Analysis I

Carl Turner

November 21, 2010

Abstract

These are notes for an undergraduate course on analysis; please send corrections, suggestions and notes to courses@suchideas.com The author's homepage for all courses may be found on his website at SuchIdeas.com, which is where updated and corrected versions of these notes can also be found.

The course materials are licensed under a permissive Creative Commons license: *Attribution-NonCommercial-ShareAlike 3.0 Unported* (see the CC website for more details).

Thanks go to Professor G. Paternain for allowing me to use his Analysis I course (Lent 2010) as the basis for these notes.

Contents

1 Limits and Convergence **3**
- 1.1 Fundamental Properties of the Real Line . . . 3
- 1.2 Cauchy Sequences . . . 5
- 1.3 Series . . . 6
- 1.4 Sequences of Non-Negative Terms . . . 7
- 1.5 Alternating Series and Absolute Convergence . . . 10

2 Continuity **13**
- 2.1 Definitions and Basic Properties . . . 13
- 2.2 Intermediate Values, Extreme Values and Inverses . . . 15

3 Differentiability **18**
- 3.1 Definitions and Basic Properties . . . 18
- 3.2 The Mean Value Theorem and Taylor's Theorem . . . 20
- 3.3 Taylor Series and Binomial Series . . . 25
- 3.4 * Comments on Complex Differentiability . . . 26

4 Power Series **28**
- 4.1 Fundamental Properties . . . 28
- 4.2 Standard Functions . . . 32
  - 4.2.1 Exponentials and logarithms . . . 32

4.2.2 Trigonometric functions . . . 35
4.2.3 Hyperbolic functions . . . 37

5 Integration **38**
5.1 Definitions and Key Properties . . . 38
5.2 Computation of Integrals . . . 44
5.3 Mean Values and Taylor's Theorem Again, Improper Integrals and the Integral Test . . . 46
5.3.1 The mean value theorem and Taylor's theorem . . . 46
5.3.2 Infinite (improper) integrals and the integral test . . . 48

6 * Riemann Series Theorem **52**

7 * Measure Theory **54**

Books

- M. Spivak, *Calculus* (1994)
- B. R. Gelbaum & J. M. H. Olmsted, *Counterexamples in Analysis* (2003)

1 Limits and Convergence

We begin with a comprehensive review of material from courses on the reals $\mathbb{R}$, including all properties of limits.

Definition 1.1. Let $a_n \in \mathbb{R}$. Then we say a_n *tends to a limit* a (written $a_n \to a$ or more explicitly $\lim_{n\to\infty} a_n = a$) as $n \to \infty$ if, given any $\epsilon > 0$, there exists some $N \in \mathbb{N}$ such that $|a_n - a| < \epsilon$ for all $n \geq N$.

This is essentially the statement that, given some small neighbourhood of a, *beyond some point*, the sequence *never* leaves that neighbourhood.

It is very important that we see that N is a function of ϵ, i.e. $N = N(\epsilon)$.

1.1 Fundamental Properties of the Real Line

We begin first by considering exactly what we mean when we talk about the real line. In essence, the following axiom expresses the key difference between the real numbers $\mathbb{R}$ and the rationals $\mathbb{Q}$.

Fact 1.2. *Suppose $a_n \in \mathbb{R}$ is an increasing, and that there is some $A \in \mathbb{R}$ upper bound on the sequence such that $a_n \leq A$ for all n. Then there is a number $a \in \mathbb{R}$ such that $a_n \to a$ as $n \to \infty$.*

That is, any increasing sequence bounded above converges to some limit in $\mathbb{R}$, so that $\mathbb{R}$ is *closed* under limit taking. In fact, this shows $\mathbb{R}$ is a *closed set.* Note we could easily restate the axiom in terms of lower bounds too, by simply negating the above.

Remark. The above axiom is equivalent to the fact that every non-empty set of real numbers bounded above has a *supremum*, or least upper bound.

Lemma 1.3. *A few fundamental properties of limits:*

- *(i)* *The limit is unique.*
- *(ii)* *If $a_n \to a$ as $n \to \infty$ and $n_1 < n_2 < \cdots$ then $a_{n_j} \to a$ as $j \to \infty$; that is, subsequences always converge to the same limit.*
- *(iii)* *If $a_n = c$ is constant, $a_n \to c$.*
- *(iv)* *If $a_n \to a$ and $b_n \to b$, then $a_n + b_n \to a + b$.*
- *(v)* *If $a_n \to a$ and $b_n \to b$, then $a_n b_n \to ab$.*
- *(vi)* *If $a_n \to a$, $a_n \neq 0$ $\forall n$ and $a \neq 0$, then $\frac{1}{a_n} \to \frac{1}{a}$.*
- *(vii)* *If $a_n \leq A$ $\forall n$ and $a_n \to a$ then $a \leq A$.*

Proof. Only proofs of 1, 2 and 5 follow; the remainder are left as an exercise.

Uniqueness. $a_n \to a$ and $a_n \to b$ means that $\forall \epsilon > 0$ we have N_1 and N_2 such that $|a_n - a| < \epsilon$ for $n > N_1$ and $|a_n - b| < \epsilon$ for $n > N_2$.

But then by the triangle inequality, $|a-b| \leq |a-a_n| + |b-a_n| < 2\epsilon$ for $n \geq \max\{N_1, N_2\}$. Hence $0 \leq |a-b| < x$ for all $x > 0$, so $|a-b| = 0$.

Subsequences. If $a_n \to a$, we have some $N(\epsilon)$ such that $|a_n - a| < \epsilon$ for $n \geq N(\epsilon)$. Then noting $n_j \geq j$ (which is clear from, say, induction), we see $\left|a_{n_j} - a\right| < \epsilon$ for $j \geq N$, so we are done.

Limits of products. Note that by the triangle inequality,

$$\begin{aligned} |a_n b_n - ab| &\leq |a_n b_n - a_n b| + |a_n b - ab| \\ &= |a_n| |b_n - b| + |b| |a_n - a| \\ &\leq (1 + |a|) |b_n - b| + |b| |a_n - a| \end{aligned}$$

for $n \geq N_1(1)$ for the last step to hold. Now letting $n \geq \max\{N_1(\epsilon), N_2(\epsilon), N_1(1)\}$ we have $|a_n b_n - ab| \leq (1+|a|)\epsilon + |b|\epsilon = \epsilon(|a| + |b| + 1)$ which can be made arbitrarily small since the bracketed term is constant, so we are done. ☐

We now prove a very simple limit which is a very useful basis for many other proofs.

Lemma 1.4. $\frac{1}{n} \to 0$ *as* $n \to \infty$.

Proof. $a_n = \frac{1}{n}$ is decreasing and is bounded below by 0. Hence this converges by the fundamental axiom, say to $a \in \mathbb{R}$. To show $a = 0$, we could progress in a number of ways. A clever method which makes use of the above properties is to note the subsequence $a_{2n} = \frac{1}{2} \cdot \frac{1}{n}$ must converge to a also, but clearly $a_{2n} = \frac{1}{2} a_n \to \frac{1}{2} a$ (we know this since this is the product of the constant sequence $b_n = \frac{1}{2}$ and a_n), so $a = \frac{1}{2} a$ as limits are unique and hence $a = 0$. ☐

Example 1.5. Find the limit of $n^{1/n} = \sqrt[n]{n}$ as $n \to \infty$.

Here are two approaches:

- Write $\sqrt[n]{n} = 1 + \delta_n$. Then $n = (1+\delta_n)^n > \frac{n(n-1)}{2}\delta_n^2$ using a binomial expansion; but then $\frac{2}{n-1} > \delta_n^2$ so $0 < \delta_n < \sqrt{\frac{2}{n-1}} \to 0$ and hence $\delta_n \to 0$.
- Assuming standard properties of e and log, we note $\sqrt[n]{n} = e^{\log n/n} \to e^0 = 1$.

Remark. If we were instead considering sequences of complex numbers $a_n \in \mathbb{C}$ we give essentially the same treatment of the limit (the $|x|$ terms are generalized from absolute value to the modulus of the complex number x; the triangle inequality still holds). Then all properties proved above still hold, except the last, since the statement is meaningless (there is no order on $\mathbb{C}$). In fact, many properties of real sequences and series are applicable to complex numbers; frequently, considering the sequences in $\mathbb{R}$ given by the real and imaginary parts of the complex numbers is a useful approach.

We now prove a theorem which is a very useful tool in proving various results in fundamental analysis.

Theorem 1.6 (The Bolzano-Weierstrass Theorem). *If the sequence $x_n \in \mathbb{R}$ is bounded by K, $|x_n| \leq K$ $\forall n$, then there is some subsequence given by $n_1 < n_2 < \cdots$ and some number $x \in \mathbb{R}$ such that $x_{n_j} \to x$ as $j \to \infty$. That is, every bounded sequence has a convergent subsequence.*

Remark. We cannot make any guarantees about the uniqueness of x; simply consider $x_n = (-1)^n$.

To see how to set about proving this, try to form an intuition of why this should be. Imagine trying to construct a sequence for which this did not hold. We would have to generate an infinite sequence where no infinite collection of points were concentrated around any one point in the bounded interval. But once we have placed sufficiently many points, we will be forced to place the next one within any distance ϵ of a point already present. Another way of looking at this is that, regardless of how we try to avoid stacking many in the same area, infinitely many must end up in some 'band' of pre-specified finite size ϵ, by a sort of infinite pigeon-hole principle. So if choose progressively narrow bands, then we can always find a 'more tightly squeezed' subsequence. The general idea of this approach is referred to as *bisection* or informally *lion-hunting*. It is formalized in the proof shown.

Proof. Let out initial interval $[a_0, b_0] = [-K, K]$, and let $c = \frac{a_0+b_0}{2}$ be its midpoint.

Now either infinitely many values lie in $[a_0, c]$ or infinitely many values lie in $[c, b_0]$ (or both).

In the first case, set $a_1 = a_0$ and $b_1 = c$, so $[a_1, b_1] = [a_0, c]$ contains infinitely many points. Otherwise, $[a_1, b_1] = [c, b_0]$.

Then proceed indefinitely to obtain sequences a_n and b_n with $a_{n-1} \leq a_n \leq b_n \leq b_{n-1}$, and $b_n - a_n = \frac{b_{n-1}-a_{n-1}}{2}$. But then these are bounded, monotonic sequences, and so $a_n \to a$ and $b_n \to b$. But taking limits of the relationship of the interval widths, $b - a = \frac{b-a}{2}$ so $a = b$.

Hence, since infinitely many x_m lie in each interval $[a_n, b_n]$, we can indefinitely select $n_j > n_{j-1}$ such that $x_{n_j} \in [a_j, b_j]$ holds $\forall j$ (only finitely many terms in the sequence are excluded by the $n_j > n_{j-1}$ condition).

Then $a_j \leq x_{n_j} \leq b_j$ holds for all j, so since $a_j, b_j \to a$, we get $x_{n_j} \to a$. □

1.2 Cauchy Sequences

Definition 1.7. A sequences $a_n \in \mathbb{R}$ is a *Cauchy* sequence if, given $\epsilon > 0$, there is some N such that $\forall n, m \geq N$, $|a_n - a_m| < \epsilon$.

That is, a sequence is Cauchy if there is some point in the sequence after which no two elements differ by more than ϵ. Note $N = N(\epsilon)$ again.

Theorem 1.8. *A sequence is convergent $\iff$ it is Cauchy.*

Proof. Showing a convergent sequence is Cauchy is fairly easy; let $a_n \in \mathbb{R}$ have limit a. Then we have $|a_n - a| < \epsilon$ for $n \geq N$. But $|a_m - a_n| \leq |a_m - a| + |a_n - a| < 2\epsilon$ if $m, n \geq N$, so we are done.

Now for the converse, we will make use of the Bolzano-Weierstrass theorem. So first, we must show a Cauchy sequence a_n is bounded.

Take $\epsilon = 1$, so $|a_n - a_m| < 1$ for all $n \geq N$. Hence, for all $n \geq N$ we have

$$|a_n| \leq |a_n - a_N| + |a_N| \leq 1 + |a_N|$$

Then $K = \max\left(\{1 + |a_N|\} \cup \{|a_i| : 1 \leq i \leq N - 1\}\right)$ is a bound for a_n.

Now by Bolzano-Weierstrass, a_n has a convergent subsequence $a_{n_j} \to a$. But then for $n \geq N(\epsilon)$

$$|a_n - a| \leq \left|a_n - a_{n_j}\right| + \left|a_{n_j} - a\right|$$

so finding some n_j large enough that both terms on the right hand side are less than ϵ, we are done. Note that this is possible because a_n is Cauchy, and $a_{n_j} \to a$ respectively. □

Therefore, for real sequences of real numbers, convergence and the Cauchy property are *equivalent*. This is the *General Principle of Convergence*. Since the real numbers have the property that the limit of any Cauchy sequence is a real number, the real numbers $\mathbb{R}$ are a *complete* space (this is analogous to the statement that $\mathbb{R}$ is *closed* above).

