frac{[f(b) - P(b)] \cdot n!}{(b-a)^n} Q_n(x)$$
$$= [f(b) - P(b)] \left(\frac{x-a}{b-a}\right)^n.$$

Now with the constructions as above, we can write down $\phi(x) = f(x) - P(x) - E(x)$. Then since $\phi(x)$ satisfies all of the conditions of our higher-order Rolle's theorem, there exists $c \in (a, b)$ such that $\phi^{(n)}(c) = 0$.

Taking the nth derivative, since P is a degree $n-1$ polynomial and thus has a zero derivative, we get

$$\phi^{(n)}(c) = f^{(n)}(c) - [f(b) - P(b)] \cdot \frac{n!}{(b-a)^n} = 0$$
$$\implies f(b) = P(b) + \frac{f^{(n)}(c)}{n!}(b-a)^n.$$

for some $c \in (a, b)$.

This result is Taylor's theorem, specially Taylor's theorem with the Lagrange form of the remainder. For completeness, this proof is given in a standalone form below.

Theorem 4.20 (Taylor's Theorem with Lagrange Remainder)**.** *Let f be continuous and $(n-1)-$times differentiable on $[a, b]$ and n-times differentiable on (a, b). Then we have*

$$f(b) = f(a) + f'(a)(b-a) + \cdots + \frac{f^{(n-1)}(a)}{(n-1)!}(b-a)^{n-1} + \frac{f^{(n)}(c)}{n!}(b-a)^n,$$

for some $c \in (a, b)$.

Proof. We define

$$\phi(x) = f(x) - \sum_{k=0}^{n-1} \frac{f^{(k)}(a)}{k!}(x-a)^k - M\left(\frac{x-a}{b-a}\right)^n,$$

where M is chosen such that $\phi(a) = \phi(b) = 0$. Then differentiating we have $\phi(a) = \phi'(a) = \cdots = \phi^{(n-1)}(a) = 0$.

Then since $\phi(a) = \phi(b)$, there exists $c_1 \in (a, b)$ such that $\phi'(c_1) = 0$. Similarity $\phi'(a) = \phi'(c_1) = 0$, and thus there exists $c_2 \in (a, c_1)$ such that $\phi''(c_2) = 0$. Continuing on in this way, we find a c_n such that $\phi^{(n)}(c_n) = 0$, where $c_n \in (a, b)$.

Then differentiating again we have

$$\phi^{(n)}(c) = f^{(n)}(c) - \frac{M \cdot n!}{(b-a)^n} = 0.$$

Now we know that M is such that $\phi(b) = 0$, so we get

$$M = f(b) - \sum_{k=0}^{n-1} \frac{f^{(k)}(a)}{k!}(b-a)^n.$$

Thus we have

$$f^{(n)}(c) = \frac{M \cdot n!}{(b-a)^n} = \left(f(b) - \sum_{k=0}^{n-1} \frac{f^{(k)}(a)}{k!}(b-a)^n \right) \cdot \frac{n!}{(b-a)^n}$$

$$\implies f(b) = f(a) + f'(a)(b-a) + \cdots + \frac{f^{(n-1)}(a)}{(n-1)!}(b-a)^{n-1} + \frac{f^{(n)}(c)}{n!}(b-a)^n,$$

where $c \in (a, b)$ as required. $\qquad\square$

So we have Taylor's theorem (which in this case can be thought of as a higher-order mean value theorem), let's try and do something with it. One of the main uses of Taylor's theorem is in giving us an infinite series which converges to the value of our function. This involves considering the terms of the degree n taylor polynomial for increasing values of n, as we shall see.

Doing this type of problem has two main steps.

1. Write down Taylor's theorem for the case of the function being considered.

2. Show that the error term tends to zero as n goes to infinity.

We see how this is done using a Tripos question from 2010 (so eh spoilers I guess).

Example 4.21 (Using Taylor's Theorem – Part IA, 2010 Paper 1, Q10). *Problem.* Suppose that $e : \mathbb{R} \to \mathbb{R}$ is a differentiable function such that $e(0) = 1$ and $e'(x) = e(x)$ for all $x \in \mathbb{R}$. Use Taylor's theorem with the remainder in Lagrange's form to prove that

$$e(x) = \sum_{n \geqslant 0} \frac{x^n}{n!} \quad \text{for all} \quad x \in \mathbb{R}$$

[No property of the exponential function may be assumed.]

Solution. We begin by noting that $e(x)$ is infinitely differentiable, with $e^{(n)}(x) = e(x)$. Indeed this is true for $n = 1$, and then it follows by induction for all n. Thus Taylor's theorem with Lagrange Remainder holds for all $n \in \mathbb{N}$.

Thus for all $n \in \mathbb{N}$ we have (using $e^{(n)}(0) = 1$ and $e^{(n)}(c) = e(c)$)

$$e(x) = \sum_{k=0}^{n-1} \frac{x^k}{k!} + \frac{e(c)x^n}{n!}$$

for some $c \in (0, x)$. And thus it suffices to show that $\frac{e(c)x^n}{n!} \to 0$ as $n \to \infty$, since that would imply that $e(x) - \sum_{k=0}^{n-1} \frac{x^k}{k!} \to 0$ as $n \to \infty$. Now $e(t)$ is differentiable and hence continuous on $[0, x]$, and thus is bounded. So it suffices to show that $x^n/n! \to 0$, which holds.

Thus

$$e(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!},$$

as required.

Now it is *not* the case that the remainder term always goes to zero. A common counterexample is $f(x) = \exp(-1/x^2)$ for $x \neq 0$ with $f(0) = 0$.

5 Power Series

Following on from our discussion of infinite series in Chapter 2, we are going to discuss a particular type of infinite series known as a *power series*. Throughout this section, we are going to work mostly in $\mathbb{C}$.

Definition 5.1 (Complex Power Series)**.** A *complex power series* is an infinite series of the form

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

where $z \in \mathbb{C}$ and $a_n \in \mathbb{C}$.

In this section we will look at some properties of power series, and we will use them to (finally) define some functions that we have alluded to quite a bit.

5.1 Radius of Convergence

Before we develop the concept of the radius of convergence, we first need to develop a lemma about the convergence of power series.

Lemma 5.2 (Absolute Convergence of Power Series)**.** *Suppose that $\sum_{n=0}^{\infty} a_n z_0^n$ converges for some $z_0 \in \mathbb{C}$. Then $\sum_{n=0}^{\infty} a_n z^n$ converges absolutely for all $z \in \mathbb{C}$ with $|z| < |z_0|$.*

Proof. Since $\sum_{n=0}^{\infty} a_n z_1^n$ converges, we have $a_n z_1^n \to 0$, implying this sequence is bounded. Then there is some $K > 0$ such that $|a_n z_1^n| < K$ for all n.

So if $|z| < |z_0|$, we have $|a_n z^n| \leq K |z/z_1|^n$. Since the geometric series $\sum_{n=0}^{\infty} |z/z_1|^n$ converges, our result follows by comparison. $\square$

With this lemma we can prove one of our key results in the study of power series, the existence of the *radius of convergence*.

Theorem 5.3 (Radius of Convergence)**.** *Suppose that $a_n \in \mathbb{C}$. Then either the power series $\sum_{n=0}^{\infty} a_n z^n$ converges for all $z \in \mathbb{C}$, or there exists a real number R with $R \geq 0$ such that*

(i) $\sum_{n=0}^{\infty} a_n z^n$ *converges absolutely if $|z| < R$,*

(ii) $\sum_{n=0}^{\infty} a_n z^n$ *diverges if $|z| > R$.*

We say that R is the radius of convergence *of the power series.*

Proof. If $\sum_{n=0}^{\infty} a_n z^n$ converges for all $z \in \mathbb{C}$, then we are done. Otherwise, there must exist $z_1 \in \mathbb{C}$ such that $\sum_{n=0}^{\infty} a_n z_1^n$ diverges. Then by our previous lemma the power series must diverge for all $z \in \mathbb{C}$ where $|z| > |z_1|$.

Now define the set $S = \{|z| : \sum_{n=0}^{\infty} a_n z^n \text{ converges}\}$. This set is bounded above by $|z_1|$, and is non-empty since the power series converges for $z = 0$. Thus it has a supremum, say $R = \sup S$.

Then by definition $\sum_{n=0}^{\infty} a_n z^n$ diverges if $|z| > R$, and then suppose $|z| < R$. Then there must be some z_0 with $|z| < |z_0| < R$ such that the power series converges for z_0 (by the definition of the supremum again). Then by our previous lemma the power series converges absolutely for z. $\square$

What's quite lovely is that when we are inside the radius of convergence, the power series converges absolutely! This means that we can avoid all of those issues with rearranging series that show up when things don't converge absolutely. This can be quite helpful when actually working with power series.

What's also quite nice is that we can employ all of those lovely results we developed in section 2 to find what the radius of convergence of a power series is. Let's look at two lemmas that come from the ratio and root tests, which can give the radius of convergence.

Lemma 5.4 (The Ratio Test for Radius of Convergence). *Let $\sum_{n=0}^{\infty} a_n z^n$ be a power series with radius of convergence R. Then if $|a_{n+1}/a_n| \to \ell$ as $n \to \infty$, we have $R = 1/\ell$ if $\ell \neq 0$, and convergence everywhere if $\ell = 0$.*

Proof. By the ratio test, we have absolute convergence if

$$\lim_{n \to \infty} \left| \frac{a_{n+1}}{a_n} \cdot \frac{z^{n+1}}{z^n} \right| < 1.$$

So if $\ell = 0$, then we have convergence everywhere. Alternatively, if $|z| < 1/\ell$ we have absolute convergence, and if $|z| > 1/\ell$ the series diverges, again by the ratio test. $\square$

Lemma 5.5 (The Root Test for Radius of Convergence). *Let $\sum_{n=0}^{\infty} a_n z^n$ be a power series with radius of convergence R. Then if $|a_n|^{1/n} \to \ell$ as $n \to \infty$, we have $R = 1/\ell$.*

Proof. By the root test, we have absolute convergence if

$$\lim_{n \to \infty} |a_n z^n|^{1/n} < 1.$$

So if $\ell = 0$, then we have convergence everywhere. Alternatively, if $|z| < 1\ell$ we have absolute convergence, and if $|z| > 1/\ell$ the series diverges, again by the root test. $\square$

Now it is time for an important conceptual point: if we have a power series with radius of convergence R, then this only specifies the convergence/divergence inside or outside the circle $|z| = R$ (the *circle of conver-*

gence). Specifically it says nothing about the behaviour on the circle. To see this, we will look at a few examples[16].