1.3 Series

The concept of infinite summation is key to much of both discrete and continuous maths, and a careful study of them is crucial to understanding how power series expansions of functions work.

Definition 1.9. Let a_n lie in $\mathbb{R}$ or $\mathbb{C}$. We say the *series* $\sum_{j=1}^{\infty} a_j$ *converges* if the sequence of partial sums $s_N = \sum_{n=1}^{N} a_j$ converges to s as $N \to \infty$.

In the case that it does, we write $\sum_{j=1}^{\infty} a_j = s$. If s_N does not converge, we say the series is *divergent*.

Note that in essence studying series is simply studying the sequences given by partial sums.

Lemma 1.10.

(i) *If $\sum_{j=1}^{\infty} a_j$ converges, we must have $a_j \to 0$ as $j \to \infty$.*

(ii) *If $\sum_{j=1}^{\infty} a_j$ and $\sum_{j=1}^{\infty} b_j$ converges, then so does $\sum_{j=1}^{\infty} (\lambda a_j + \mu b_j)$ $\forall \lambda, \mu \in \mathbb{C}$.*

(iii) *Suppose there is some N such that $a_j = b_j$ for $j \geq N$; then $\sum_{j=1}^{\infty} a_j$ and $\sum_{j=1}^{\infty} b_j$ both converge or diverge; that is, initial terms do not affect convergence properties.*

Proof. We only prove the first two parts; the last is left as an easy exercise.

Firstly, note that all convergent sequences are Cauchy; so the partial sums must satisfy $|s_N - s_M| < \epsilon$ for all sufficiently large N, M. Let $M = N - 1$; then we see $|a_N| < \epsilon$ for sufficiently large N, so $a_j \to 0$.

For the second part, we have

$$\begin{aligned} s_N &= \sum_{j=1}^{N} (\lambda a_j + \mu b_j) \\ &= \lambda \sum_{j=1}^{N} a_j + \mu \sum_{j=1}^{N} b_j \\ &= \lambda c_N + \mu d_N \end{aligned}$$

where c_N and d_N are the convergent sums of a_j and b_j. By the properties of limits, we have $\sum_{j=1}^{\infty} (\lambda a_j + \mu b_j) = \lambda \sum_{j=1}^{\infty} a_j + \mu \sum_{j=1}^{\infty} b_j$ being convergent. $\square$

Lemma 1.11. *The geometric series given by summing $a_n = r^n$ is convergent for all $r \in \mathbb{C}$ such that $|r| < 1$, and divergent for $|r| \geq 1$.*

Proof. The convergence is easily seen by writing $\sum_1^N r^n = 1 + r + r^2 + \cdots + r^N = \frac{1-r^{N+1}}{1-r} \to \frac{1}{1-r}$.

The divergence is equally clear, as $r^n \not\to 0$. $\square$

1.4 Sequences of Non-Negative Terms

When we restrict ourselves to $a_n \geq 0$, we obtain several useful criteria for convergence.

Theorem 1.12 (Comparison test). *Suppose $0 \leq b_n \leq a_n$ for all n. Then if $\sum a_n$ converges, so does $\sum b_n$.*

Proof. Write $s_N = \sum^N a_n$ and $t_N = \sum^N b_n$ for the partial sums; then both are increasing, and obviously $t_N \leq s_N$. Then s_N is bounded above since it converges. Thus t_N is strictly increasing and also bounded above, and hence converges. $\square$

Remark. In fact, since initial terms do not affect convergence, we only need $0 \leq b_n \leq a_n$ for all $n \geq N$ for some N.

Theorem 1.13 (Ratio test). *Suppose $a_n > 0$ is a strictly positive sequence such that $\frac{a_{n+1}}{a_n} \to l$. Then*

- *if $l < 1$, $\sum a_n$ converges;*
- *if $l > 1$, $\sum a_n$ diverges.*

Remark. Note that the test is inconclusive if $l = 1$; see below for examples of both convergent and divergent series with this property.

Proof. Suppose $l < 1$; then take $r \in (l, 1)$. Then there is some N such that for all $n \geq N$, $\frac{a_{n+1}}{a_n} < r$, so that

$$\begin{aligned} a_n &= \frac{a_n}{a_{n-1}} \cdot \frac{a_{n-1}}{a_{n-2}} \cdot \dots \cdot \frac{a_{N+1}}{a_N} a_N \\ &< r^{n-N} a_N \\ &= r^n \underbrace{\left(\frac{a_N}{r^N}\right)}_{k} \end{aligned}$$

But since $r < 1$, we know $0 < \sum_{n=N}^{\infty} a_n < \sum_{n=N}^{\infty} kr^n$ where the upper bound converges (it is a multiple of a geometric series with $|r| < 1$); hence by the comparison test, $\sum a_n$ converges. (Note $a_n > 0$ so the partial sums form an increasing sequence.) □

Now if $l > 1$, there is some N such that $\frac{a_{n+1}}{a_n} > 1$ for all $n \geq N$. But then $a_n > a_N$ for all $n > N$. Hence $a_n \not\to 0$ and the series is divergent.

Theorem 1.14 (Root test). *Suppose $a_n \geq 0$ is a non-negative sequence. Then if $\sqrt[n]{a_n} \to l$,*

- *if $l < 1$, $\sum a_n$ converges;*
- *if $l > 1$, $\sum a_n$ diverges;*

Remark. It is possible, if $l = 1$ and the limit is exceeded infinitely many times, to note that then $a_n > 1$ infinitely many times, so $a_n \not\to 0$ and the series diverges, but in general this is not a useful case.

Proof. As before, we compare the series with a suitable geometric series.

If $l < 1$, we know that for all $n \geq N$ we have $\sqrt[n]{a_n} < r$ for any $r \in (l, 1)$. Hence $a_n < r^n < 1$; then since $\sum_{n=N}^{\infty} r^n$ converges, by comparison we are done.

If $l > 1$, then $a_n \not\to 0$ so the series diverges. □

Example 1.15. We take two examples to show the ratio test and root test cannot give conclusive information for $l = 1$.

- $\sum \frac{1}{n}$ diverges, as may be seen by noting

$$\begin{aligned}(1)+\left(\frac{1}{2}\right)+\left(\frac{1}{3}+\frac{1}{4}\right)+\left(\frac{1}{5}+\cdots+\frac{1}{8}\right)+\cdots &> (1)+\left(\frac{1}{2}\right)+\left(\frac{1}{4}+\frac{1}{4}\right)+\left(\frac{1}{8}+\cdots+\frac{1}{8}\right)+\cdots \\ &= 1+\frac{1}{2}+\frac{1}{2}+\frac{1}{2}+\cdots\end{aligned}$$

  Now in the ratio test, we get $\frac{n}{n+1} \to 1$, and in the root test we get $\left(\frac{1}{n}\right)^{1/n} \to 1$ (see Example 1.5).

- Consider now $\sum \frac{1}{n^2}$. We have $\frac{1}{n^2} < \frac{1}{n(n-1)} = \frac{1}{n-1} - \frac{1}{n}$ for $n > 1$. Hence $\sum_2^N \frac{1}{n^2} < \sum_2^N \left(\frac{1}{n-1} - \frac{1}{n}\right) = \frac{1}{1} - \frac{1}{N}$, so $\sum_2^N \frac{1}{n^2} < 2$, and the series converges.

  Now in the ratio test, we get $\frac{n^2}{(n+1)^2} = \left(\frac{n}{n+1}\right)^2 \to 1$, and in the root test we have $\left(\frac{1}{n^2}\right)^{1/n} = \left(\frac{1}{n^{1/n}}\right)^2 \to 1$.

This shows the importance of considering borderline cases for convergence tests; they should almost always be treated separately. We finish with an example to illustrate some simple uses of the ratio and root tests:

Example 1.16.

- $\sum_1^\infty \frac{n^{2010}}{2^n}$. The ratio test gives $\left(\frac{n+1}{n}\right)^{2010} 2^{n-(n+1)} \to \frac{1}{2} < 1$, so the series converges (note all terms are positive).
- $\sum_1^\infty \left(\frac{1}{\log n}\right)^n$. The root test gives $a_n^{1/n} = \frac{1}{\log n} \to 0 < 1$, so the series converges.

Theorem 1.17 (Cauchy condensation test)**.** *Let a_n be a decreasing sequence of positive terms. Then $\sum a_n$ converges iff $\sum 2^n a_{2^n}$ does.*

Proof. This proof has echoes of the proof that the harmonic series diverges given above; note that $a_{2^k} \le a_{2^{k-1}+i} \le a_{2^{k-1}}$ where $i \in \left[1, 2^{k-1}\right]$.

Suppose initially that $\sum a_n$ converges. Then

$$\begin{aligned}2^{n-1}a_{2^n} &= a_{2^n} + a_{2^n} + \cdots + a_{2^n} \\ &\le a_{2^{n-1}+1} + a_{2^{n-1}+2} + \cdots + a_{2^{n-1}+2^{n-1}-1} + a_{2^n} \\ &= \sum_{m=2^{n-1}+1}^{2^n} a_m\end{aligned}$$

from which it follows that

$$\begin{aligned}\sum_{n=1}^{N} 2^{n-1} a_{2^n} &\leq \sum_{n=1}^{N} \sum_{m=2^{n-1}+1}^{2^n} a_m \\ &= \sum_{m=2}^{2^N} a_m\end{aligned}$$

Thus $\sum_{n=1}^{N} 2^n a_{2^n}$ is bounded above by the upper bound of $2 \sum a_m$, and since it is increasing it converges.

Now conversely, assume $\sum_{n=1}^{N} 2^n a_{2^n}$ converges. Then similarly

$$\begin{aligned}\sum_{m=2^{n-1}+1}^{2^n} a_m &\leq 2^{n-1} a_{2^{n-1}} \\ \sum_{m=2}^{2^N} a_m &\leq \sum_{n=1}^{N} 2^{n-1} a_{2^{n-1}}\end{aligned}$$

so that $\sum a_m$ is bounded above (since it is increasing) and because it is increasing, it must converge. □

Lemma 1.18. *The sum* $\sum_1^\infty \frac{1}{n^k}$ *converges iff* $k > 1$.

Proof. Clearly for $k \leq 0$, this diverges. Now for $k > 0$, $a_n = \frac{1}{n^k}$ is decreasing. So consider $\sum 2^n a_{2^n} = \sum 2^n \frac{1}{2^{nk}} = \sum 2^{n-nk} = \sum \left(2^{1-k}\right)^n$ which is a geometric series that converges iff $2^{1-k} < 1 = 2^0$, i.e. $k > 1$. Hence by Cauchy's condensation test, the original sum converges iff $k > 1$. □

1.5 Alternating Series and Absolute Convergence

An *alternating series* is one with terms of alternating sign. The following test gives us some very useful information for analyzing such series, in the case that the absolute value of the terms is decreasing.

Theorem 1.19. *Let* $(a_n)_1^\infty$ *be a positive,* decreasing *series. Then* $\sum_1^\infty (-1)^{n+1} a_n$ *converges iff* $a_n \to 0$.

Proof. Clearly if $a_n \not\to 0$, this series diverges. So we need to show that if $a_n \to 0$ the sum converges.

Write $s_n = a_1 - a_2 + \cdots + (-1)^{n+1} a_n$ for the nth partial sum. Now

$$s_{2n} = (a_1 - a_2) + (a_3 - a_4) + \cdots + (a_{2n-1} - a_{2n}) \geq 0$$

is an increasing series (each bracketed term is positive). But

$$s_{2n} = a_1 - (a_2 - a_3) - (a_4 - a_5) - \cdots - (a_{2n-2} - a_{2n-1}) - a_{2n} \leq a_1$$

so it is bounded above, and we have $s_{2n} \to s$. Now since $s_{2n-1} = s_{2n} + a_{2n} \to s + 0 = s$, we have $s_n \to s$. □

Remark. This is a special case of *Abel's test*.

Note that it is therefore in some sense much 'easier' to find a convergent series if we have alternating signs when we are summing. This motivates the following definition:

Definition 1.20. Take $a_n \in \mathbb{R}, \mathbb{C}$. If $\sum |a_n|$ is convergent, then $\sum a_n$ is *absolutely convergent.*

If $\sum a_n$ converges but $\sum |a_n|$ does not, then $\sum a_n$ is *conditionally convergent.*

Theorem 1.21. *If $\sum |a_n|$ converges, then $\sum a_n$ converges.*

Proof. We can do this very concisely as follows for the real case: $0 \leq (a_n + |a_n|) \leq 2|a_n|$. Thus since $\sum 2|a_n| \to 2a$, by comparison, $\sum (a_n + |a_n|)$ converges. Thus $\sum (a_n + |a_n|) - \sum |a_n| = \sum a_n$ also converges.

Another proof for the real case considers $v_n = \max\{0, a_n\} \geq 0$ and $w_n = \max\{0, -a_n\} \geq 0$. Note $a_n = v_n - w_n$. Then $|a_n| = v_n + w_n \geq v_n \geq w_n \geq 0$. So by comparison, $\sum v_n$ and $\sum w_n$ both converge, and therefore so does $\sum a_n$.

Now consider the complex case, $a_n = x_n + iy_n \in \mathbb{C}$ where $x_n, y_n \in \mathbb{R}$. Since $|x_n|, |y_n| \leq |a_n|$, by comparison $\sum |x_n|$ and $\sum |y_n|$ both converge absolutely and hence $\sum (x_n + iy_n) = \sum a_n$ does.

Alternatively, we can prove the whole theorem making use of Cauchy sequences. Let $s_n = \sum_{i=1}^{n} a_i$ be the partial sums. Then for $q \geq 0$

$$|s_{n+q} - s_n| = \left| \sum_{i=n}^{n+q} a_i \right| \leq \sum_{i=n}^{n+q} |a_i| = d_{n+q} - d_n$$

writing $d_n = \sum_{i=1}^{n} |a_i|$ for the (convergent) partial sums of the absolutely convergent series. But since d_n is convergent, it is Cauchy so that we can find N such that $\forall n \geq N$, $d_{n+q} - d_n < \epsilon$. But then $|s_{n+q} - s_n| \leq d_{n+q} - d_n < \epsilon$ so s_n is Cauchy and hence convergent. □

The name 'conditionally convergent' indicates that in fact, if we re-order the terms in the sum, we can alter the limit of the sum. In fact, the *Riemann series theorem* or *Riemann rearrangement theorem* shows that we can actually obtain any limit at all by this process - we can even make the sum diverge. (See the extra section 6.)

Example 1.22. Given that $1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\cdots=\log 2$, find the value of $\left(1-\frac{1}{2}-\frac{1}{4}\right)+\left(\frac{1}{3}-\frac{1}{6}-\frac{1}{8}\right)+\cdots$, which has the same terms, but re-arranged.

Note the groups of three are of the form $\left(\frac{1}{2k-1}-\frac{1}{2(2k-1)}\right)-\frac{1}{4k}=\frac{1}{2(2k-1)}-\frac{1}{2(2k)}$, so the sum equals $\frac{1}{2}\left(1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\cdots\right)=\frac{1}{2}\log 2$.

Theorem 1.23. *If $\sum a_n$ is absolutely convergent, then every series consisting of the same terms rearranged has the same limit.*

Proof. We give a proof for real sequences only; let $\sum b_n$ be a rearrangement of $\sum a_n$, where $s_n=\sum_{i=1}^n a_i \to s$, and $t_n=\sum_{i=1}^n b_i$.

Now given n, there is some q such that s_q contains every term appearing in t_n.

If $a_n \geq 0$, then $t_n \leq s_q \leq s$, and $t_n \to t$ for some $t \leq s$. But by symmetry, reversing the argument, $s \leq t$ so $s=t$.

Now if a_n has some sign, then we take $v_n=\max\{0,a_n\} \geq 0$ and $w_n=\max\{0,-a_n\} \geq 0$ which must converge as series as above, so $a_n=v_n-w_n$, and similarly choose $b_n=x_n-y_n$ for $x_n,y_n \geq 0$. Then by the above, $\sum^\infty v_n=\sum^\infty x_n$ and $\sum^\infty w_n=\sum^\infty y_n$, so that $t_n=\sum(x_n-y_n) \to \sum^\infty(v_n-w_n)=\lim_{n\to\infty} s_n=s$. $\square$

2 Continuity

In general, we consider $f : E \to \mathbb{C}$, where $E \subseteq \mathbb{C}$ is non-empty.

2.1 Definitions and Basic Properties

We shall provide two definitions of continuity at a point, and then prove their equivalence.

> **Definition 2.1.** f is *continuous* at $a \in E$ if, given any sequence $z_n \in E$ such that $z_n \to a$, then $f(z_n) \to f(a)$.

This first definition is stating that no matter how we choose to approach a point a, the value of the function will always approach the value at the point.

> **Definition 2.2.** f is *continuous* at $a \in E$ if, given $\epsilon > 0$, there exists some $\delta > 0$ such that if $z \in I$ and $|z - a| < \delta$, then $|f(z) - f(a)| < \epsilon$.

This is the well-known epsilon-delta (ϵ-δ) formulation of continuity, stating that given any non-trivial neighbourhood around the function's value at a point, there is a non-trivial neighbourhood of E such that the function's values lie in the given neighbourhood.

Proposition 2.3. *These two definitions are equivalent.*

Proof.

$\Longleftarrow$: We know that given $\epsilon > 0$ there is some $\delta > 0$ with $|f(z) - f(a)| < \epsilon$ when $|z - a| < \delta$. Take $z_n \in E$ tending to a. Then there is some N such that $|z_n - a| < \delta$ for $n \geq N$; and hence $|f(z_n) - f(a)| < \epsilon$ for $n \geq N$. Hence $f(z_n) \to f(a)$.

$\Longrightarrow$: Suppose that given any sequence $z_n \in E$ such that $z_n \to a$, then $f(z_n) \to f(a)$, but that the second definition does not hold; i.e. we have some $\epsilon > 0$ such that $\forall \delta > 0$ there is some $z \in E$ with $|z - a| < \delta$ and $|f(z) - f(a)| \geq \epsilon$. Now take $\delta_n = \frac{1}{n}$, and a sequence z_n with $|z_n - a| < \delta_n$ such that $|f(z_n) - f(a)| \geq \epsilon$. Now since $\delta_n \to 0$, $z_n \to a$; but then by assumption, $f(z_n) \to f(a)$, contradicting the fact that $|f(z_n) - f(a)| \geq \epsilon$. Hence the ϵ-δ continuity condition must hold.

□

Now we can use these definitions interchangeably.

Proposition 2.4. *Given $f, g : E \to \mathbb{C}$ continuous at $a \in E$. Then $f + g$, fg and λf (for all $\lambda \in \mathbb{C}$) are also continuous at a. If $g \neq 0$, then $\frac{1}{g}$ is continuous at a.*

Proof. Since the corresponding statements are true of all sequences, using the first definition, these must hold. ☐

Corollary 2.5. *All polynomials are continuous everywhere in $\mathbb{C}$, because $f(z) = z$ clearly is (using the second definition, we see $\delta = \epsilon$). Also, rational functions (quotients of polynomials) are continuous at every point where the denominator does not vanish.*

Note that we say a function is a continuous on a set $S \subseteq E$ iff it is continuous at every point in S.

Theorem 2.6. *Let $f : A \to \mathbb{C}$, $g : B \to \mathbb{C}$ be two functions such that $f(A) \subseteq B$. Then if f is continuous at a and g is continuous at $f(a)$, then the composition $g \circ f$ is continuous at a.*

Proof. Take $z_n \in A$ with $z_n \to a$.

As f is continuous, the sequence $z_n' = f(z_n) \in B$ converges $z_n' \to f(a)$.

As g is continuous, $z_n'' = g(z_n') = g \circ f(z_n)$ also converges to $g(f(a))$. ☐

Example 2.7.

(i) Investigate the continuity of $f(x) = \begin{cases} \sin \frac{1}{x} & x \neq 0 \\ 0 & x = 0 \end{cases}$ (assuming that $\sin x$ is a continuous function).

At $x \neq 0$, this is continuous because $\frac{1}{x}$ is a rational function, and $\sin \frac{1}{x}$ is a composition of continuous functions.

At $x = 0$, we need to consider the behaviour of the function $\sin \frac{1}{x}$ as $x \to 0$. In fact, it is clear that $\sin \frac{1}{x}$ does not tend to 0 because it will take the value 1 on the sequence $x_n = \frac{1}{(2n+\frac{1}{2})\pi} \to 0$. So f is discontinuous at $x = 0$.

(ii) Investigate the continuity of $f(x) = \begin{cases} x \sin \frac{1}{x} & x \neq 0 \\ 0 & x = 0 \end{cases}$ (assuming that $\sin x$ is a continuous function).

As above, for $x \neq 0$ this is definitely continuous. Now for $x = 0$, note that $|f(x_n)| \leq |x_n|$, so $f(x_n) \to 0$ for any sequence $x_n \to 0$. Hence as $f(x_n) = 0$, f is also continuous here.

Definition 2.8. Let $f : E \to \mathbb{C}$, where $E \subseteq \mathbb{C}$. Take $a \in \mathbb{C}$ where a is not isolated in E (i.e. there is some non-constant sequence $z_n \in E$ such that $z_n \to a$ - note that a is not necessarily in E itself). Then we say $f(z)$ has the *limit* l, written

$$\lim_{z \to a} f(z) = l$$

or $f(z) \to l$ as $z \to a$, if given $\epsilon > 0$ there is some $\delta > 0$ such that $\forall z \in E$ with $0 < |z-a| < \delta$, $|f(z) - l| < \epsilon$.

Remark.

(i) As was shown for the definitions of continuity $\lim_{z \to a} f(z) = l \iff$ for all sequences $z_n \in E$ ($z_n \to a$ but $z_n \neq a$) $f(z_n) \to l$.

(ii) If $a \in E$ then $\lim_{z \to a} f(z) = f(a) \iff f$ is continuous at a.

The above definition has the expected properties of uniqueness, linearity, multiplicativity, the $\lim \frac{1}{g(z)} = \frac{1}{\lim g(z)}$ if $\lim g(z) \neq 0$. This is all clear from the same arguments as before.

The final result we show in this section is about *uniform continuity*, an important concept which we will use in showing continuous functions are (Riemann) integrable.

Lemma 2.9 (Uniform continuity)**.** *If $f : [a,b] \to \mathbb{R}$ is continuous, then given $\epsilon > 0$ there is some $\delta > 0$ such that if $|x-y| < \delta$, for two points in this interval, then $|f(x) - f(y)| < \epsilon$.*

This is essentially stating that not only does a function continuous at every point have some suitable function $\delta(x, \epsilon)$, but that there is a function $\delta(\epsilon)$ which 'works for all x'.

Proof. Suppose there is not such a δ. Then there is some $\epsilon > 0$ such that for all $\delta > 0$ we can find x, y with $|x-y| < \delta$ but $|f(x) - f(y)| \geq \epsilon$.

Take $\delta = \frac{1}{n}$, and take x_n, y_n satisfying the above. Now by Bolzano-Weierstrass, we have some subsequence $x_{n_k} \to c$ where $c \in [a,b]$.

But then consider the corresponding subsequence in y:

$$\begin{aligned} |y_{n_k} - c| &\leq |x_{n_k} - y_{n_k}| + |x_{n_k} - c| \\ &< \frac{1}{n_k} + |x_{n_k} - c| \\ &\to 0 \end{aligned}$$

as $n_k \to \infty$ and therefore $y_{n_k} \to c$.

But $|f(x_{n_k}) - f(y_{n_k})| \geq \epsilon$ where ϵ is fixed, whilst since f is continuous, $f(x_{n_k}) \to f(c)$ and $f(y_{n_k}) \to f(c)$. This is a contradiction, so we are done. □

2.2 Intermediate Values, Extreme Values and Inverses

We now consider some of the consequences of continuity, which will prove very useful in analyzing differentiability properties too.

Theorem 2.10 (The Intermediate Value Theorem)**.** *For a continuous function $f : [a,b] \to \mathbb{R}$, such that $f(a) \neq f(b)$; then f takes on every value lying between $f(a)$ and $f(b)$.*

Proof. Without loss of generality, assume $f(a) < f(b)$, so that we want to find $f^{-1}(\eta)$ where $\eta \in (f(a), f(b))$.

Consider $S = \{x \in [a,b] : f(x) < \eta\} \neq \emptyset$, as $a \in S$. Also, S is bounded above (by b) and hence $c = \sup S$ exists.

Given $n > 0$ an integer, $c - \frac{1}{n}$ is not an upper bound for S, as c is a least upper bound. So therefore, if $c \neq a$ there is a sequence $x_n \in S \subseteq [a,b]$ such that $x_n > c - \frac{1}{n}$. Then $x_n \to c$ as $n \to \infty$ (as $x_n \le c$), and also $x_n \in S$ implies $f(x_n) < \eta$. By continuity of f, $f(x_n) \to f(c)$. Hence $f(c) \le \eta$. (If $c = a$ it is immediate that $f(c) = f(a) \le \eta$.)