Example 5.6 (Convergence on $|z| = R$). Consider the following power series:

$$\text{(i)} \sum_{n=0}^{\infty} z^n, \quad \text{(ii)} \sum_{n=1}^{\infty} \frac{z^n}{n} \quad \text{(iii)} \sum_{n=1}^{\infty} \frac{z^n}{n^2}.$$

All of these have a radius of convergence $R = 1$, however for $|z| = 1$ for each power series we have:

1. Divergence if $|z| = 1$.

2. Convergence[17] if $|z| = 1$ with $z \neq 1$, and divergence if $z = 1$.

3. Convergence if $|z| = 1$.

Thus we can't say anything about the case $|z| = R$ and each case really has to be treated separately.

5.2   Differentiating Power Series

So things are generally quite nice inside the circle of convergence. One of the nice things we can do is differentiate! Even better, we can differentiate power series just like we do polynomials! The following results you need to know (including the lemmas), but the proofs are not examinable (feel free to skip them).

We are going to show that the power series $\sum_{n=0}^{\infty} a_n z^n$ with radius of convergence R has the derivative $\sum_{n=0}^{\infty} n a_n z^{n-1}$, whenever $|z| < R$. Before showing this, we will need a lemma which gives us that this second power series converges for $|z| < R$.

Lemma 5.7. *Let* $f(z) = \sum_{n=0}^{\infty} a_n z^n$ *be a power series whose radius of convergence is* R. *Then if* $0 < r < R$, *the power series* $\sum_{n=1}^{\infty} n|a_n| r^{n-1}$ *converges.*

Proof. Pick w such that $r < |w| < R$. Then the power series $\sum_{n=0}^{\infty} a_n w^n$ converges, so the terms $|a_n w^n|$ are bounded above by say M. Then we have

$$n|a_n| r^n = n|a_n w^n| \frac{r^n}{|w|^n} \leq M n \frac{r^n}{|w|^n}.$$

But then the series $\sum_{n=1}^{\infty} M n (r/|w|)^n$ converges by the ratio test, and thus by the comparison test $\sum_{n=1}^{\infty} n|a_n| r^{n-1}$ converges. $\square$

---

[16]This should remind you of the deadly sin mentioned in section 2, where the root or ratio test said nothing if the limit was 1.

[17]Have a look at Example Sheet 1 for this result

Applying this lemma twice also gives us that the power series $\sum_{n=2}^{\infty} n(n-1)|a_n|r^{n-2}$ converges. We are now ready to prove our theorem about differentiating power series.

Theorem 5.8 (Differentiating Power Series). *Let $f(z) = \sum_{n=0}^{\infty} a_n z^n$ be a power series whose radius of convergence is R. Then f is differentiable at all points with $|z| < R$, and $f'(z) = \sum_{n=0}^{\infty} n a_n z^{n-1}$.*

Proof. Using the difference of powers factorisation, we have

$$\frac{f(z+h) - f(z)}{h} = \sum_{n=0}^{\infty} a_n \frac{(z+h)^n - z^n}{h}$$

$$= \sum_{n=1}^{\infty} a_n \left[(z+h)^{n-1} + (z+h)^{n-2} z + \cdots + z^{n-1} \right].$$

Then, since $\sum_{n=1}^{\infty} n a_n z^{n-1}$ converges by our lemma, we have

$$\frac{f(z+h) - f(z)}{h} - \sum_{n=1}^{\infty} n a_n z^{n-1} = \sum_{n=1}^{\infty} a_n \left(\sum_{j=0}^{n-1} z^j (z+h)^{n-1-j} - z^{n-1} \right)$$

$$= \sum_{n=1}^{\infty} a_n \left(\sum_{j=0}^{n-1} z^j \left[(z+h)^{n-1-j} - z^{n-1-j} \right] \right).$$

Now we want to bound the terms of this inside sum. Choose r such that $|z| < r < R$ and h such that $|z| + |h| < r$. Then employing the difference of powers factorisation again we get

$$|(z+h)^{n-1-j} - z^{n-1-j}| = |h((z+h)^{n-2-j} + z(z+h)^{n-3-j} + \cdots + z^{n-2-j})|$$

$$\leq |h|(n-1-j)r^{n-2-j}.$$

Applying this bound, assuming the right hand sum converges we have

$$\left| \frac{f(z+h) - f(z)}{h} - \sum_{n=1}^{\infty} n a_n z^{n-1} \right| \leq \sum_{n=1}^{\infty} |a_n| \left(\sum_{j=0}^{n-1} r^j |h|(n-1-j)r^{n-2-j} \right)$$

$$= |h| \sum_{n=1}^{\infty} |a_n| r^{n-2} \left(\sum_{j=0}^{n-1} j \right)$$

$$= \frac{1}{2} |h| \sum_{n=2}^{\infty} |a_n| n(n-1) r^{n-2}.$$

Using our previous lemma twice shows that the right hand sum converges, and thus

$$\frac{f(z+h) - f(z)}{h} - \sum_{n=1}^{\infty} n a_n z^{n-1} \to 0 \quad \text{as} \quad h \to 0,$$

that is, $f'(z) = \sum_{n=0}^{\infty} n a_n z^{n-1}$. $\qquad \square$

5.3 Exponential, Trigonometric & Hyperbolic Functions

With all that out of the way, let's define some functions! You'll hopefully be familiar with all of these anyway, but we will now be able to define them properly.

We are going to need one little lemma before we proceed.

Lemma 5.9 (Constant Value Theorem Over $\mathbb{C}$). *If $F : \mathbb{C} \to \mathbb{C}$ has $F'(z) = 0$ for all $z \in \mathbb{C}$, then $F(z)$ is constant for all z.*

Proof. For some $z \in \mathbb{C}$, we define the functions $u_z, v_z : \mathbb{R} \to \mathbb{R}$ such that $F(tz) = u_z(t) + iv_z(t)$. By the chain rule $F'(tz)z = u'_z(t) + iv'_z(t) = 0$ for any $t \in \mathbb{R}$. Then comparing real and imaginary parts, this implies $u'_z(t) = 0$ and $v'_z(t) = 0$, for all $t \in \mathbb{R}$. So u_z and v_z are constant.

Then for any $z \in \mathbb{C}$ we have $F(z) = u_z(1) + iv_z(1) = u_z(0) + iv_z(0) = F(0)$, so F is constant too. $\square$

5.3.1 The Exponential & Logarithm Function

We will begin with one of the most important functions, the exponential function.

Definition 5.10 (The Exponential Finction). We define the *exponential function* $\exp : \mathbb{C} \to \mathbb{C}$ by the power series

$$\exp(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!}.$$

When we state a definition using power series, we pretty much always have to check convergence straight away so that we do not end up writing complete nonsense. For this power series, we can see by the ratio test that it converges for all $z \in \mathbb{C}$. By the result in the previous section, we can also see that it's differentiable.

Now we can prove all of the properties that you already are aware of.

Proposition 5.11 (Properties of the Exponential Function). *Let $z, w \in \mathbb{C}$. Then the following hold.*

 (i) $\exp'(z) = \exp(z)$.

 (ii) $\exp(z)\exp(w) = \exp(z + w)$.

 (iii) $\exp(x) > 0$ for $x \in \mathbb{R}$.

 (iv) $\exp(x)$ is strictly increasing on $\mathbb{R}$.

 (v) $\exp(x) \to \infty$ as $x \to \infty$, $\exp(x) \to 0$ as $x \to -\infty$.

 (vi) $\exp : \mathbb{R} \to (0, \infty)$ is a bijection.

Proof. We prove each individually.

(i) This follows from differentiating the power series.

(ii) Let $a, b \in \mathbb{C}$ and define $F(z) = \exp(a + b - z) \exp(z)$. Then $F'(z) = -\exp(a + b - z) \exp(z) + \exp(a + b - z) \exp(z) = 0$, and thus by the constant value theorem, F is constant. So $\exp(a + b - z) \exp(z) = F(0) = \exp(a + b)$, and setting $z = b$ gives $\exp(a) \exp(b) = \exp(a + b)$.

(iii) Clearly $\exp(x) > 0$ for all $x \geq 0$, and $\exp(0) = 1$. Then $\exp(0) = \exp(x - x) = \exp(x) \exp(-x) = 1$, and thus $\exp(-x) > 0$ for $x > 0$.

(iv) Differentiating, $\exp'(x) = \exp(x) > 0$, and thus exp is strictly increasing on $\mathbb{R}$.

(v) Truncating the power series, $\exp(x) > 1 + x$ for $x > 0$. Thus if $x \to \infty$, $\exp(x) \to \infty$. Then $\exp(-x) = 1/\exp(x)$, so $\exp(x) \to 0$ as $x \to -\infty$.

(vi) Injectivity follows directly from exp being strictly increasing. For surjectivity, take $y > 0$. Then by (v) there exists $a, b \in \mathbb{R}$ such that $\exp(a) < y < \exp(b)$, and by the intermediate value theorem there is $c \in (a, b)$ such that $\exp(c) = y$. $\qquad\square$

Since the exponential function is a bijection $\mathbb{R} \to \mathbb{R}^+$, it also has a well defined inverse.