Now note $c \neq b$ as otherwise $f(c) = f(b) \le \eta$, contradicting $\eta < f(b)$. Hence for all n sufficiently large $c + \frac{1}{n} \in [a,b]$, and $c + \frac{1}{n} \to c$. So by continuity, $f\left(c + \frac{1}{n}\right) \to f(c)$. Since $c + \frac{1}{n} > c$, $c + \frac{1}{n} \notin S$. So $f\left(c + \frac{1}{n}\right) \ge \eta$, and hence $f(c) \ge \eta$.

Therefore, $f(c) = \eta$. □

This theorem has many different areas of application, one of which we now present in the form of an example:

Example 2.11. Show there is an nth root of any positive number y, for $n \in \mathbb{N}$.

Consider $f(x) = x^n$, $x \ge 0$. Then f is continuous as it is polynomial.

Consider f in the interval $[0, 1+y]$. Then $0 = f(0) < y < f(1+y) = (1+y)^n$. Then by the intermediate value theorem, there is some $c \in (0, 1+y)$ such that $f(c) = c^n = y$ and hence there is a positive nth root. In fact, it is unique, since if say $d < c$, $y = d^n < c^n = y$.

We now consider the boundedness of a function f.

Theorem 2.12. *A continuous function $f : [a,b] \to \mathbb{R}$ on a closed interval is bounded; i.e. $\exists k : |f(x)| \le k$ for all $x \in [a,b]$.*

Proof. Suppose there is not such a k. Then for any $n \in \mathbb{N}$, there is some $x_n \in [a,b]$ with $|f(x_n)| > n$. By Bolzano-Weierstrass, there is some convergent subsequence $x_{n_j} \to x$, and $x \in [a,b]$. But then $\left|f\left(x_{n_j}\right)\right| > n_j$ and hence $f\left(x_{n_j}\right) \to \infty$ contradicting continuity, since $f\left(x_{n_j}\right) \to f(x)$. □

Theorem 2.13. *A continuous function $f : [a,b] \to \mathbb{R}$ on a closed interval is bounded and achieves its bounds; i.e. there exist $x_1, x_2 \in [a,b]$ such that $f(x_1) \le f(x) \le f(x_2)$ for all $x \in [a,b]$.*

Proof. We offer two proofs.

(i) Let $M = \sup\{f(x) : x \in [a,b]\}$, which we know is defined by the above.

Now take $M - \frac{1}{n} < M$. By the definition of the supremum, there is some sequence $c_n \in [a, b]$ such that $M - \frac{1}{n} \le f(c_n) \le M$.

By Bolzano-Weierstrass, there is some convergent subsequence $c_{n_j} \to x_2$, and so, since these points lie in $[a, b]$, $M - \frac{1}{n_j} \le f\left(c_{n_j}\right) \le M$, and hence $M = f(x_2)$ by continuity.

The same argument applies to find a suitable x_1.

(ii) We have $M = \sup\{f(x) : x \in [a, b]\} < \infty$ as above. Suppose for a contradiction that $f(x) < M$ for all $x \in [a, b]$.

Consider $g(x) = \frac{1}{M - f(x)}$ in $[a, b]$, a continuous function by hypothesis. But by the previous theorem applied to g, there is some bound such that $|g(x)| \le k$ for all $x \in [a, b]$. Since $g > 0$, $\frac{1}{M - f(x)} \le k$ and hence $\frac{1}{k} \le M - f(x)$. Then $f(x) \le M - \frac{1}{k}$ for all $x \in [a, b]$. But this contradicts the definition of M as the supremum of the values of f, as $f(x) \le M - \frac{1}{k} < M$.

□

Finally, we investigate the existence of an inverse functions.

Definition 2.14. $f : [a, b] \to \mathbb{R}$ is *(strictly) increasing* if $x_1 > x_2$ implies $f(x_1) \ge f(x_2)$ ($f(x_1) > f(x_2)$).

Theorem 2.15. *Let $f : [a, b] \to \mathbb{R}$ be continuous and strictly increasing. Then if $f([a, b]) = [c, d]$, the function $f : [a, b] \to [c, d]$ is bijective and $g = f^{-1} : [c, d] \to [a, b]$ is also continuous and strictly increasing.*

Proof. First we prove the necessary properties of f:

Injectivity: Since f is strictly increasing, $f(x_1) = f(x_2)$ implies $x_1 = x_2$.

Surjectivity: This follows from the intermediate value theorem.

Hence f is bijective, and has an inverse.

Monotonicity: Take $y_1 = f(x_1) < y_2 = f(x_2)$; then as f is strictly increasing, we must have $x_1 < x_2$.

Continuity: Consider $k = f(h)$, for $h \in (a, b)$. Given $\epsilon > 0$, let $k_1 = f(h - \epsilon)$ and $k_2 = f(h + \epsilon)$. We have $k_1 < k < k_2$ and $h - \epsilon < g(y) < h + \epsilon$ for any $y \in (k_1, k_2)$. So taking $\delta = \min\{|k_2 - k|, |k - k_1|\}$, we have that whenever $|y - k| < \delta$, $|g(y) - h| < \epsilon$. For the endpoints, reproduce the appropriate half of the argument.

□

3 Differentiability

In general we consider $f : E \subseteq \mathbb{C} \to \mathbb{C}$, but we are chiefly concerned with functions $f : [a, b] \to \mathbb{R}$.

3.1 Definitions and Basic Properties

Definition 3.1. f is *differentiable* at x, with *derivative* $f'(x)$, iff

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

Remark.

(i) We could also write $h = y - x$ so that $\lim_{h \to 0} \frac{f(x+h)-f(x)}{h} = f'(x)$.

(ii) Consider $\epsilon(h) = \frac{f(x+h)-f(x)-hf'(x)}{h}$.

We see $f(x+h) = f(x) + hf'(x) + \epsilon(h)h$, where $\lim_{h \to 0} \epsilon(h) = 0$.

Hence an equivalent way of stating that f is differentiable at x is to say that there is a function $\epsilon(h)$ such that $f(x+h) = f(x) + hf'(x) + h\epsilon(h)$ with $\lim_{h \to 0} \epsilon(h) = 0$, where $f'(x)$ is the derivative of f.

We also write $f(x+h) = f(x) + hf'(x) + \tilde{\epsilon}(h)$ where $\frac{\tilde{\epsilon}(h)}{h} \to 0$ as $h \to 0$, or

$$f(x) = f(a) + (x-a) f'(a) + \epsilon(x)(x-a)$$

with $\epsilon(x) \to 0$.

(iii) If f is differentiable at x, then it is continuous at x, since $f(x+h) = f(x) + hf'(x) + h\epsilon(h)$ with $\epsilon(h) \to 0$ as $h \to 0$ implies that $\lim_{h \to 0} f(x+h) = f(x)$.

Example 3.2. Assess the differentiability $f : \mathbb{R} \to \mathbb{R}$ maps $x \mapsto |x|$.

- $x > 0$: $\lim_{h \to 0} \frac{f(x+h)-f(x)}{h} = \lim_{h \to 0} \frac{x+h-x}{h} = 1 = f'(x)$
- $x < 0$: $\lim_{h \to 0} \frac{f(x+h)-f(x)}{h} = \lim_{h \to 0} \frac{-(x+h)+x}{h} = -1 = f'(x)$
- $x = 0$: Consider $\lim_{h \to 0} \frac{f(h)-f(0)}{h} = \lim_{h \to 0} \frac{f(h)}{h}$. The one-sided limits $\lim_{h \to 0^+} \frac{f(h)}{h} = 1$ and $\lim_{h \to 0^-} \frac{f(h)}{h} = -1$ do not agree, so this limit does not exist, even though f is continuous at 0.

Hence this function is differentiable everywhere except at $x = 0$.

Proposition 3.3.

(i) *If $f(x) = c$ $\forall c \in E$ then it is differentiable on E and $f'(x) = 0$.*

(ii) *If f, g differentiable at x then $f + g$ is differentiable, $(f+g)' = f' + g'$.*

(iii) *If f, g differentiable at x then fg is differentiable, $(fg)' = f'g + fg'$.*

(iv) *If f is differentiable at x, and $f(t) \neq 0$ for all $t \in E$, then $\frac{1}{f}$ is differentiable and $\left(\frac{1}{f}\right)' = \frac{-f'}{f^2}$.*

Proof.

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

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

(iii) We have

$$\begin{aligned}\frac{f(x+h)\,g(x+h)-f(x)\,g(x)}{h} &= \frac{f(x+h)\left[g(x+h)-g(x)\right]-g(x)\left[f(x+h)-f(x)\right]}{h} \\ &= f(x+h)\,\frac{g(x+h)-g(x)}{h}-g(x)\,\frac{f(x+h)-f(x)}{h} \\ &\to f(x)\,g'(x)-g(x)f'(x)\end{aligned}$$

by continuity and differentiability.

(iv) We have

$$\begin{aligned}\frac{1/f(x+h)-1/f(x)}{h} &= \frac{f(x)-f(x+h)}{hf(x+h)f(x)} \\ &= \frac{f(x)-f(x+h)}{h}\times\frac{1}{f(x+h)f(x)} \\ &\to -f'(x)\times\frac{1}{\left[f(x)\right]^2}\end{aligned}$$

$\square$

Remark. From the last two properties, we get $\left(\frac{f}{g}\right)' = \frac{f'g-fg'}{g^2}$.

Proposition 3.4. *If $f(x) = x^n$, $f'(x) = nx^{n-1}$, for $n \in \mathbb{Z}$.*

Proof. First, note that if $n = 0$, $f(x) = 1$ so $f'(x) = 0$ as required. Also, if $n = 1$, so $f(x) = x$, then $f'(x) = \lim \frac{x+h-x}{h} = 1$.

Then we proceed for positive integers by induction; $(x \cdot x^n)' = (x)' \cdot x^n + x \cdot (x^n)' = x^n + x \cdot nx^{n-1} = (n+1)\,x^n$.

For negative powers, we have $(x^{-n})' = \left(\frac{1}{x^n}\right)' = \frac{-nx^{n-1}}{x^{2n}} = \frac{-n}{x^{n+1}} = -nx^{-n-1}$. $\square$

It follows that all polynomials and rational functions are differentiable, except where the denominator vanishes.

Theorem 3.5 (The chain rule). *Let $f : U \to \mathbb{C}$ be differentiable at $a \in U$, such that $f(x) \in V$ $\forall x \in U$, and let $g : V \to \mathbb{R}$ is differentiable at $f(a)$, then $g \circ f : U \to \mathbb{R}$ is differentiable at a and $(g \circ f)' = (g' \circ f)\, f'$.*

Proof. Since f is differentiable at a,

$$f(x) = f(a) + (x-a) f'(a) + (x-a)\epsilon_f(x)$$

for some function with $\lim_{x\to a}\epsilon_f(x) = 0$. Also,

$$g(y) = g(f(a)) + (y - f(a))\, g'(f(a)) + (y - f(a))\,\epsilon_g(y)$$

where $\lim_{y\to f(a)}\epsilon_g(y) = 0$.

Now at $y = f(x)$,

$$\begin{aligned}
g(f(x)) &= g(f(a)) + (f(x) - f(a))\left[g'(f(a)) + \epsilon_g(f(x))\right] \\
&= g(f(a)) + \left[(x-a) f'(a) + \epsilon_f(x)(x-a)\right]\left[g'(f(a)) + \epsilon_g(f(x))\right] \\
&= g(f(a)) + (x-a)\, g'(f(a))\, f'(a) + (x-a)\underbrace{\left[\epsilon_f(x)\, g'(f(a)) + \epsilon_g(f(x))\left[f'(a) + \epsilon_f(x)\right]\right]}_{\sigma(x)}
\end{aligned}$$

So $g(f(x)) = g(f(a)) + (x-a)\, g'(f(a))\, f'(a) + (x-a)\,\sigma(x)$, and we need only show $\lim_{x\to a}\sigma(x) = 0$.

Note that we have functions $\epsilon_f(x)$ and $\epsilon_g(y)$ with an inherent ambiguity at $x = a$ and $y = f(a)$ respectively; since their values are arbitrary, we can set $\epsilon_f(a) = \epsilon_g(f(a)) = 0$ and then both functions are continuous at these points. Consequently, $\sigma(x)$ is continuous at $x = a$, and since $\sigma(a) = 0$, it follows that $\lim_{x\to a}\sigma(x) = 0$. □

Up to this point, everything holds in full generality for $f : E \subseteq \mathbb{C} \to \mathbb{C}$. We now look in more detail at the case of functions $f : [a, b] \to \mathbb{R}$.

3.2 The Mean Value Theorem and Taylor's Theorem

We now move on to consider what analytic use we can make of the differentiability of a function. We first prove a useful existence theorem:

Theorem 3.6 (Rolle's Theorem)**.** *Let $f : [a, b] \to \mathbb{R}$ be continuous, and further, differentiable on (a, b). If $f(a) = f(b)$ then there is some $c \in (a, b)$ such that $f'(c) = 0$.*

Proof. We know we have $M = \max_{x\in[a,b]} f(x)$ and $m = \min_{x\in[a,b]} f(x)$ achieved at some point. Let $k = f(a) = f(b)$.

If $M = m = k$ then f is constant, and so any point $c \in (a, b)$ has $f'(c) = 0$.

Otherwise $M > k$ or $m < k$. We give the argument for $M > k$, but the other argument is very similar.

We have c such that $f(c) = M$, $c \neq a, b$. By differentiability, we can write $f(c + h) = f(c) + h\left(f'(c) + \epsilon(h)\right)$ where $\epsilon(h) \to 0$ as $h \to 0$.

Because of this, if $f'(c) > 0$, then $h[f'(c) + \epsilon(h)] > 0$ for $h > 0$ small enough; then $f(c+h) > f(c) = M$, a contradiction.

Similarly, if $f'(c) < 0$, then $h[f'(c) + \epsilon(h)] < 0$ for $h < 0$ small enough; then $f(c+h) > f(c) = M$, also a contradiction.

Hence $f'(c) = 0$. □

Theorem 3.7 (The Mean Value Theorem). *Let $f : [a,b] \to \mathbb{R}$ be continuous, and differentiable on (a,b). Then there is some $c \in (a,b)$ such that $f(b) - f(a) = f'(c)[b-a]$.*

Proof. Consider $\phi(x) = f(x) - kx$, where we choose k so that $\phi(a) = \phi(b)$, that is, $k = \frac{f(b)-f(a)}{b-a}$.

Then by Rolle's Theorem, there is some $c \in (a,b)$ such that $\phi'(c) = 0$, that is, $f'(c) = k = \frac{f(b)-f(a)}{a-b}$ as required. □

Remark. We can rephrase this by letting $b = a + h$; then $f(a+h) = f(a) + hf'(a+\theta h)$ for some $\theta \in (0,1)$.

This theorem is essentially the first step to grasping the behaviour of $f(a+h)$ in terms of its derivatives. It has the following important consequences.

Corollary 3.8. *If $f : [a,b] \to \mathbb{R}$ is continuous and differentiable, then:*

- *if $f'(x) > 0$ $\forall x \in (a,b)$, then f is strictly increasing;*
- *if $f'(x) \geq 0$ $\forall x \in (a,b)$, then f is increasing;*
- *if $f'(x) = 0$ $\forall x \in (a,b)$, then f is constant.*

Proof. By the mean value theorem, $f(y) = f(x) + f'(c)(y-x)$ for any $x, y \in (a,b)$ and some $c \in (a,b)$.

So $y > x$ implies

- $f(y) > f(x)$ if $f'(c) > 0$;
- $f(y) \geq f(x)$ if $f'(c) \geq 0$;
- $f(y) = f(x)$ if $f'(c) = 0$.

□

We can interpret the last statement as giving the first definite solution to a differential equation.

Theorem 3.9 (Inverse function theorem). *Given $f : [a,b] \to \mathbb{R}$ continuous, and differentiable on (a,b), with $f'(x) > 0$ for all $x \in (a,b)$.*

Let $c = f(a)$ and $d = f(b)$. Then $f : [a, b] \to [c, d]$ is a bijection and $f^{-1} : [c, d] \to [a, b]$ is continuous on $[c, d]$ and differentiable on (c, d), with

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

Proof. Since $f'(x) > 0$ for all x in the interval, by the above corollary, f is strictly increasing. Hence, by Theorem 2.15, there is a continuous inverse $g = f^{-1}$.

We need to prove that g is differentiable with derivative $g'(y) = \frac{1}{f'(g(y))} = \frac{1}{f'(x)}$ where $y = f(x)$.

If $h \neq 0$ is sufficiently small, then there is a unique $k \neq 0$ such that $y + k = f(x + h)$, so that $g(y + k) = x + h$.

Now $\frac{g(y+k)-g(y)}{k} = \frac{x+h-x}{k} = \frac{h}{f(x+h)-f(x)}$. Then as $k \to 0$, $g(y + k) = x + h \to x = g(y)$ as $h \to 0$. Thus

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

□

Example 3.10. Rational powers of integers.

Let $f(x) = x^q$ for q a positive integer, $x \geq 0$, so its inverse $g(x) = x^{1/q}$.

$f' = qx^{q-1} > 0$ if $x > 0$. The inverse rule shows g is differentiable on $(0, \infty)$, and $g' = \frac{1}{q\left(x^{1/q}\right)^{q-1}} = \frac{1}{q}x^{\frac{1}{q}-1}$.

Hence letting $f(x) = x^{p/q}$ for p any integer, and q a positive integer. We find

$$\begin{aligned} g'(x) &= \left(\left(x^{1/q}\right)^p\right)' \\ &= p\left(x^{1/q}\right)^{p-1} \frac{1}{q}x^{\frac{1}{q}-1} \\ &= \frac{p}{q}x^{\frac{p}{q}-1} \end{aligned}$$

.

Hence for all $r \in \mathbb{Q}$, $(x^r)' = rx^{r-1}$. Once we define x^r for $r \in \mathbb{R}$, we will see this also holds here. (See Definition 4.15.)

Theorem 3.11 (Cauchy's mean value theorem)**.** *Suppose $f, g : [a, b] \to \mathbb{R}$ are continuous, and differentiable on (a, b). Then there is a $t \in (a, b)$ such that $[f(b) - f(a)]\, g'(t) = f'(t)\,[g(b) - f(a)]$.*

Proof. Consider the function $h : [a, b] \to \mathbb{R}$ given by

$$h(x) = \begin{vmatrix} 1 & 1 & 1 \\ f(a) & f(x) & f(b) \\ g(a) & g(x) & g(b) \end{vmatrix}$$

which is continuous and differentiable as f, g. Now $h(a) = h(b) = 0$ (since two columns are identical in each case). Hence, by Rolle's theorem, there is some $t \in (a, b)$ such that $h'(t) = 0$.

Expanding the determinant and differentiating, we see that $f'(t)g(b) - g'(t)f(b) + f(a)g'(t) - f'(t)g(a) = 0$ gives the above equality. $\square$

Corollary 3.12 (L'Hôpital's Rule)**.** *Suppose $f, g : [a, b] \to \mathbb{R}$ are continuous, and differentiable on (a, b), and that $f(a) = g(a) = 0$. Then if $\lim_{x \to a} \frac{f'(x)}{g'(x)} = l$ exists, then $\lim_{x \to a} \frac{f(x)}{g(x)} = l$.*

Proof. Note that by Cauchy's mean value theorem, that for all $x \in (a, b)$, we have some $t \in (a, x)$ such that $f(x)g'(t) = f'(t)g(x)$.

Now since $\lim_{x \to a} \frac{f'(x)}{g'(x)} = l$ exists, $g'(t) \neq 0$ in some neighbourhood (a, x), and in this interval $\frac{f'(t)}{g'(t)} g(x) = f(x)$. Since $g'(t) \neq 0$ for any t in this interval, it is impossible that $g(x) = g(a) = 0$ (by Rolle's theorem). So $\frac{f'(t)}{g'(t)} = \frac{f(x)}{g(x)}$ in this interval.

Thus $\lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{x \to a} \frac{f'(t)}{g'(t)} = \lim_{t \to a} \frac{f'(t)}{g'(t)} = l$. $\square$

Example 3.13. $\lim_{x \to 0} \frac{\sin x}{x} = \lim_{x \to 0} \frac{\cos x}{1} = \cos 0 = 1$.

We now consider the first of three forms of Taylor's theorem that we will derive in this course:

Theorem 3.14 (Taylor's theorem with Lagrange's remainder)**.** *Suppose f and its derivatives up to order $n-1$ are continuous on $[a, a+h]$, and $f^{(n)}$ exists for $x \in (a, a+h)$. Then there is a $\theta \in (0, 1)$ such that*

$$f(a+h) = f(a) + hf'(a) + \cdots + \frac{h^{n-1}}{(n-1)!} f^{(n-1)}(a) + \frac{h^n}{n!} f^{(n)}(a + \theta h)$$

Remark. $R_n = \frac{h^n}{n!} f^{(n)}(a + \theta h)$ is *Lagrange's* form of the remainder; in the case $n = 1$ this is precisely the mean value theorem.

Proof. Consider, for $t \in [0, h]$, the function

$$\phi(t) = f(a+t) - f(a) - tf'(a) - \cdots - \frac{t^{n-1}}{(n-1)!} f^{(n-1)}(a) - \frac{t^n}{n!} B$$

We will choose B so that $\phi(h) = 0$. Clearly, $\phi(0) = 0$.

We apply Rolle's theorem n times. Applying it to ϕ gives $h_1 \in (0, h)$ with $\phi'(h_1) = 0$. But $\phi'(0) = f'(a) - f'(a) = 0$. Applying Rolle again on $[0, h_1]$ gives $h_2 \in (0, h_1)$ with $\phi''(h_2) = 0$.

Also, note $\phi(0) = \phi'(0) = \cdots = \phi^{(n-1)}(0) = 0$. Hence continually applying Rolle gives $0 < h_n < \cdots < h_1 < h$ such that $\phi^{(i)}(h_i) = 0$.

Then $\phi^{(n)}(t) = f^{(n)}(a+t) - B$, so $\phi^{(n)}(h_n) = 0 = f^{(n)}(a + h_n) - B$. Hence since $h_n < h$ we can write $h_n = \theta h$, so then $B = f^{(n)}(a + \theta h)$, giving the statement of the theorem. $\square$

Theorem 3.15 (Taylor's theorem with Cauchy's remainder)**.** *Let f be as above, and assume $a = 0$ for simplicity. Then*

$$f(h) = f(0) + f'(0)h + \cdots + \frac{h^{n-1} f^{(n-1)}(0)}{(n-1)!} + R_n$$

where $R_n = (1-\theta)^{n-1} f^{(n)}(\theta h) \frac{h^n}{(n-1)!}$ for some $\theta \in (0,1)$.

Proof. Define, for $t \in [0,h]$, the function

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

noting that

$$\begin{aligned} g'(t) &= -f'(t) + [f'(t) - (h-t) f''(t)] + \left[(h-t) f''(t) - \frac{(h-t)^2}{2} f'''(t)\right] + \cdots - \frac{(h-t)^{n-1}}{(n-1)!} f^{(n)}(t) \\ &= -\frac{(h-t)^{n-1}}{(n-1)!} f^{(n)}(t) \end{aligned}$$

Then setting $\phi(t) = g(t) - \left(\frac{h-t}{h}\right)^p g(0)$ for $p \in \mathbb{Z}$, $1 \le p \le n$.

$\phi(0) = 0$ and $\phi(h) = 0$ (using the original definition of g), so applying Rolle gives $\phi'(\theta h) = 0$ for some $\theta \in (0,1)$.

Now we compute $\phi'(\theta h) = g'(\theta h) + p\left(\frac{h-\theta h}{h}\right)^{p-1} \frac{1}{h} g(0) = g'(\theta h) + p\frac{(1-\theta)^{p-1}}{h} g(0)$. Hence using $\phi'(\theta h) = 0$ we have

$$\phi'(\theta h) = 0 \quad = \quad \frac{-h^{n-1}(1-\theta)^{n-1}}{(n-1)!} f^{(n)}(\theta h) + p\frac{(1-\theta)^{p-1}}{h}\left[f(h) - f(0) - \cdots - \frac{h^{n-1}}{(n-1)!} f^{(n-1)}(0)\right]$$

which on rearrangement gives

$$f(h) = f(0) + f'(0)h + \cdots + \frac{h^{n-1} f^{(n-1)}(0)}{(n-1)!} + \underbrace{\frac{h^n (1-\theta)^{n-1}}{(n-1)! p (1-\theta)^{p-1}} f^{(n)}(\theta h)}_{R_n}$$

Now note that we have a general form of the remainder for $1 \le p \le n$ (note θ depends on p). We choose $p = 1$ to complete the proof of Cauchy's form. $\square$

In fact, by choosing $p = n$ we can also recover Lagrange's form of the remainder.

Remark. Note that we may re-translate to the $[x, x+h]$. Analogous results hold on intervals $[x-h, x]$ for $h > 0$.

3.3 Taylor Series and Binomial Series

Having established this result for general n, it is obviously tempting to consider the limit $n \to \infty$. In fact, it is clear that this would immediately give the Taylor expansion of the function f *if it converged*. To ensure that this is the case, we simply require $R_n \to 0$ in this limit.