Definition 5.12 (The Logarithm Function)**.** We define the *logarithm function* $\log : (0, \infty) \to \mathbb{R}$ to be the inverse of exp.

By the inverse function theorem, this is a differentiable function, and its derivative is given by $\log'(t) = 1/t$ for $t > 0$.

Remark. As defined here, the logarithm function is an inverse of $\exp : \mathbb{R} \to (0, \infty)$, not over the whole of $\mathbb{C}$. In general, exp is *not* a bijection, and thus it having an inverse doesn't make sense (with the tools we have developed, this is remedied in later courses).

Using both the exponential and logarithm functions, we can define powers in the general case. For $x > 0$ and $\alpha \in \mathbb{R}$, we define $x^\alpha = \exp(\alpha \log x)$. With this we get the normal 'rules of indices' being properties of the exponential function, and it it not hard to see that this definition matches how powers would be previously defined for $\alpha \in \mathbb{Q}$. This also gives us a new shorthand for the exponential function: $\exp(z) = e^z$.

Aside: Multiplying Infinite Series

In Theorem 5.11 we proved a result involving multiplying two exponential functions together. We also managed to do this in a way that avoided using the power series definition of the function. Despite this, I think it's worth taking a moment to look properly at the multiplication of infinite series[18].

When we talk about trying multiplying infinite series, we are really talking about taking two infinite series so that they can be combined into a new infinite series[19], which hopefully converges to the product of the two original series.

So how can we do this? A natural approach would be to just 'expand everything', but unfortunately that's not really how infinite series work, so we can't do that. Still, we can use this idea as inspiration for a more mathematically sound approach.

Being quite informal and hand-wavy, let's consider two power series $\sum_{n=0}^{\infty} a_n z^n$ and $\sum_{n=0}^{\infty} b_n z^n$, multiply them term by term and then group by powers of z. Doing this, we get something like

$$(a_0 + a_1 z + a_2 z^2 + a_3 z^3 + \cdots) \cdot (b_0 + b_1 z + b_2 z^2 + b_3 z^3 + \cdots)$$
$$= a_0 b_0 + (a_0 b_1 + a_1 b_0)z + (a_0 b_2 + a_1 b_1 + a_2 b_2)z^2 + (a_0 b_3 + a_1 b_2 + a_2 b_1 + a_3 b_0)z^3 + \cdots.$$

Taking inspiration from this, for two sequence a_n and b_n we can define their *convolution* to be the sequence c_n where

$$c_n = a_0 b_n + a_1 b_{n-1} + \cdots + a_n b_0.$$

It is this construction that will form the basis of our product of infinite series.

Definition (Cauchy Product)**.** Given two infinite series $\sum_{n=0}^{\infty} a_n$ and $\sum_{n=0}^{\infty} b_n$, their *Cauchy product* is the infinite series

$$\sum_{n=0}^{\infty} c_n,$$

where $c_n = a_0 b_n + a_1 b_{n-1} + \cdots + a_n b_0$.

The Cauchy product does give us a way to obtain a 'product' of infinite series, but before we go any further some important caveats need to be pointed out. The most notable is the following: the infinite series obtained with the Cauchy product *does not necessarily converge*[20].

---

[18]This section is completely not-examinable and the proofs are relatively long, so I suggest that you skip it if you are using these notes for revision.

[19]If we don't care about obtaining a new infinite series, this discussion can be completely ignored.

[20]It's also possible for two divergent series to have a convergent Cauchy product. Try coming up with an example as an exercise.

Example (Convergent Series Whose Cauchy Product Diverges)**.** Consider the convergent[21] series $\sum_{n=0}^{\infty} \frac{(-1)^n}{\sqrt{n+1}}$. Then the Cauchy product of this series with itself is given by

$$\sum_{n=0}^{\infty} \sum_{k=0}^{n} \frac{(-1)^k}{\sqrt{k+1}} \frac{(-1)^{n-k}}{\sqrt{n-k+1}}.$$

But then since

$$(k+1)(n-k+1) \leq \left(\frac{n}{2}+1\right)^2 - \left(\frac{n}{2}-k\right)^2 \leq \left(\frac{n}{2}+1\right)^2,$$

the terms of the Cauchy product satisfy

$$\left| \sum_{k=0}^{n} \frac{(-1)^k}{\sqrt{k+1}} \frac{(-1)^{n-k}}{\sqrt{n-k+1}} \right| \geq \sum_{k=0}^{n} \frac{1}{n/2+1} = \frac{n+1}{n/2+1},$$

and thus they don't tend to 0 so the series cannot converge. This implies that we have two convergent series whose Cauchy product does not converge.

This is obviously not a very good thing to happen, and we will spend the rest of this aside trying to deal with this possible issue. Luckily enough, this issue can be completely avoided if we know that one of the series being 'multiplied' converges absolutely[22].

Theorem (Mertens' Theorem)**.** *Suppose that $\sum_{n=0}^{\infty} a_n = A$ converges absolutely, and $\sum_{n=0}^{\infty} b_n = B$ converges. Then their Cauchy product $\sum_{n=0}^{\infty} c_n$ converges, and*

$$\sum_{n=0}^{\infty} c_n = AB.$$

Proof. Let the partial sums of the series be

$$A_n = \sum_{k=0}^{n} a_k, \quad B_n = \sum_{k=0}^{n} b_k, \quad C_n = \sum_{k=0}^{n} c_k.$$

Then we have

$$\begin{aligned} C_n &= a_0 b_0 + (a_0 b_1 + a_1 b_0) + (a_0 b_2 + a_1 b_1 + a_2 b_0) + \cdots + (a_0 b_n + \cdots + a_{n-1} b_n) \\ &= a_0 B_n + a_1 B_{n-1} + \cdots + a_n B_0 \\ &= (a_0 B + \cdots + a_n B) - (a_0 B + \cdots + a_n B) + a_0 B_n + a_1 B_{n-1} + \cdots + a_n B_0 \\ &= A_n B + a_0 (B_n - B) + a_1 (B_{n-1} - B) + \cdots + a_n (B_0 - B). \end{aligned}$$

---

[21] Exercise!

[22] This is another reason why absolute convergence is a great property to have.

We wish to show $C_n \to AB$, and we will do so using three bounds. Given $\epsilon > 0$, since $A_n \to A$, there is an integer L such that $n \geq L$ implies

$$|A_n - A| \leq \frac{\epsilon/3}{|B|}.$$

Then since $\sum_{k=0}^{\infty} |a_k|$ converges and $B_n \to B$, there is an integer M such that $n \geq M$ implies

$$|B_n - B| \leq \frac{\epsilon/3}{\sum_{k=0}^{\infty} |a_k|}.$$

Also, since $\sum_{k=0}^{\infty} a_k$ converges, $a_n \to 0$, and thus there is an integer N such that $n \geq N$ implies that

$$|a_n| \leq \frac{\epsilon/3}{\sum_{k=0}^{M-1} |B_k - B|}.$$

Combining these[23], for $N \geq \max\{L, M + N\}$ we obtain

$$|C_n - AB| = \left| (A_n - A)B + \sum_{k=0}^{n} a_{n-k}(B_k - B) \right|$$

$$\leq |A_n - A||B| + \sum_{k=0}^{M-1} |a_{n-k}||B_k - B| + \sum_{k=M}^{n} |a_{n-k}||B_k - B| \leq \epsilon$$

Thus $\sum_{n=0}^{\infty} c_n = AB$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

We can apply this result to power series. Recalling that power series converge absolutely inside their circle of convergence, we can safely multiply power series using the Cauchy product.

Corollary (Multiplying Power Series). *Let $\sum_{n=0}^{\infty} a_n z^n$ and $\sum_{n=0}^{\infty} b_n z^n$ be power series with radius of convergence R_1 and R_2 respectively. Then if $|z| < \min\{R_1, R_2\}$ we have*

$$\left(\sum_{n=0}^{\infty} a_n z^n \right) \cdot \left(\sum_{n=0}^{\infty} b_n z^n \right) = \left(\sum_{n=0}^{\infty} c_n z^n \right),$$

where $c_n = a_0 b_n + a_1 b_{n-1} + \cdots + a_n b_0$.

So to summarise: when inside the circle of convergence, we can multiply power series in the natural way, and the result will be as we expect. Using this corollary, we could prove $\exp(a + b) = \exp(a)\exp(b)$ directly (though admittedly the other proof is much faster).

---

[23]We use $M + N$ because we will need $n - (M - 1) \geq N$ to make the bounds work.

5.3.2 The Trigonometric Functions

We can use the exponential function to define the familiar trigonometric functions.

Definition 5.13 (Sine & Cosine)**.** We define the *sine* and *cosine functions* by the power series

$$\sin z = \frac{e^{iz} - e^{-iz}}{2i} = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \frac{z^7}{7!} + \cdots$$

$$\cos z = \frac{e^{iz} + e^{-iz}}{2} = 1 - \frac{z^2}{2} + \frac{z^4}{4!} - \frac{z^6}{6!} + \cdots$$

where $z \in \mathbb{R}$.

Stating these definitions in terms of the exponential function make lots of identities quite easy to derive.

Proposition 5.14 (Properties of sin and cos)**.** *Let $z, w \in \mathbb{R}$. Then the following hold.*

(i) $\sin' z = \cos z$, $\cos' z = -\sin z$.

(ii) $\sin^2 z + \cos^2 z = 1$.

(iii) $|\sin z|, |\cos z| \leq 1$.

(iv) $\cos(z + w) = \cos z \cos w - \sin z \sin w$.

(v) $\sin(z + w) = \sin z \cos w + \cos z \sin w$.

Proof. We prove each individually, mainly by computation.

(i) For sin we have $\sin' z = \left[(e^{iz} - e^{-iz}/2i\right]' = (e^{iz} + e^{-iz})/2 = \cos z$, and for cos we have $\cos' z = \left[(e^{iz} + e^{-iz})/2\right]' = (e^{-iz} - e^{iz})/2i = -\sin z$.