Definition 3.16. The *Taylor series expansion* of a function f which is infinitely differentiable on the interval $[x, x+h]$ is

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

if the remainder term in Taylor's theorem tends to 0. The *Maclaurin series expansion* is the $x = 0$ case.

Remark. Note that the requirement that the $R_n \to 0$ is stronger than the statement that the series converges. The following example illustrates why.

Example 3.17. Show the function $f(x) = \exp\left(-1/x^2\right)$ for $x \neq 0$, $f(0) = 0$, is infinitely differentiable, and calculate $f^{(n)}(0)$ for all n. What does this mean in terms of Taylor series expansions for f?

Firstly, note that

$$\begin{aligned} \lim_{h\to 0} \frac{f(h) - f(0)}{h} &= \lim_{h\to 0} \frac{e^{-1/h^2}}{h} \\ &= \lim_{y\to\pm\infty} \frac{y}{e^{y^2}} \\ &= 0 \end{aligned}$$

since exponentials always grow faster than polynomials.

Thus we have

$$f'(x) = \begin{cases} \frac{2}{x^3} e^{-1/x^2} & x \neq 0 \\ 0 & x = 0 \end{cases}$$

More generally, for $x \neq 0$, $f^{(n)}(x) = p\left(\frac{1}{x}\right) e^{-1/x^2}$ for some polynomial p. Hence $f^{(n)}(x) \to 0$ as $x \to 0$ for all n, by a similar argument to the above. That is, f is infinitely differentiable, with all derivatives equal to 0. Hence an attempt to write a Taylor series expansion would give $f(h) = \sum_{n=0}^{\infty} 0 = 0$, which is clearly wrong.

But now note that the remainder terms, roughly$h^n f^{(n)}(x)$, contain terms like $\frac{h^n}{(\theta h)^{3n}} = h^{-2n}\theta^{-3n}$, which grow as $n \to \infty$. Hence $R_n(h)$ does not necessarily converge to 0, so the Taylor series does not necessarily converge to the actual value of f.

We now move on to consider a particularly useful example of a Taylor series, the binomial series.

Theorem 3.18 (Binomial series expansion). *For $|x| < 1$,*

$$(1+x)^r = \sum_{n=0}^{\infty} \binom{r}{n} x^n$$

where $\binom{r}{n} := \frac{r(r-1)\cdots(r-n+1)}{n!}$, $\binom{r}{0} = 1$, are the generalized binomial coefficients.

Proof. Clearly $f^{(n)}(x) = r \cdot (r-1) \cdots (r-n+1) \cdot (1+x)^{r-n}$, and $\frac{f^{(n)}(0)}{n!} = \binom{r}{n}$.

Consider Lagrange's form of the remainder:

$$\begin{aligned} R_n &= x^n (1+\theta x)^{r-n} \binom{r}{n} \\ &= (1+\theta x)^{r-n} \left[\binom{r}{n} x^n \right] \end{aligned}$$

Now $\binom{r}{n} x^n$ we can analyze using the convergence of the series itself - note that if $a_n = \binom{r}{n} x^n$, $\left|\frac{a_{n+1}}{a_n}\right| = \left|\frac{(r-n)x}{n+1}\right| \to |x| < 1$, so by the ratio test this converges. Hence $a_n = \binom{r}{n} x^n \to 0$.

Now including $(1+\theta x)^{r-n}$, we consider separately the two cases

$0 < x < 1$: Here, $1 + \theta x > 1$ so $(1+\theta x)^{r-n} < 1$ for $n > r$, and thus $|R_n| < \left|\binom{r}{n} x^n\right| \to 0$.

$-1 < x < 0$: In this case, we must use something more inventive. Making use of our alternative form, due to Cauchy, for the remainder, we get

$$\begin{aligned} R_n &= \frac{(1-\theta)^{n-1} r (r-1) \cdots (r-n+1) (1+\theta x)^{n-r} x^n}{(n-1)!} \\ &= r \binom{r-1}{n-1} \frac{(1-\theta)^{n-1}}{(1+\theta x)^{n-r}} x^n \\ &= r \left(\frac{1-\theta}{1+\theta x} \right)^{n-r} (1-\theta)^{r-1} \binom{r-1}{n-1} x^n \\ &= r \left(\frac{1-\theta}{1+\theta x} \right)^{n-1} (1+\theta x)^{r-1} \binom{r-1}{n-1} x^n \end{aligned}$$

Now $\frac{1-\theta}{1+\theta x} < 1$, and if $r - 1 \geq 0$, $(1-\theta)^{r-1}$ is bounded, whilst if $r - 1 < 0$, $(1+\theta x)^{r-1}$ is bounded; so in either case, these expressions are bounded by a constant multiplied by $\left|\binom{r-1}{n-1} x^n\right|$, which converges to 0, so we are done.

□

3.4 * Comments on Complex Differentiability

We can use the same definitions for differentiability in $\mathbb{C}$, but we have to be aware that we can take the limit in any direction in the complex plane, rather than just from the positive and negative directions.

As a result, *holomorphic* functions (infinitely differentiable) or *analytic* functions (equal to their power series) on complex domains have much stronger constraints than those on the real line. In fact, it turns out that *all holomorphic functions are analytic*, which has profound consequences for complex analysis. Cauchy's integral formula is a wonderful demonstration of this. The following example illustrates how differentiability is stronger for complex domains.

Example 3.19. Let $f : \mathbb{C} \to \mathbb{C}$ map $z \mapsto \overline{z}$. Then letting $z_n = z + \frac{1}{n}$ and $w_n = z + \frac{i}{n}$, we see that

$$\frac{f(z_n) - f(z)}{z_n - z} = \frac{1/n}{1/n} = 1$$

whilst

$$\frac{f(w_n) - f(z)}{w_n - z} = \frac{\overline{i}/n}{i/n} = -1$$

so the limit does not in general exist, and the function is nowhere differentiable, despite appearing quite smooth.

Note that the corresponding 2-dimensional map $f : \mathbb{R}^2 \to \mathbb{R}^2$ $f(x, y) = (x, -y)$ is differentiable.

4 Power Series

We consider functions given by some infinite sum $\sum_{n=0}^{\infty} a_n z^n$ for $a_n \in \mathbb{C}$ and $z \in \mathbb{C}$.

4.1 Fundamental Properties

Lemma 4.1. *If $\sum_{n=0}^{\infty} a_n z_0^n$ converges and $|z| < |z_0|$, then $\sum_{n=0}^{\infty} a_n z^n$ converges* absolutely.

Proof. Since $\sum_{n=0}^{\infty} a_n z_0^n$ converges, $a_n z_0^n \to 0$ as $n \to \infty$, and in particular it is bounded by some K.

Now

$$\begin{aligned} |a_n z^n| &= \left| a_n z_0^n \cdot \frac{z^n}{z_0^n} \right| \\ &\leq K \left| \frac{z^n}{z_0^n} \right| \end{aligned}$$

but as $|z| < |z_0|$, the sum on the right of

$$\sum_{n=0}^{\infty} |a_n z^n| \leq K \sum_{n=0}^{\infty} \left| \frac{z^n}{z_0^n} \right|$$

is a convergent geometric series, and hence the sum converges, by comparison. □

Theorem 4.2. *A power series either*

(i) *converges absolutely $\forall z \in \mathbb{C}$; or*

(ii) *converges absolutely $\forall z \in \mathbb{C}$ inside a circle $|z| = R$ and diverges $\forall z \in \mathbb{C}$ outside it; or*

(iii) *converges for $z = 0$ only.*

Definition 4.3. We the circle $|z| = R$ to be the *circle of convergence*, and R is the *radius of convergence*. We allow $R = 0$ and $R = \infty$ for the other cases.

Proof. Since we are investigating the properties of a borderline, we are motivated to define $S = \{x \in \mathbb{R} : x \geq 0 \text{ and } \sum a_n x^n \text{ converges}\}$. Then since $0 \in S$, S is non-empty, we either have S unbounded or $R = \sup S$ exists. Now note that by the above, we know that if $x_1 \in S$ the entire interval $[0, x_1] \subset S$, so in the former case setting $R = \infty$ we are done. Also, if $R = 0$, we are done.

Now otherwise, we need to show that for $|z| < R$, $\sum a_n z^n$ converges absolutely, whilst for $|z| > R$, it diverges.

But we can now use the known property again; since R is a least upper bound, for any $|z| < R$ there is some $x \in (|z|, R)$ such that the series converges; then we know $\sum a_n z^n$ converges absolutely.

Similarly, if the series converged for some z with $|z| > R$, then we would have some $x \in (R, |z|)$ such that the series converged, which is a contradiction. □

Having established this incredibly useful property of power series, we can introduce methods to compute R. In fact, since these are just regular series, it is very natural to just turn to our main two tests from the section on series.

Lemma 4.4. *If* $\left|\frac{a_{n+1}}{a_n}\right| \to l$ *then* $R = \frac{1}{l}$ *(with* $l = 0$ *corresponding to* $R = \infty$ *and vice versa).*

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

$$\lim_{n\to\infty} \left|\frac{a_{n+1}z^{n+1}}{a_n z}\right| = l\,|z| < 1$$

Note this also shows that, if $|z| > \frac{1}{l}$, the limit is greater than 1 so the series diverges (the terms do not tend to 0).

The cases $l = 0, \infty$ are easy to show using this logic. □

Lemma 4.5. *If* $|a_n|^{1/n} \to l$ *then* $R = \frac{1}{l}$ *(with* $l = 0$ *corresponding to* $R = \infty$ *and vice versa).*

Proof. Identical to the above. □

With these simple tests, it is now very easy and natural to work with particular power series.

Example 4.6. The following are simple examples of the behaviour of power series, including boundary behaviour. (We omit the range of the series for clarity.)

(i) $\sum \frac{z^n}{n!}$, the exponential series, has $\left|\frac{n!}{(n+1)!}\right| = \frac{1}{n+1} \to 0$ so $R = \infty$. Thus the exponential series converges everywhere.

(ii) $\sum z^n$, the geometric series, has $R = 1$ as before. Note that at the boundary $|z| = 1$ the series diverges.

(iii) $\sum n! z^n$ clearly has $R = 0$, as can be verified by noting $\left|\frac{(n+1)!}{n!}\right| = n + 1 \to \infty$.

(iv) $\sum \frac{z^n}{n}$. Here, $\left|\frac{n}{n+1}\right| \to 1$, so $R = 1$. The boundary behaviour here is more complicated. If $z = 1$, then this is the harmonic series, which diverges. However, if $z \neq 1$ one can use Abel's test with $b_n = \frac{1}{n} \to 0$ and $a_n = z^n$, since $s_N = \sum_{i=1}^{N} z^i = \frac{1-z^{N+1}}{1-z} + 1$ is bounded in N. Thus $|s_N| \leq 1 + \frac{1+|z|^{N+1}}{|1-z|} \leq 1 + \left|\frac{z}{1-z}\right|$ is bounded in N, so if $|z| = 1$ but $z \neq 1$ this converges.

(v) $\sum \frac{z^n}{n^2}$ has $R = 1$. This converges everywhere on the boundary, as it converges absolutely ($\sum \frac{1}{n^2}$ converges).

(vi) $\sum nz^n$ also has $R = 1$, but on $|z| = 1$ the terms $|nz^n| = n \to \infty$ so this diverges everywhere on the boundary.

From this it is clear that, in general, the behaviour on the boundary is more complicated than elsewhere, as we might have expected based on the ratio and root tests.

We would now ideally like to merge our work on power series into that on differentiability. The following theorem expresses another very fundamental result:

Theorem 4.7. *Suppose $\sum a_n z^n$ has radius of convergence R, and define*

$$f(z) := \sum_{n=0}^{\infty} a_n z^n$$

for $|z| < R$. Then f is differentiable and further

$$f'(z) = \sum_{1}^{\infty} a_n n z^{n-1}$$

where $|z| < R$.

Remark. Iterating this theorem shows that f can be differentiated infinitely many times, term by term (just like a polynomial), so long as we remain strictly within its radius of convergence.

This is not particularly difficult to prove, though we use a comparatively long derivation.

Lemma 4.8. *We use the following two results to reach the conclusion.*

(i) $\binom{n}{r} \leq n(n-1)\binom{n-2}{r-2}$

(ii) $\left|(z+h)^n - z^n - nhz^{n-1}\right| \leq n(n-1)(|z| + |h|)^{n-2}|h|^2$ *for all $h, z \in \mathbb{C}$.*

Then: if $\sum a_n z^n$ has radius of convergence R, then so do $\sum n a_n z^{n-1}$ and $\sum n(n-1) a_n z^{n-2}$.