(ii) $\sin^2 z + \cos^2 z = -(e^{2iz} + e^{-2iz} + 2)/4 + (e^{2iz} + e^{-2iz} - 2)/4 = 1$, as required.

(iii) This follows from the above result.

(iv) Expanding the right hand expression gives

$$\cos z \cos w - \sin z \sin w$$
$$= \frac{\left(e^{iz} + e^{-iz}\right)\left(e^{iw} + e^{-iw}\right)}{4} + \frac{\left(e^{iz} - e^{-iz}\right)\left(e^{iw} - e^{-iw}\right)}{4}$$
$$= \frac{e^{i(z+w)} + e^{-i(z+w)}}{2}$$
$$= \cos(z + w).$$

(v) Differentiating the above result with respect to z, we have $-\sin z \cos w - \cos z \sin w = -\sin(z+w)$, giving us our desired result. $\square$

One of the more notable properties of sin and cos is that they are periodic, which is what we are going to show next. To establish this, we can try evaluating one of them (say cos) at some values near where we expect a root to be (guided by our previous knowledge of these functions).

Lemma 5.15 (Smallest Root of cos). *There is a smallest positive x such that $\cos(x) = 0$.*

Proof. We begin by computing the sign of the derivative over $(0, 2)$. Since $\cos'(x) = -\sin(x)$, we can use the inequality $\frac{x^{2n-1}}{(2n-1)!} > \frac{x^{2n+1}}{(2n+1)!}$ to bound

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots > 0,$$

so cos has a negative derivative and must be decreasing on $(0, 2)$.

Evaluating the power series for cos at $\sqrt{2}$ and $\sqrt{3}$, we can see that the function changes sign

$$\cos\sqrt{2} = 1 - \frac{2}{2} + \frac{2^2}{4!} - \frac{2^3}{6!} + \cdots > 0, \cos\sqrt{3} = 1 - \frac{3}{2} + \frac{3^2}{4!} - \frac{3^3}{6!} + \cdots < 0.$$

Then by the intermediate value theorem, there must be a root in the interval $(\sqrt{2}, \sqrt{3})$. Also since cos is decreasing on $(0, 2)$, this must also be the smallest positive root. $\square$

The existence of this root gives us a way to define π.

Definition 5.16 (π). π is the smallest positive real number such that $\cos \pi/2 = 0$.

We can easily determine the value of $\sin \pi/2$ too. We have $\sin^2 \pi/2 + \cos^2 \pi/2 = 1$, so $\sin^2 \pi/2 = 1$, and knowing from the proof of Theorem 5.15 that $\sin(x) > 0$ on $(0, 2)$, we can deduce that $\sin \pi/2 = 1$. These results are all we need to show that sin and cos are periodic, and also that the two functions are just translated versions of each-other.

Proposition 5.17 (Periodicity of sin and cos). *The functions* sin *and* cos *are periodic with a period of 2π, and are related as follows.*

$$\begin{aligned}
\sin(z + \pi/2) &= \cos z, & \cos(z + \pi/2) &= -\sin z, \\
\sin(z + \pi) &= -\sin z, & \cos(z + \pi) &= -\cos z, \\
\sin(z + 2\pi) &= \sin z, & \cos(z + 2\pi) &= \cos z.
\end{aligned}$$

Proof. This follows directly from $\sin(\pi/2) = 1$, $\cos(\pi/2) = 0$ and the angle addition formulas. $\square$

Remark (Periodicity of exp). If we use the power series for sin and cos, we can write $\exp(iz) = \cos z + i \sin z$, from which the above result implies that exp is periodic with a period of 2π.

With sin and cos have been defined properly, we can then define tan, sec, csc and all of those fun functions in the normal way.

5.3.3   The Hyperbolic Functions

The last set of functions we will touch on is the (probably) slightly less familiar hyperbolic functions. As with the standard trigonometric functions, we can state the definitions using the exponential function.

Definition 5.18 (Hyperbolic Sine & Cosine). We define the *hyperbolic sine* and *hyperbolic cosine functions* by the power series

$$\sinh z = \frac{e^z - e^{-z}}{2} = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \frac{x^7}{7!} + \cdots$$

$$\cosh z = \frac{e^z + e^{-z}}{2} = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \frac{x^6}{6!} + \cdots$$

where $z \in \mathbb{R}$.

Remark. The basic hyperbolic and trigonometric functions are related by $\cosh z = \cos(iz)$ and $\sinh z = -i \sin(iz)$. Remembering this can allow you to adapt facts about sin and cos to facts about sinh and cosh.

We will state a few properties for convenience.

Proposition 5.19 (Properties of sinh & cosh). *Let $z, w \in \mathbb{R}$. Then the following hold.*

(i) $\sinh' z = \cosh z$, $\cosh' z = \sin z$.

(ii) $\cosh^2 z - \sinh^2 z = 1$.

(iii) $\sinh(z + w) = \sinh z \cosh w + \cosh z \sinh w$.

(iv) $\cosh(z + w) = \cosh z \cosh w + \sinh z \sinh w$.

Proof. We will use the observation made in the above remark.

(i) For sinh we have $\sinh' z = [-i \sin(iz)]' = \cos(iz) = \cosh z$, and for cosh we have $\cosh' z = [\cos(iz)]' = -i \sin(iz) = \sinh z$.

(ii) $\cosh^2 z - \sinh^2 z = \cos(iz)^2 - (-i \sin(iz))^2 = \cos^2(iz) + \sin^2(iz) = 1$.

(iii) $\sinh(z + w) = -i \sin(iz + iw) = \sinh(z) \cosh(w) + \cosh(z) \sinh(w)$.

(iv) $\cosh(z + w) = \cos(iz + iw) = \cosh(z) \cosh(w) + \sinh(z) \sinh(w)$. $\square$

6 Integration

We are now prepared, both technically and emotionally, to talk about *integrals*. The idea behind integration is to assign some sort of 'area' to sets. The first issue we run into is that is not quite clear what 'area' actually means (in mathematical terms), so to avoid this difficulty, we will instead define the notion of an integral (specifically the *Riemann integral*, there are other types) and we will use this as our definition of area (not vice versa).

Throughout this section, we will deal with the integration of bounded functions defined on bounded intervals, $f : [a, b] \to \mathbb{R}$. We will discuss a slight generalisation later on.

6.1 Dissections, Upper & Lower Sums, and Riemann Integrals

As with most ideas in analysis, integration depends fundamentally on a 'limiting process'. For Riemann integrals, this is conducted by considering *dissections* of the interval in which we will be integrating over.

Definition 6.1 (Dissection). A *dissection* $\mathcal{D}$ of a finite interval $[a, b]$ is a finite subset of $[a, b]$ containing the endpoints. We typically write $\mathcal{D} = \{x_0, x_1, x_2, \ldots, x_n\}$ with $a = x_0 \leq x_1 \leq \cdots \leq x_n = b$.

For a given dissection, because we are dealing with bounded functions we can sensibly define the lower and upper sums of a function as follows.

Definition 6.2 (Upper & Lower Sums). Let $f : [a, b] \to \mathbb{R}$ be a bounded function, and let $\mathcal{D}$ be a dissection of $[a, b]$. We define the *upper sum* $S_{\mathcal{D}}(f)$ and *lower sum* $s_{\mathcal{D}}(f)$ by

$$S_{\mathcal{D}}(f) = \sum_{j=1}^{n} (x_j - x_{j-1}) \sup_{x \in [x_{j-1}, x_j]} f(x),$$

$$s_{\mathcal{D}}(f) = \sum_{j=1}^{n} (x_j - x_{j-1}) \inf_{x \in [x_{j-1}, x_j]} f(x).$$
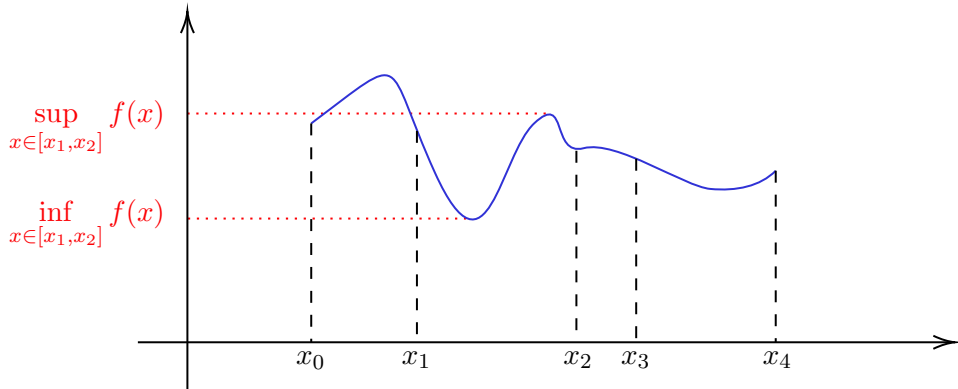
Let's make some observations. By construction we have $S_{\mathcal{D}}(f) \geq s_{\mathcal{D}}(f)$, and we can also see intuitively that replacing a dissection by a more elaborate one will decrease the upper sum and increase the lower sum. Indeed, if we say that $\mathcal{D}'$ *refines* $\mathcal{D}$ when $\mathcal{D}' \supseteq \mathcal{D}$, then we can formalise this intuition as follows.