Proof. For the first stage, we see

$$\frac{\binom{n}{r}}{\binom{n-2}{r-2}} = \frac{n!\,(n-r)!\,(r-2)!}{r!\,(n-r)!\,(n-2)!} = \frac{n(n-1)}{r(r-1)} \leq n(n-1)$$

Then letting $c = (z+h)^n - z^n - nhz^{n-1} = \sum_{r=2}^{n} z^{n-r}h^r$,

$$\begin{aligned}
|c| &\leq \sum_{r=2}^{n} \binom{n}{r} |z|^{n-r} |h|^r \\
&\leq \sum_{r=2}^{n} n(n-1) \binom{n-2}{r-2} |z|^{n-r} |h|^r \\
&= n(n-1)|h|^2 \sum_{r=2}^{n} \binom{n-2}{r-2} |z|^{n-r} |h|^{r-2} \\
&= n(n-1)|h|^2 (|z|+|h|)^{n-2}
\end{aligned}$$

Now let $z \in \mathbb{C}$ satisfy $0 < |z| < R$. Then choose $R_0 \in (|z|, R)$, so $a_n R_0^n \to 0$ as $n \to \infty$ and so this is bounded by some K. Then

$$\begin{aligned}
\left|na_n z^{n-1}\right| &= \frac{n}{|z|} |a_n z^n| \\
&\leq \frac{nK}{|z|} \left|\frac{z}{R_0}\right|^n
\end{aligned}$$

and, by the ratio test, $\sum n \left|\frac{z}{R_0}\right|^n$ converges.

Then by comparison, $\sum na_n z^{n-1}$ converges absolutely.

But further, $|a_n z^n| \leq n|a_n z^n| = \frac{1}{|z|} n \left|a_n z^{n-1}\right|$, so by comparison, if $\sum na_n z^{n-1}$ converges absolutely, so does $\sum a_n z^n$. Thus the radii of convergence are identical.

We can use much the same argument to see that $\sum n(n-1) a_n z^{n-2}$ also has the same radius of convergence. $\square$

We are now ready to prove the original theorem:

Proof. Since we have shown $\sum na_n z^{n-1}$ has radius of convergence R, it defines some function $g(z)$ in this circle.

Consider the quantity $\frac{f(z+h)-f(z)-hg(z)}{h}$; if we can show it has limit 0 as $h \to 0$, we are done.

It equals

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

and by the above lemma, the sum terms are bounded by $|a_n| n(n-1)(|z|+|h|)^{n-2}|h|^2$. Take some $r > 0$ such that $|z| + r < R$; then these are ultimately bounded by $|a_n| n(n-1)(|z|+r)^{n-2}|h|^2$ and the corresponding series converges, as $|z|+r$ lies within the boundary of absolute convergence, to some value A.

Then taking h small enough, $\frac{|f(z+h)-f(z)-hg(z)|}{|h|} \leq \frac{1}{|h|} A|h|^2 = A|h| \to 0$ as $h \to 0$. $\square$

4.2 Standard Functions

The above work allows us to justify introducing functions using power series.

4.2.1 Exponentials and logarithms

We already know $\sum \frac{z^n}{n!}$ has infinite radius of convergence, so we can make the following familiar definition:

> **Definition 4.9.** The *exponential map* $e : \mathbb{C} \to \mathbb{C}$ is defined by
>
> $$e\left(z\right) = \sum_{0}^{\infty} \frac{z^n}{n!}$$

Now by the above theorem, we know $e'(z) = e(z)$ also has infinite radius of convergence.

Proposition 4.10. *If $F : \mathbb{C} \to \mathbb{C}$ is differentiable, and $F'(z) = 0$ for all $z \in \mathbb{C}$, then F is constant.*

Proof. Consider F parametrized on the line 0 to z, i.e. $g(t) = F\left(tz\right)$, $g : [0, 1] \to \mathbb{C}$.

Write $g(t) = u(t) + iv(t)$ for real-valued functions u, v.

Then by the chain rule, $g'(t) = zF'(tz) = 0$; but then $g' = u' + iv' = 0$, so $u' = v' = 0$ and hence u and v are both constant, by Proposition 3.2.

Then $g(0) = g(1)$ and hence $F(z) = F(0)$. □

Remark. This proof can be generalized to any *path-connected* domain in $\mathbb{C}$ etc. (i.e. for any domain containing a path between any two points).

Theorem 4.11. *We prove several key properties of the exponential map, mainly restricting ourselves to $\mathbb{R}$.*

(i) *$e : \mathbb{R} \to \mathbb{R}$ is differentiable everywhere, with $e' = e$.*

(ii) *$e\left(a + b\right) = e\left(a\right) e(b)$ for all $a, b \in \mathbb{C}$.*

(iii) *$e\left(x\right) > 0$ for $x \in \mathbb{R}$.*

(iv) *e is strictly increasing.*

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

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

Proof. We prove these sequentially.

(i) This is already established in more generality for $\mathbb{C}$.

(ii) Let $F : \mathbb{C} \to \mathbb{C}$ map $z \mapsto e(a+b-z)\,e(z)$. Then it is easy to note $F'(z) = 0$; hence $F(b) = e(a)\,e(b) = F(0) = e(a+b)\,e(0) = e(a+b)$ by the power series.

(iii) By definition, $e(x) > 0$ if $x \geq 0$, and $e(0) = 1$. Then by the above, $e(x)\,e(-x) = e(0) = 1$, so that $e(-x) = 1/e(x) > 0$.

(iv) $e'(x) = e(x) > 0$, so by Proposition 3.2 e is strictly increasing.

(v) By definition, $e(x) > 1 + x$ for $x > 0$, so $e(x) \to \infty$ as $x \to \infty$. If $x \to \infty$, $e(-x) = \frac{1}{e(x)} \to 0$.

(vi) e is strictly increasing, so this is injective. Surjectivity follows from the fact that there are arbitrarily small and large values $e(a)$ and $e(b)$, and the fact that the continuity of e means the intermediate value theorem holds.

$\square$

Remark. We have in fact shown that $e : (\mathbb{R}, +) \to ((0,\infty), \times)$ is a group isomorphism.

Since $e : \mathbb{R} \to (0,\infty)$ is bijective, there is a function $l : (0,\infty) \to \mathbb{R}$ such that $e(l(t)) = t$ for all $t \in (0,\infty)$, and $l(e(x)) = x$ for all $x \in \mathbb{R}$.

Definition 4.12. The function $\log : (0,\infty) \to \mathbb{R}$ is the inverse of e.

Theorem 4.13.

(i) $\log : (0,\infty) \to \mathbb{R}$ *is a bijection and* $\log(e(x)) = x$, $e(\log(t)) = t$ *for all acceptable ranges of* x, t.

(ii) $\log$ *is differentiable and* $\log'(t) = \frac{1}{t}$.

(iii) $\log(xy) = \log(x) + \log(y)$ *for all* $x, y \in (0,\infty)$.

Proof.

(i) This is true by definition.

(ii) log is differentiable by the inverse rule, Theorem 3.9, and further

$$\log'(t) = \frac{1}{e'(\log(t))} = \frac{1}{e(\log(t))} = \frac{1}{t}$$

(iii) log is the inverse of a group homomorphism, and therefore is a homomorphism.

$\square$

We can now define a more general function for $\alpha \in \mathbb{R}$ for $x > 0$:

$$r_\alpha(x) := e(\alpha \log x)$$

so that $r_1(x) = x$ etc.

Theorem 4.14. *Suppose $x, y > 0$, and $\alpha, \beta \in \mathbb{R}$*

(i) $r_\alpha(xy) = r_\alpha(x) r_\alpha(y)$.

(ii) $r_{\alpha+\beta}(x) = r_\alpha(x) r_\beta(x)$.

(iii) $r_\alpha(r_\beta(x)) = r_{\alpha\beta}(x)$.

Proof.

(i) Straightforwardly

$$\begin{aligned} r_\alpha(xy) &= e(\alpha \log(xy)) \\ &= e(\alpha \log x + \alpha \log y) \\ &= e(\alpha \log x) e(\alpha \log y) \\ &= r_\alpha(x) r_\alpha(y) \end{aligned}$$

(ii) Left as an exercise.

(iii) Finally,

$$\begin{aligned} r_\alpha(r_\beta(x)) &= e(\alpha\beta \log x) \\ &= r_{\alpha\beta}(x) \end{aligned}$$

□

Now for $n \in \mathbb{N}$, $r_n(x) = r_{\underbrace{1+\cdots+1}_{n}}(x) = x \cdot x \cdot \cdots \cdot x = x^n$. Also, since $r_0(x) = 1 = r_n(x) r_{-n}(x)$, $r_{-n}(x) = x^{-n}$.

For $q > 0$ a positive integer, $\left[r_{1/q}(x)\right]^q = r_1(x) = x$, so $r_{1/q}(x) = x^{1/q}$.

So $r_{p/q}(x) = x^{p/q}$.

Definition 4.15. We define exponentiation for $\alpha \in \mathbb{R}$ by $x^\alpha := r_\alpha(x)$.

By the above, this definition is entirely compatible with the definition for $x \in \mathbb{Q}$.

Definition 4.16. The *base of natural logarithms* $e \in \mathbb{R}$ is $e = \sum_0^\infty \frac{1}{n!}$, so that $\log e = 1$. $\exp(x) = e(x) = e^x$.

Hence $e^x = r_x(e) = \exp(x \cdot \log e) = \exp x$.

So $x^\alpha = e^{\alpha \log x}$.
Then $(x^\alpha)' = \alpha \frac{1}{x} e^{\alpha \log x} = \frac{\alpha}{x} x^\alpha = \alpha x^{\alpha - 1}$.
Also, $f(x) = a^x$ for $a > 0$ and $a \in \mathbb{R}$ gives $f(x) = e^{x \log a}$ and $f'(x) = (\log a) a^x$.

Proposition 4.17. *Exponentials grow faster than powers; i.e.* $\frac{e^x}{x^k} \to \infty$ *as* $x \to \infty$ *for all* $k \in \mathbb{R}$.

Proof. Take $n > k$ and consider the power series expansion; $e^x > \frac{x^n}{n!}$ so $\frac{e^x}{x^k} > \frac{x^{n-k}}{n^k} \to \infty$ as $x \to \infty$. □

4.2.2 Trigonometric functions

Continuing this process, we define the two important trigonometric functions.

Definition 4.18. The *sine* and *cosine* functions are given by

$$\begin{aligned} \sin z : &= \sum_0^\infty \frac{(-1)^n z^{2n+1}}{(2n+1)!} \\ \cos z &= \sum_0^\infty \frac{(-1)^n z^{2n}}{(2n)!} \end{aligned}$$

One may fairly easily verify these both have infinite radius of convergence.

By the above, therefore, these are both differentiable, and differentiating term by term, $\sin' z = \cos z$ and $\cos' z = -\sin z$.

Now we can easily derive Euler's Formula:

$$\begin{aligned} e^{iz} &= \sum_0^\infty \frac{(iz)^n}{n!} \\ &= \sum_0^\infty \frac{(iz)^{2n}}{(2n)!} + \sum_0^\infty \frac{(iz)^{2n+1}}{(2n+1)!} \\ &= \cos z + i \sin z \end{aligned}$$

Similarly, $e^{-iz} = \cos z - i \sin z$.
Then $\cos z = \frac{e^{iz} - e^{-iz}}{2}$ and $\sin z = \frac{e^{iz} - e^{-iz}}{2i}$.
So $\cos(-z) = \cos z$, $\sin(-z) = -\sin(z)$, $\cos(0) = 1$ and $\sin(0) = 0$.
Also, using $e^{a+b} = e^a e^b$, we get (for all $z, w \in \mathbb{C}$)

$$\begin{aligned} \sin(z+w) &= \sin z \cos w + \cos z \sin w \\ \cos(z+w) &= \cos z \cos w - \sin z \sin w \end{aligned}$$

Further, $\sin^2 z + \cos^2 z = 1$ for all $z \in \mathbb{C}$.
If $x \in \mathbb{R}$, this shows $|\sin x|, |\cos x| \leq 1$. (If $z \in \mathbb{C}$, neither function is bounded.)

Proposition 4.19. *There is a smallest positive number* ω *with* $\sqrt{2} < \frac{\omega}{2} < \sqrt{3}$ *such that* $\cos\frac{\omega}{2} = 0$, *and hence* $\sin\frac{\omega}{2} = 1$.

Proof. If $0 < x < 2$, $\sin x > 0$, since $\sin x = \left(x - \frac{x^3}{3!}\right) + \left(\frac{x^5}{5!} - \frac{x^7}{7!}\right) + \cdots$.

But $(\cos x)' = -\sin x < 0$ for $x \in (0, 2)$, so $\cos x$ is strictly decreasing function in this range.