Lemma 6.3 (Refinement Lemma). *If $f : [a, b] \to \mathbb{R}$ is a bounded function and $\mathcal{D}, \mathcal{D}'$ are dissections of $[a, b]$ such that $\mathcal{D} \subseteq \mathcal{D}'$, then*

$$s_{\mathcal{D}}(f) \leq s_{\mathcal{D}'}(f) \leq S_{\mathcal{D}'}(f) \leq S_{\mathcal{D}}(f).$$

Proof. It suffices to show that this holds when $\mathcal{D}'$ and $\mathcal{D}$ differ by a single element (since then we can just repeatedly apply this argument to obtain the general case). Suppose that $\mathcal{D}' = \{x_0, \ldots, x_n\}$, and that $\mathcal{D} = \mathcal{D}' \backslash \{x_i\}$ where $i \neq 0, n$. Then let

$$A = \sup_{x \in [x_{i-1}, x_i]} f(x), \quad \text{and} \quad B = \sup_{x \in [x_i, x_{i+1}]} f(x).$$

We can then establish the right hand side of the inequality with

$$\begin{aligned} S_{\mathcal{D}}(f) - S_{\mathcal{D}'}(f) &= (x_{i+1} - x_{i-1}) \max\{A, B\} - (x_i - x_{i-1})A - (x_{i+1} - x_i)B \\ &\geq \max\{A, B\}[(x_{i+1} - x_{i-1}) - (x_i - x_{i-1}) - (x_{i+1} - x_i)] = 0. \end{aligned}$$

Then the left hand side of the inequality follows similarly (taking infima). $\square$

A direct consequence of this lemma is that lower sums can *never* exceed upper sums.

Lemma 6.4 (Key Integration Property). *If $f : [a, b] \to \mathbb{R}$ is a bounded function and $\mathcal{D}_1, \mathcal{D}_2$ are dissections of $[a, b]$ then*

$$s_{\mathcal{D}_1}(f) \leq S_{\mathcal{D}_2}(f).$$

Proof. This follows from the previous lemma by noting that $s_{\mathcal{D}_1}(f) \leq s_{\mathcal{D}_1 \cup \mathcal{D}_2}(f)$, and $S_{\mathcal{D}_1 \cup \mathcal{D}_2}(f) \leq S_{\mathcal{D}_2}(f)$. $\square$

We are now (finally) ready to say what it means for a function to be Riemann integrable.

Definition 6.5 (Upper & Lower Integrals). For a bounded function $f : [a, b] \to \mathbb{R}$ we define the *upper integral* as $I^*(f) = \inf_{\mathcal{D}} S_{\mathcal{D}}(f)$ and the *lower integral* as $I_*(f) = \sup_{\mathcal{D}} s_{\mathcal{D}}(f)$.

Note that Theorem 6.4, along with the boundedness of the function, guarantees that the upper and lower integrals always exist.

Definition 6.6 (Riemann Integrable)**.** If $f : [a,b] \to \mathbb{R}$ is a bounded function and $I^*(f) = I_*(f)$, then we say f is *Riemann integrable* and write

$$\int_a^b f(x)\mathrm{d}x = I^*(f).$$

So this is what it means for a function to be Riemann integrable[24]. In general using this definition directly on a function to show that it's integrable is quite tricky. As an example, we can attempt to show that $f(x) = x$ is integrable.

Example 6.7 ($f(x) = x$ is Integrable)**.** Consider the function $f(x) = x$ defined on the interval $[a,b]$. Then for any dissection $\mathcal{D} = \{x_0, x_1, \ldots, x_n\}$ of $[a,b]$ we have the upper sum given by

$$\begin{aligned} S_{\mathcal{D}}(f) &= \sum_{j=1}^{n} (x_j - x_{j-1})x_j \\ &= \sum_{j=1}^{n} \left(\frac{x_j^2}{2} - \frac{x_{j-1}^2}{2} + \frac{x_j^2}{2} + \frac{x_{j-1}^2}{2} - x_{j-1}x_j \right) \\ &= \sum_{j=1}^{n} \left(\frac{x_j^2}{2} - \frac{x_{j-1}^2}{2} \right) + \frac{1}{2} \sum_{j=1}^{n} (x_j - x_{j-1})^2 \\ &= \left(\frac{b^2}{2} - \frac{a^2}{2} \right) + \frac{1}{2} \sum_{j=1}^{n} (x_j - x_{j-1})^2 . \end{aligned}$$

If $x_j - x_{j-1} < \delta$ for all j, then we can obtain the upper bound

$$S_{\mathcal{D}}(f) \leq \left(\frac{b^2}{2} - \frac{a^2}{2} \right) + \frac{1}{2} \sum_{j=1}^{n} \delta\, (x_j - x_{j-1}) = \left(\frac{b^2}{2} - \frac{a^2}{2} \right) + \frac{\delta}{2}(b - a).$$

So given $\epsilon > 0$, there is some dissection $\mathcal{D}$ such that $S_{\mathcal{D}}(f) \leq (b^2 - a^2)/2 + \epsilon$, and a similar argument also gives $s_{\mathcal{D}}(f) \geq (b^2 - a^2)/2 - \epsilon$. So by our key integration property, we have

$$\frac{1}{2}(b^2 - a^2) - \epsilon \leq I^*(f), I_*(f) \leq \frac{1}{2}(b^2 - a^2) + \epsilon,$$

for any $\epsilon > 0$, and thus $I^*(f) = I_*(f) = (b^2 - a^2)/2$. This show that $f(x) = x$ is integrable over $[a,b]$, with

$$\int_a^b x\mathrm{d}x = \frac{b^2}{2} - \frac{a^2}{2}.$$

---

[24]We will regularly drop the 'Riemann', and when referring to 'integrability' we will always be talking only about Riemann integrability.

There are a few ways that we can streamline this argument. Later on we will develop more general integrability results, but the following criterion is also quite useful for showing particular functions are integrable.

Theorem 6.8 (Riemann's Integrability Criterion). *Let $f : [a, b] \to \mathbb{R}$ be a bounded function. Then f is Riemann integrable if and only if given any $\epsilon > 0$, we can find a dissection $\mathcal{D}$ with*

$$S_{\mathcal{D}}(f) - s_{\mathcal{D}}(f) < \epsilon.$$

Proof. If f is Riemann integrable, then given $\epsilon > 0$ there must exist dissections $\mathcal{D}_1$, $\mathcal{D}_2$ such that

$$S_{\mathcal{D}_1}(f) < \int_a^b f(x)\mathrm{d}x + \frac{\epsilon}{2} \quad \text{and} \quad s_{\mathcal{D}_2}(f) > \int_a^b f(x)\mathrm{d}x - \frac{\epsilon}{2}.$$

Then taking $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$ we have $S_{\mathcal{D}}(f) \le S_{\mathcal{D}_1}(f)$, and $s_{\mathcal{D}}(f) \ge s_{\mathcal{D}_2}(f)$, and thus

$$S_{\mathcal{D}}(f) - s_{\mathcal{D}}(f) \le S_{\mathcal{D}_1}(f) - s_{\mathcal{D}_2}(f) < \epsilon.$$

Otherwise if for any $\epsilon > 0$ we can find a dissection $\mathcal{D}$ such that $S_{\mathcal{D}}(f) - s_{\mathcal{D}}(f) < \epsilon$, then we would have

$$\inf_{\mathcal{D}'} S_{\mathcal{D}'}(f) - \sup_{\mathcal{D}'} s_{\mathcal{D}'}(f) \le S_{\mathcal{D}}(f) - s_{\mathcal{D}}(f) < \epsilon,$$

that is, $I^*(f) - I_*(f) < \epsilon$ for all $\epsilon > 0$. So we must have $I^*(f) = I_*(f)$, implying f is Riemann integrable. $\square$

6.2   Properties of Integration

We now need to prove many of the properties of integrals that you are (most likely) already familiar with. The first result we would like to prove is that an arbitrary linear combination of integrable functions is integrable, but we will need to prove some smaller results first.

The proofs below are a little bit fiddly, but they all follow the same basic idea: simplify the problem as much as possible, think a bit about what's going on with the suprema/infima, and use our key property of integration, Riemann's criterion and so on.

Proposition 6.9 (Adding Integrable Functions). *If $f, g : [a, b] \to \mathbb{R}$ are Riemann integrable, then so is $f + g$ and*

$$\int_a^b f(x) + g(x)\mathrm{d}x = \int_a^b f(x)\mathrm{d}x + \int_a^b g(x)\mathrm{d}x.$$

Proof. We begin by noting the general fact

$$\sup_{x\in[x_{j-1},x_j]}(f(x)+g(x)) \le \sup_{x\in[x_{j-1},x_j]}f(x) + \sup_{x\in[x_{j-1},x_j]}g(x),$$

which directly implies that $S_{\mathcal{D}}(f+g) \le S_{\mathcal{D}}(f)+S_{\mathcal{D}}(g)$. A similar fact about infima implies $s_{\mathcal{D}}(f+g) \ge s_{\mathcal{D}}(f) + s_{\mathcal{D}}(g)$.

Now let $\mathcal{D}_1$ and $\mathcal{D}_2$ be dissections, and let $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$. Then we have

$$S_{\mathcal{D}}(f+g) \le S_{\mathcal{D}}(f) + S_{\mathcal{D}}(g) \le S_{\mathcal{D}_1}(f) + S_{\mathcal{D}_2}(g),$$

which implies that $I^*(f+g) \le I^*(f)+I^*(g)$. Similarly we have $I_*(f+g) \ge I_*(f) + I_*(g)$.

Since f and g are integrable, their upper and lower integrals are equal, so combining the two previous inequalities gives

$$\int_a^b f(x)\mathrm{d}x + \int_a^b g(x)\mathrm{d}x \le I_*(f+g) \le I^*(f+g) \le \int_a^b f(x)\mathrm{d}x + \int_a^b g(x)\mathrm{d}x,$$

so $I_*(f+g) = I^*(f+g)$, showing $f+g$ is integrable, with the integral given by $\int_a^b f(x) + g(x)\mathrm{d}x = \int_a^b f(x)\mathrm{d}x + \int_a^b g(x)\mathrm{d}x$. $\qquad\square$

Proposition 6.10 (Negating Integrable Functions). *If $f : [a,b] \to \mathbb{R}$ is Riemann integrable, then so is $-f$, and*

$$\int_a^b (-f(x))\mathrm{d}x = -\int_a^b f(x)\mathrm{d}x.$$

Proof. Let $\mathcal{D}$ be a dissection of $[a,b]$. Then since $\sup_x -f(x) = -\inf_x f(x)$, we have

$$S_{\mathcal{D}}(-f) = \sum_{j=1}^n (x_j - x_{j-1}) \sup_{x\in[x_{j-1},x_j]}(-f(x))$$

$$= -\sum_{j=1}^n (x_j - x_{j-1}) \inf_{x\in[x_{j-1},x_j]} f(x)$$

$$= -s_{\mathcal{D}}(f).$$

Similarly $s_{\mathcal{D}}(-f) = -S_{\mathcal{D}}(f)$. This then implies that $I^*(-f) = -I_*(f)$ and $I_*(-f) = -I^*(f)$. But then f is integrable, so $I^*(-f) = I_*(-f) = -I^*(f)$. Thus $-f$ is also integrable, and $\int_a^b(-f(x))\mathrm{d}x = -\int_a^b f(x)\mathrm{d}x$. $\qquad\square$

Proposition 6.11 (Scaling Integrable Functions). *If $\lambda \in \mathbb{R}$ and $f : [a,b] \to \mathbb{R}$ is Riemann integrable, then so is λf and*

$$\int_a^b \lambda f(x)\mathrm{d}x = \lambda \int_a^b f(x)\mathrm{d}x.$$

Proof. Without loss of generality[25], we may assume that $\lambda \geq 0$. Then if $\mathcal{D}$ is a dissection of $[a, b]$ we have

$$S_{\mathcal{D}}(\lambda f) = \sum_{j=1}^{n}(x_j - x_{j-1}) \sup_{x \in [x_{j-1}, x_j]} \lambda f(x) = \lambda S_{\mathcal{D}}(f),$$

and similarly $s_{\mathcal{D}}(\lambda f) = \lambda s_{\mathcal{D}}(f)$. Then $I^*(\lambda f) = \lambda I^*(f)$ and $I_*(\lambda f) = \lambda I_*(f)$. Since f is integrable, we then have $I^*(\lambda f) = I_*(\lambda f) = \lambda I^*(f)$, so λf is integrable and $\int_a^b \lambda f(x)\mathrm{d}x = \lambda \int_a^b f(x)\mathrm{d}x$. $\square$

To summarise these results, if $f, g : [a, b] \to \mathbb{R}$ are integrable and $\lambda, \mu \in \mathbb{R}$ then $\lambda f + \mu g$ is integrable and

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

Let's now turn our attention[26] to another set of integral properties which are still natural but probably ever-so-slightly less familiar: integral inequalities. The first result is one you'd expect to be true.

Proposition 6.12 (Basic Integral Inequality). *If $f, g : [a, b] \to \mathbb{R}$ are Riemann integrable and $f(x) \leq g(x)$ for all $x \in [a, b]$, then*

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

Proof. If $f \leq g$ then for all dissections $\mathcal{D}$ we have $S_{\mathcal{D}}(f) \leq S_{\mathcal{D}}(g)$, and hence $\int_a^b f(x)\mathrm{d}x = \inf_{\mathcal{D}} S_{\mathcal{D}}(f) \leq \inf_{\mathcal{D}} S_{\mathcal{D}}(g) = \int_a^b g(x)\mathrm{d}x$. $\square$

The next result shows that $|f|$ is integrable if f is, and also something like a triangle inequality for integrals.

Proposition 6.13 (Integral Triangle Inequality). *If $f : [a, b] \to \mathbb{R}$ is Riemann integrable, then so is $|f|$ and*

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

Proof. We define $f_+(x) = \max\{f(x), 0\}$. We will showing that $f_+(x)$ is integrable. For some interval I we have

$$\sup_{x \in I} f_+(x) = \max \left\{0, \sup_{x \in I} f(x)\right\}, \qquad \inf_{x \in I} f_+(x) = \max \left\{0, \inf_{x \in I} f(x)\right\},$$

---

[25]If $\lambda < 0$, we can consider $|\lambda| f$, then by the previous lemma $-|\lambda| f = \lambda f$ will be integrable.

[26]Don't get too sad – we still have to show that the product of integrable functions is integrable, but we will come back to that once we have shown $|f|$ is integrable

and then considering the cases $\sup_{x \in I} f(x) \geq 0$ and $\sup_{x \in I} f(x) \leq 0$, we find that the following inequality holds:

$$\sup_{x \in I} f_+(x) - \inf_{x \in I} f_+(x) = \max\left\{0, \sup_{x \in I} f(x)\right\} - \max\left\{0, \inf_{x \in I} f(x)\right\}$$
$$\leq \sup_{x \in I} f(x) - \inf_{x \in I} f(x).$$

Then since f is integrable, given $\epsilon > 0$ there is a dissection $\mathcal{D}$ such that

$$S_{\mathcal{D}}(f_+) - s_{\mathcal{D}}(f_+) \leq S_{\mathcal{D}}(f) - s_{\mathcal{D}}(f) < \epsilon,$$

and thus f_+ is integrable. We can write $|f| = 2f_+ - f$, and thus $|f|$ is also integrable.

We also have $-|f| \leq f \leq |f|$, so $\left| \int_a^b f(x)dx \right| \leq \int_a^b |f(x)|dx.$ $\qquad \square$

We can use this property to prove that the product of two integrable functions fg is integrable. To do this, the expansion $2fg = (f+g)^2 - f^2 - g^2$ can be used, so all we need to show is that f^2 is integrable if f is. Also since $f^2 = |f| \cdot |f|$, we can work only with the case that f is positive.

Unlike some of the previous results, we will not write down what the value of this integral will be (there's not really a nice rule). We will only be proving integrability.

Proposition 6.14 (Multiplying Integrable Functions)**.** *If $f, g : [a,b] \to \mathbb{R}$ are Riemann integrable, then so is fg.*

Proof. We will first show that if $f : [a,b] \to \mathbb{R}$ is an integrable function such that $f(x) \geq 0$ for all $x \in [a,b]$, then f^2 is integrable.

Since f is integrable (and hence bounded), we can choose some $K > 0$ such that $|f(x)| \leq K$ for all $x \in [a,b]$. Then given any $\epsilon > 0$, there exists some dissection $\mathcal{D}$ where $S_{\mathcal{D}}(f) - s_{\mathcal{D}}(f) < \epsilon/2K$.

Using $\mathcal{D}$, we define $M_j = \sup_{x \in [x_{j-1}, x_j]} f(x)$ and $m_j = \inf_{x \in [x_{j-1}, x_j]} f(x)$. Then we have

$$S_{\mathcal{D}}(f^2) - s_{\mathcal{D}}(f^2) = \sum_{j=1}^{n} (x_j - x_{j-1})(M_j^2 - m_j^2)$$
$$= \sum_{j=1}^{n} (x_j - x_{j-1})(M_j + m_j)(M_j - m_j)$$
$$\leq 2K(S_{\mathcal{D}}(f) - s_{an\mathcal{D}}(f)) < \epsilon,$$

showing that f^2 is integrable.

Since $f^2 = |f| \cdot |f|$, and $|f| \geq 0$, this argument shows that f^2 is integrable for any integrable function f. Using this and previous properties then shows that $[(f+g)^2 - f^2 - g^2]/2 = fg$ is integrable. $\qquad \square$

6.3 Integrable Functions

At this point we have our definition of integrability, along with some basic properties of integrals. However, we have so far only really seen that $f(x) = x$ and $f(x) = |x|$ are integrable. In this section we will prove some results that will expand this list quite widely.

We have so only looked at bounded functions, and a natural question is whether bounded is a sufficient criterion for integrability. This sadly is not the case, and there is a standard counterexample due to Dirichlet.

Example 6.15 (A Bounded Non-Integrable Function). The function $f : [0, 1] \to \mathbb{R}$ given by

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

is not Riemann integrable.

To see this, note that for any dissection $\mathcal{D}$, since $\mathbb{Q}$ is dense in $\mathbb{R}$, we have $s_{\mathcal{D}}(f) = 0$ and $S_{\mathcal{D}}(f) = 1$.

So let's look at some types of integrable functions. The first is that monotonic functions are integrable, which can be shown by taking an evenly spaced dissection causing the sum to telescope.

Proposition 6.16 (Monotonic Functions are Integrable). *Let $f : [a, b] \to \mathbb{R}$ be monotonic. Then f is Riemann integrable.*

Proof. Without loss of generality, assume that f is increasing. Then given $\epsilon > 0$, for any integer $n > [(b-a)(f(b)-f(a))]/\epsilon$ we can define the dissection $\mathcal{D} = \{x_0, x_1, \dots, x_n\}$ of $[a, b]$, where $x_0 = a$, and $x_j - x_{j-1} = (b-a)/n$. Then we have

$$S_{\mathcal{D}}(f) - s_{\mathcal{D}}(f) = \sum_{j=1}^{n}(x_j - x_{j-1}) \left[\sup_{x \in [x_{j-1}, x_j]} f(x) - \inf_{x \in [x_{j-1}, x_j]} f(x) \right]$$

$$= \sum_{j=1}^{n} \frac{b-a}{n} [f(x_j) - f(x_{j-1})]$$

$$= \frac{(b-a)(f(b) - f(a))}{n} < \epsilon,$$

and thus f is integrable. $\square$

An interesting consequence of this is that if we can write $f(x) = f_1(x) - f_2(x)$ with f_1, f_2 are increasing, then f is integrable. Such functions are known as 'functions of bounded variation'.

The next result we would like to show is that every continuous function is Riemann integrable on some bounded interval. Before jumping into the

proof, let's take a moment to reflect on what might be needed[27]. For some continuous function, we want to get $S_\mathcal{D}(f) - s_\mathcal{D}(f) < \epsilon$ with some dissection. What this comes down to is finding some way to bound

$$\sup_{x \in I} f(x) - \inf_{x \in I} f(x),$$

under some sufficiently small interval I. Now we know that f is continuous, and continuity tells us that there if the interval is of size at most some δ (which can depend on where the interval is) then we can get $|f(x) - f(y)| < \epsilon$. This works for one interval, but we need to construct a dissection, so we would need to make this type of argument work for intervals which cover the whole of $[a, b]$.

What would be really nice is if we could find some δ that would work no matter where the sub-interval in $[a, b]$ is. If this was the case, then the dissection could just be made up of points which are about a distance of δ apart. It turns out that this is always possible, and gives us the notion of *uniform continuity*.[28]

Definition 6.17 (Uniform Continuity). Let $A \subset \mathbb{R}$ and $f : A \to \mathbb{R}$. We say that f is *uniformly continuous* on A if given any $\epsilon > 0$ we can find a $\delta > 0$ such that for $x, y \in A$, whenever $|x - y| < \delta$ we have $|f(x) - f(y)| < \epsilon$.

Note that unlike the previous definition, we have no knowledge of where in the set A the points x and y are, just how far apart they are. It turns out that on closed bounded intervals (like what we have when we talk about integration, at least in this section), continuous functions are always uniformly continuous. The proof of this is quite natural – we use Bolzano-Weierstrass to show that a contradiction to uniform continuity is a contradiction to continuity.

Theorem 6.18 (Continuous Functions are Uniformly Continuous). *Let $f : [a, b] \to \mathbb{R}$ be continuous. Then f is uniformly continuous.*

Proof. Suppose f was not uniformly continuous, that is, there exists some $\epsilon > 0$ such that for all $\delta > 0$ there is some x, y with $|x - y| < \delta$ such that $|f(x) - f(y)| \geq \epsilon$.

Taking $\delta = 1/n$, we can find some sequences $x_n, y_n \in [a, b]$ such that $|x_n - y_n| < 1/n$, and $|f(x_n) - f(y_n)| \geq \epsilon$. Then by Bolzano-Weierstrass, we can find some convergent subsequence $x_{n_j} \to x$ for some $x \in [a, b]$. But then $|x_{n_j} - y_{n_j}| < 1/n_j$ for all j, so we must have $y_{n_j} \to x$ also.

Then $|f(x_{n_j}) - f(y_{n_j})| \geq \epsilon$ for every j, which implies that $f(x_{n_j})$ and $f(y_{n_j})$ cannot converge to the same value. But then by continuity we have

---

[27]This is not the only possible proof! You can prove things without uniform continuity, but this is a more natural (and more slick) argument.

[28]The 'uniform' part of the definition is part of a larger class of definitions that you will come across in later analysis courses.

$f(x_{n_j}) \to f(x)$ and $f(y_{n_j}) \to f(x)$, which is a contradiction. Thus f must be uniformly continuous. $\qquad\square$

With this result, we can prove that continuous functions are integrable in the way we discussed previously.

Theorem 6.19 (Continuous Functions are Integrable). *Let $f : [a,b] \to \mathbb{R}$ be continuous. Then f is Riemann integrable.*

Proof. Given $\epsilon > 0$, since f is uniformly continuous, there is some $\delta > 0$ such that $|f(x) - f(y)| < \epsilon/(b-a)$ whenever $|x - y| < \delta$ and $x, y \in [a,b]$.

Now choose some integer $n \geq (b-a)/\epsilon$, and define the dissection $\mathcal{D} = \{x_0, x_1, \ldots, x_n\}$ with $x_j = a + j(b-a)/n$. Then we have

$$\sup_{x\in[x_{j-1},x_j]} f(x) - \inf_{x\in[x_{j-1},x_j]} f(x) \leq \frac{\epsilon}{(b-a)},$$

for all $1 \leq j \leq n$, and thus

$$S_{\mathcal{D}}(f) - s_{\mathcal{D}}(f) = \sum_{j=1}^{n} (x_j - x_{j-1}) \left[\sup_{x\in[x_{j-1},x_j]} f(x) - \inf_{x\in[x_{j-1},x_j]} f(x) \right]$$

$$\leq \sum_{j=1}^{n} \frac{b-a}{n} \cdot \frac{\epsilon}{b-a} = \epsilon,$$

and thus f is Riemann integrable. $\qquad\square$

This theorem shows that most of the functions we deal with regularly (and all of the ones discussed in the previous chapter) are integrable as well as continuous.

Also, we have so far spent relatively little time on the 'evaluating integrals' side of things, but knowing a function is integrable can help in determining the actual value of the integral. A straightforward example is shown below.

Example 6.20 (Integral of $f(x) = 1$). Consider the function $f : [a,b] \to \mathbb{R}$ defined by $f(x) = 1$. This function is continuous and hence integrable. Then taking the dissection $\mathcal{D} = \{a, b\}$, we have $S_{\mathcal{D}}(f) = s_{\mathcal{D}}(f) = b - a$, so the value of the integral is

$$\int_a^b 1 \mathrm{d}x = b - a.$$

6.4   The Fundamental Theorem of Calculus

With all of that out of the way, we can now prove a central result in the theory of integration: the fundamental theorem of calculus. The fundamental theorem of calculus connects the theory that we built up relating to differentiation and integration, showing that (in some sense) they are 'inverses'.

Theorem 6.21 (Fundamental Theorem of Calculus)**.** *Let* $f : [a, b] \to \mathbb{R}$ *be continuous. For* $x \in [a, b]$, *we define*

$$F(x) = \int_a^x f(x) \mathrm{d}x.$$

Then F *is differentiable, with derivative* $F'(x) = f(x)$ *for every* x.

Proof. For $|h| > 0$, we note that

$$\frac{F(x + h) - F(x)}{h} = \frac{1}{h} \int_x^{x+h} f(t) \mathrm{d}t.$$

Now given $\epsilon > 0$, since f is continuous at x there is some $\delta > 0$ such that when $|x - y| < \delta$ we have $|f(x) - f(y)| < \epsilon$. Hence if $0 < |h| < \delta$ we have

$$\left| \frac{1}{h} \int_x^{x+h} f(t) \mathrm{d}t - f(x) \right| = \left| \frac{1}{h} \int_x^{x+h} f(t) - f(x) \mathrm{d}t \right|$$

$$\leq \frac{1}{|h|} \left| \int_x^{x+h} |f(t) - f(x)| \right|$$

$$\leq \frac{1}{|h|} \cdot \epsilon |h| = \epsilon,$$

hence $\lim_{h \to 0} \frac{F(x+h) - F(x)}{h} = f(x)$, as required. $\square$

Another way to view this theorem (which is quite useful in evaluating integrals) is the corollary below.

Corollary 6.22 (Fundamental Theorem of Calculus – Again)**.** *Suppose that* $f : [a, b] \to \mathbb{R}$ *has a continuous derivative. Then*

$$\int_a^b f'(t) \mathrm{d}t = f(b) - f(a).$$

Proof. Define $g(x) = \int_a^x f'(t) \mathrm{d}t$. Then by the fundamental theorem of calculus, $g'(x) = f'(x)$. Also $g(a) = 0$.

Let $h(x) = g(x) - f(x)$. Then $h'(x) = 0$, so by the mean value theorem, $h(x)$ is constant. Taking the value at a, we then have $h(a) = -f(a) = g(x) - f(x) = -f(a)$ for all x. So if $x = b$, this becomes $g(b) = f(b) - f(a)$, which is our desired result. $\square$

One of the most useful applications of the fundamental theorem of calculus is that we can translate results about derivatives to results about integrals. This works in both coming up with the integrals of functions (by thinking about which functions result in a given function when differentiated), but also in the development of general methods for evaluating integrals. Two results we will look at are integration by parts, which comes from the product rule of differentiation, and integration by substitution, which comes from the chain rule.

Corollary 6.23 (Integration By Parts)**.** *Suppose that the derivatives f' and g' exist and are continuous on $[a, b]$. Then*

$$\int_a^b f'(x)g(x)\mathrm{d}x = f(b)g(b) - f(a)g(a) - \int_a^b f(x)g'(x)\mathrm{d}x.$$

Proof. By the product rule $(fg)' = f'g + fg'$. Then by the fundamental theorem of calculus we have $f(b)g(b) - f(a)g(a) = \int_a^b f'(x)g(x)\mathrm{d}x + \int_a^b f(x)g'(x)\mathrm{d}x$. □

Corollary 6.24 (Integration By Substitution)**.** *Let $g : [\alpha, \beta] \to [a, b]$ with $g(\alpha) = a$ and $g(\beta) = b$, let g' exist and be continuous on $[\alpha, \beta]$. Also let $f : [a, b] \to \mathbb{R}$ be continuous. Then*

$$\int_a^b f(x)\mathrm{d}x = \int_\alpha^\beta f(g(t))g'(t)\mathrm{d}t.$$

Proof. Set $F(x) = \int_a^x f(t)\mathrm{d}t$, and let $h(t) = F(g(t))$, which is defined since g takes values in $[a, b]$. Then by the fundamental theorem of calculus we have

$$\int_\alpha^\beta f(g(t))g'(t)\mathrm{d}t = \int_\alpha^\beta F'(g(t))g'(t)\mathrm{d}t$$

Then by the chain rule, we have

$$\int_\alpha^\beta F'(g(t))g'(t)\mathrm{d}t = \int_\alpha^\beta h'(t)\mathrm{d}t = h(\beta) - h(\alpha) = F(b) - F(a) = \int_a^b f(x)\mathrm{d}x.$$

□

6.5 Taylor's Theorem Again

With some results now built up around integration, we are going to spend the next two sections discussing two applications of integration to topics we have discussed earlier. The first will be revisiting Taylor's theorem, and after that we will look back at the convergence of infinite series.

A nice application of integration by parts (and by extension the fundamental theorem of calculus) is in proving Taylor's theorem. You may recall from our discussion of differentiation that the proof of Taylor's theorem with Lagrange remainder was somewhat cumbersome. Using integration, we can derive another form of Taylor's theorem (this time with the 'integral remainder'). This other form can then even be used to obtain Lagrange's form of the remainder, under some slightly stronger assumptions about our function[29].

---

[29]This is why we don't just throw out the old proof! We need a stronger continuity condition than before - specifically $f^{(n)}$ being continuous, not just existing.

Theorem 6.25 (Taylor's Theorem with Integral Remainder). *Let f be n-times continuously differentiable on $[a, b]$. Then we have*

$$f(b) = f(a) + f'(a)(b-a) + \cdots + \frac{f^{(n-1)}(a)}{(n-1)!}(b-a)^{n-1} + \int_a^b \frac{f^{(n)}(t)}{(n-1)!}(b-t)^{n-1} \mathrm{d}t$$

Proof. We do induction on n. For $n = 1$, the result becomes $f(b) - f(a) = \int_a^b f'(t)\mathrm{d}t$, which is true by the fundamental theorem of calculus.

Now if f is $(n+1)$-times continuously differentiable, integrating by parts we have

$$\int_a^b \frac{f^{(n)}(t)}{(n-1)!}(b-t)^{n-1}\mathrm{d}t = \left[-\frac{f^n(t)}{n!}(b-t)^n \right]_a^b + \int_a^b \frac{f^{(n+1)}(t)}{n!}(b-t)^n\mathrm{d}t$$

$$= \frac{f^{(n)}(a)}{n!}(b-a)^n + \int_a^b \frac{f^{(n+1)}(t)}{n!}(b-t)^n\mathrm{d}t.$$

Thus if the result is true for n, it is also true for $n+1$, so we are done by induction. $\qquad\square$

With this result in hand, we can try and re-obtain the Lagrange form of the remainder. Looking back at the form of this, we really want to some way take the $f^{(n)}(t)$ term 'out of the integral', turning it into some scale factor $f^{(n)}(c)$, for some $c \in (a, b)$.

The way we worked with this type of idea before was using the mean value theorem, so we are going to prove an analogue of the mean value theorem for integrals. To do this we will (naturally) use the mean value theorem along with the fundamental theorem of calculus.

Lemma 6.26 (Integral Mean Value Theorem). *Let $f, g : [a, b] \to \mathbb{R}$ be continuous with $g(x) \neq 0$ for all $x \in (a, b)$. Then there exists $c \in (a, b)$ such that*

$$\int_a^b f(x)g(x)\mathrm{d}x = f(c)\int_a^b g(x)\mathrm{d}x.$$

Proof. Let $F(x) = \int_a^x f(t)g(t)\mathrm{d}t$ and $G(x) = \int_a^x g(t)\mathrm{d}t$. By Cauchy's mean value theorem, there exists some $c \in (a, b)$ such that

$$(F(b) - F(a))G'(c) = F'(c)(G(b) - G(a)),$$

which by the fundamental theorem of calculus implies

$$\left(\int_a^b f(t)g(t)\mathrm{d}t \right) g(c) = f(c)g(c)\int_a^b g(t)\mathrm{d}t.$$

Then sine $g(c) \neq 0$, the result follows. $\qquad\square$

And now we can write down a weaker form of Taylor's theorem with Lagrange remainder.

Theorem 6.27 (Taylor's Theorem with Lagrange Remainder – Weaker).
Let f be n-times continuously differentiable on $[a, b]$. Then we have

$$f(b) = f(a) + f'(a)(b - a) + \cdots + \frac{f^{(n-1)}(a)}{(n-1)!}(b - a)^{n-1} + \frac{f^{(n)}(c)}{n!}(b - a)^n,$$

for some $c \in (a, b)$.

Proof. By Taylor's theorem with integral remainder, we have

$$f(b) = f(a) + f'(a)(b-a) + \cdots + \frac{f^{(n-1)}(a)}{(n-1)!}(b-a)^{n-1} + \int_a^b \frac{f^{(n)}(t)}{(n-1)!}(b-t)^{n-1}\mathrm{d}t.$$

Then the integral mean value theorem gives us that

$$\int_a^b \frac{f^{(n)}(t)}{(n-1)!}(b-t)^{n-1}\mathrm{d}t = f^{(n)}(c)\int_a^b \frac{(b-t)^{n-1}}{(n-1)!}\mathrm{d}t = \frac{f^{(n)}(c)}{n!}(b-a)^n,$$

for some $c \in (a, b)$, giving our result. $\square$

Now, to just emphasize one last time: this version of the result imposes *an extra continuity condition*, and *does not replace the result and proof we had earlier.* That said, it's still a perfectly good result on its own.

6.6   Improper Integrals & The Integral Test

So far our definition of Riemann integration applies only to functions that are bounded over some closed interval. In this subsection we are going to try and extend this definition in a natural way to discuss the integration of functions that are not necessarily bounded and integration over an interval that may not be closed (and may be infinite).

Definition 6.28 (Improper Integrals). Suppose $f : [a, b) \to \mathbb{R}$ is integrable (and bounded) on every interval $[a, R]$, where $b \in \mathbb{R} \cup \{\infty\}$, and

$$\int_a^R f(x)\mathrm{d}x \to \ell \quad \text{as} \quad R \to b.$$

Then we say that the *improper integral* $\int_a^b f(x)\mathrm{d}x$ converges, and that its value is ℓ. Otherwise, we say that it diverges.

Let's have a look at some examples.

Example 6.29 (Integral of $1/x^k$ from 1 to ∞). We will show that the improper integral $\int_1^\infty \frac{1}{x^k}\mathrm{d}x$ converges if and only if $k > 1$.

If $k \neq 1$, then considering the integral of $1/x^k$ over $[1, R]$, we have

$$\int_1^R \frac{1}{x^k}\mathrm{d}x = \frac{x^{1-k}}{1-k}\Big|_1^R$$

$$= \frac{R^{1-k} - 1}{1 - k},$$

which tends to a finite limit as $R \to \infty$, namely $-1/(1 - k)$, if and only if $k > 1$.

Otherwise, if $k = 1$, then over $[1, R]$ we have

$$\int_1^R \frac{1}{x} \mathrm{d}x = \log R,$$

which does not tend to a finite limit as $R \to \infty$. Thus the integral converges if and only if $k > 1$.

Example 6.30 (Integral of $1/\sqrt{x}$ from 0 to 1). We will show that the improper integral $\int_0^1 \frac{1}{\sqrt{x}} \mathrm{d}x$ converges.

For any $\delta > 0$, we have

$$\int_\delta^1 \frac{1}{\sqrt{x}} \mathrm{d}x = 2\sqrt{x}\big|_\delta^1 = 2 - 2\sqrt{\delta} \to 2 \quad \text{as} \quad \delta \to 0,$$

and thus the improper integral converges to 2.

Note the example above shows that it is possible to have an improper integral which converges when the function is not bounded on the interval being integrated over.[30]

The definition of improper integrals also gives us a sensible way to say if $\int_{-\infty}^\infty f(x)\mathrm{d}x$ converges. If for some a we have $\int_{-\infty}^a f(x)\mathrm{d}x = \ell_1$ and $\int_a^\infty f(x)\mathrm{d}x = \ell_2$, then we say that $\int_{-\infty}^\infty f(x)\mathrm{d}x = \ell_1 + \ell_2$.

Remark (Warning). This is a strictly stronger notion than saying that $\int_{-R}^R f(x)\mathrm{d}x$ converges to some limit as $R \to \infty$.

It is with improper integrals that we get the last application of integration to our previously discussed topics: infinite series.

Theorem 6.31 (The Integral Comparison Test). *If $f : [1, \infty) \to \mathbb{R}$ is a non-negative decreasing function with $f(x) \to 0$ as $x \to \infty$ then*

$$\sum_1^\infty f(n) \quad \text{and} \quad \int_1^\infty f(x)\mathrm{d}x$$

either both diverge or both converge.

---

[30]It is also possible to have a function f that is unbounded on the interval $[1, \infty)$, and yet $\int_1^\infty f(x)\mathrm{d}x$ converges. If you haven't seen this before, try coming up with an example (as a hint, one way to do it is by drawing triangles in a somewhat clever way).

Proof. Since f is decreasing and is non-negative, we have for $n \geq 2$ that

$$\int_n^{n+1} f(x)\mathrm{d}x \leq f(n) \leq \int_{n-1}^n f(x)\mathrm{d}x.$$

That is, since $\int_1^2 f(x)\mathrm{d}x \leq f(1)$, we have

$$\int_1^{N+1} f(x)\mathrm{d}x \leq \sum_{n=1}^N f(x) \leq \int_1^N f(x)\mathrm{d}x + f(1),$$

and then our result follows as $N \to \infty$. $\qquad\square$

The integral comparison test is useful because in many cases it is easier to evaluate an integral than it is to evaluate a sum (of course this isn't always true, and there's many examples in both directions). This test also gives us a way to determine the convergence of sums that previously required tools such as Cauchy's condensation test[31]. An example is shown below.

Example 6.32 (Using the Integral Test)**.** We will prove that $\sum_{n=2}^\infty \frac{\log n}{n^2}$ converges using the integral test.

We compute using integration by parts that

$$\int_2^R \frac{\log x}{x^2}\mathrm{d}x = -\left.\frac{\log x}{x}\right|_2^R + \int_2^R \frac{1}{x^2}\mathrm{d}x$$
$$= -\frac{\log R}{R} + \frac{\log 2}{2} - \frac{1}{R} + \frac{1}{2}.$$

This converges to $(1 + \log 2)/2$ as $R \to \infty$, and thus $\int_2^\infty \log x/x^2\mathrm{d}x$ and $\sum_{n=2}^\infty \log n/n^2$ both converge by the integral test.

---

[31]Though it can be rather cumbersome to evaluate the integrals in certain cases, in which case reaching for Cauchy's condensation test first might be still helpful. Using the integral test also is a straightforward way to derive Cauchy's condensation test