Now

$$\begin{aligned}\cos\sqrt{2} &= \left(1 - \frac{2}{2!}\right) + \underbrace{\left(\frac{2^2}{4!} - \frac{2^3}{6!}\right)}_{>0} + \underbrace{\left(\frac{2^4}{8!} - \frac{2^5}{10!}\right)}_{>0} + \cdots \\ &> 0\end{aligned}$$

and

$$\begin{aligned}\cos x &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \underbrace{\left(\frac{x^6}{6!} - \frac{x^8}{8!}\right)}_{>0} - \underbrace{(\cdots)}_{>0} - \cdots \\ &< 1 - \frac{x^2}{2!} + \frac{x^4}{4!} \\ \cos\sqrt{3} &< 1 - \frac{3}{2!} + \frac{3^2}{4!} = -\frac{1}{8} < 0\end{aligned}$$

hence by the intermediate value theorem, there is a unique ω such that $\frac{\omega}{2} \in \left(\sqrt{2}, \sqrt{3}\right)$ such that $\cos\frac{\omega}{2} = 0$.

Then $\left|\sin\frac{\omega}{2}\right| = 1$, but $\sin\frac{\omega}{2} > 0$ so $\sin\frac{\omega}{2} = 1$. □

Definition 4.20. We define the constant *pi* to be the number such that $\pi := \omega$.

We can now prove several results about periodicity:

Theorem 4.21.

(i) $\sin\left(z + \frac{\pi}{2}\right) = \cos z$, $\cos\left(z + \frac{\pi}{2}\right) = -\sin z$.

(ii) $\sin(z + \pi) = -\sin z$, $\cos(z + \pi) = -\cos z$.

(iii) $\sin(z + 2\pi) = \sin z$, $\cos(z + 2\pi) = \cos z$.

Proof. These are all immediate from the addition formulae, and the known values of sin and cos. □

Remark. We also get the periodicity of the exponential function, $e^{z+2\pi i} = e^z$, and Euler's Identity,

$$\begin{aligned}e^{i\pi} &= \cos\pi + i\sin\pi \\ e^{i\pi} &= -1\end{aligned}$$

The other trigonometric functions are defined in the usual way as quotients etc. of sin and cos.

4.2.3 Hyperbolic functions

Definition 4.22. The *hyperbolic cosine* and *hyperbolic sine* are

$$\begin{aligned} \cosh z &= \frac{e^z + e^{-z}}{2} \\ \sinh z &= \frac{e^z - e^{-z}}{2} \end{aligned}$$

It follows that $\cosh z = \cos iz$ and $i \sinh z = \sin iz$, and also $\cosh' z = \sinh z$, and $\sinh' z = \cosh z$. Also, $\cosh^2 z - \sinh^2 z = 1$.

5 Integration

Having rigorously defined differentiation, it is natural to try to do the same for integration. The approach defined here is to define the usual Riemann integral. However, the definitions do not immediately seem related. It is actually in many respects easier to introduce the *Lebesgue* integral, but that is beyond the scope of this course.

5.1 Definitions and Key Properties

We consider *bounded* functions $f : [a, b] \to \mathbb{R}$.

> **Definition 5.1.** A *dissection* or *partition* of the interval $[a, b]$ is a finite subset $\mathcal{D}$ of $[a, b]$ which contains a and b. We write
>
> $$\mathcal{D} = \{a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b\}$$
>
> The *upper and lower sums* of f with respect to $\mathcal{D}$ are
>
> $$S(f, \mathcal{D}) = \sum_{j=1}^{n} (x_j - x_{j-1}) \sup_{x \in [x_{j-1}, x_j]} f(x)$$
>
> and
>
> $$s(f, \mathcal{D}) = \sum_{j=1}^{n} (x_j - x_{j-1}) \inf_{x \in [x_{j-1}, x_j]} f(x)$$
>
> respectively.

These two definitions are simply the sums of the (signed) areas of the rectangles formed by taking one side as each interval along the real axis, and extending the other to the largest and lowest points reached by the function over the corresponding interval.

Clearly $S(f, \mathcal{D}) \geq s(f, \mathcal{D})$.

> **Lemma 5.2.** *If $\mathcal{D}$ and $\mathcal{D}'$ are dissections with $\mathcal{D}' \supseteq \mathcal{D}$. Then*
>
> $$S(f, \mathcal{D}) \geq S(f, \mathcal{D}') \geq s(f, \mathcal{D}') \geq s(f, \mathcal{D})$$

Proof. Suppose $\mathcal{D}'$ has a single extra point, say $y \in (x_{r-1}, x_r)$.

Note that

$$\sup_{x \in [x_{r-1}, y]} f(x), \sup_{x \in [y, x_r]} f(x) \leq \sup_{x \in [x_{r-1}, x_r]} f(x)$$

Hence

$$(x_r - x_{r-1}) \inf_{x \in [x_{r-1}, x_r]} f(x) \geq (x_r - y) \inf_{x \in [x_{r-1}, y]} f(x) + (x_r - y) \inf_{x \in [y, x_r]} f(x)$$

Therefore, $S(f,\mathcal{D}) \geq S(f,\mathcal{D}')$. Then if there is more than one point difference, repeat this process, adding each individual point one at a time.

The central inequality is true, as noted above. The proof for the final inequality is totally analogous. □

Lemma 5.3. *If $\mathcal{D}_1$ and $\mathcal{D}_2$ are any two dissections, then*

$$S(f,\mathcal{D}_1) \geq S(f,\mathcal{D}_1 \cup \mathcal{D}_2) \geq s(f,\mathcal{D}_1 \cup \mathcal{D}_2) \geq s(f,\mathcal{D}_2)$$

Proof. Since $\mathcal{D}_1 \cup \mathcal{D}_2 \supseteq \mathcal{D}_1, \mathcal{D}_2$, the two outer inequalities must hold, whilst the central inequality was already known. □

The important result here is that the upper sum of any dissection is greater than or equal to the lower sum of any other dissection.

Definition 5.4. The *upper and lower integrals* are defined as

$$\begin{aligned} I^\star(f) &= \inf_{\mathcal{D}} S(f,\mathcal{D}) \\ I_\star(f) &= \sup_{\mathcal{D}} s(f,\mathcal{D}) \end{aligned}$$

respectively, where the infimum and supremum are both taken over all dissections of the interval.

Remark. These are both well-defined because f is bounded by K; i.e. $S(f,\mathcal{D}) \geq -K(b-a)$ and $s(f,\mathcal{D}) \leq K(b-a)$ for all $\mathcal{D}$. Also, we know $I_\star(f) \leq I^\star(f)$.

Definition 5.5. We say a bounded function f is *Riemann integrable*, or in this course simply *integrable*, if

$$I^\star(f) = I_\star(f)$$

In this case we write

$$\int_a^b f(x)\,\mathrm{d}x = I^\star(f) = I_\star(f)$$

or simply

$$\int_a^b f$$

Remark. It is useful to extend the notation slightly to allow $\int_a^a f = 0$, and $\int_b^a f = -\int_a^b f$.

Having defined the integral, we should clearly investigate which functions are Riemann integrable. Firstly, we should establish that there are certainly bounded functions which are not Riemann integrable.

Example 5.6 (due to Dirichlet)**.** We define the (evidently bounded) function $f : [0,1] \to \mathbb{R}$

$$f(x) = \begin{cases} 1 & x \in \mathbb{Q} \cap [0,1] \\ 0 & x \text{ irrational on } [0,1] \end{cases}$$

Then for any partition $\mathcal{D}$, the upper sum is 1 whilst the lower sum is 0, since every integral contains both a rational and irrational point. Thus $I^\star = 1$ and $I_\star = 0$, so f is not Riemann integrable.

However, there are two very broad classes of function which are Riemann integrable, namely monotonic and continuous functions. To prove these, we establish a useful criterion for integrability.

Theorem 5.7 (Riemann)**.** *A bounded function $f : [a,b] \to \mathbb{R}$ is integrable iff given $\epsilon > 0$ there is a dissection $\mathcal{D}$ such that $S(f,\mathcal{D}) - s(f,\mathcal{D}) < \epsilon$.*

Proof.

$\Longrightarrow$: Assume f is Riemann integrable. Then given $\epsilon > 0$, by the definition of $I^\star(f)$ there is some dissection $\mathcal{D}_1$ such that $S(f,\mathcal{D}_1) - I^\star < \frac{\epsilon}{2}$, and similarly $\mathcal{D}_2$ such that $s(f,\mathcal{D}_2) - I_\star > \frac{\epsilon}{2}$. The result follows on taking $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$, since $I^\star = I_\star$, and using the inequalities proved above.

$\Longleftarrow$: Now assume that for any $\epsilon > 0$ we have such a dissection. Then we have $0 \leq I^\star(f) - I_\star(f) < \epsilon$ for all $\epsilon > 0$, so $I^\star = I_\star$.

$\square$

We can now show that all monotonic and continuous functions are integrable, noting that both are automatically bounded.

Theorem 5.8. *If $f : [a,b] \to \mathbb{R}$ is monotonic, then f is integrable on $[a,b]$.*

Proof. Assume, without loss of generality, that f is increasing.

Take $\mathcal{D}$ to be any partition of $[a,b]$. Since f is increasing,

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

and

$$s(f,\mathcal{D}) \quad = \quad \sum_{j=1}^{n} f(x_{j-1})(x_j - x_{j-1})$$

Thus $S(f,\mathcal{D}) - s(f,\mathcal{D}) = \sum_{j=1}^{n}[f(x_j) - f(x_{j-1})](x_j - x_{j-1})$.
Now consider $\mathcal{D}_n = \left\{a, a+\frac{b-a}{n}, \cdots, a+\frac{(b-a)(n-1)}{n}, b\right\}$ for $n > 0$. Clearly,

$$\begin{aligned} S(f,\mathcal{D}) - s(f,\mathcal{D}) &= \frac{b-a}{n}\sum_{j=1}^{n}[f(x_j) - f(x_{j-1})] \\ &= \frac{b-a}{n}[f(b) - f(a)] \end{aligned}$$

Hence given ϵ, we can always take n sufficiently large that this is less than ϵ. Thus by Riemann's criterion, we are done. □

To establish the result for continuous functions, we need uniform continuity (see Lemma 2.9) that we can find some $\delta(\epsilon)$ such that whenever x and y lie in the function's domain, and $|x-y| < \delta$, we have $|f(x) - f(y)| < \epsilon$.

Theorem 5.9. *If $f : [a,b] \to \mathbb{R}$ is continuous, then f is integrable.*

Proof. Consider the partition $\mathcal{D}$ with points $x_j = \frac{b-a}{n}j + a$ for $0 \leq j \leq n$ (where $n > 0$ is an integer). Then we have

$$S(f,\mathcal{D}) - s(f,\mathcal{D}) = \frac{b-a}{n}\sum_{j=1}^{n}\left(\sup_{x\in[x_{j-1},x_j]} f(x) - \inf_{x\in[x_{j-1},x_j]} f(x)\right)$$

Let $\epsilon > 0$ be given, and consider the corresponding δ given by uniform continuity. Now take n large enough so that $\frac{b-a}{n} < \delta$. Then $\forall x, y \in [x_{j-1}, x_j]$, $|f(x) - f(y)| < \epsilon$, and (e.g. because bounds are achieved for continuous functions on closed intervals) it follows that

$$\begin{aligned} S(f,\mathcal{D}) - s(f,\mathcal{D}) &< \frac{b-a}{n}\sum_{j=1}^{n}\epsilon \\ &= \epsilon(b-a) \end{aligned}$$

Hence, by Riemann's criterion, we are done. □

Having established that not all functions are integrable, but that monotonic and continuous functions always are, it is useful to have an example of a function which does not fit into either of these categories, but which is still integrable. In fact, we shall modify Example 5.6.

Example 5.10. $f(x) = \begin{cases} \frac{1}{q} & \text{if } x = \frac{p}{q} \in [0,1] \text{ in its lowest form} \\ 0 & \text{if } x \text{ is irrational} \end{cases}$

Before we attempt to prove integrability of this function, it is worth considering what we would expect the actual value of any such integral to be. Since $\frac{1}{q}$ is in general 'small' for the 'vast majority' of rational numbers, we expect that there is near no area under the corresponding graph, and hence that the integral should be 0. In fact, this approach leads fairly naturally to a method of proof.

Clearly $s(f, \mathcal{D}) = 0$ for $\mathcal{D}$ any partition, so $I_\star = 0$. Therefore, by Riemann's criterion, it is sufficient to find some $\mathcal{D}_\epsilon$ for all $\epsilon > 0$ such that $S(f, \mathcal{D}) < \epsilon$.

Now let $\epsilon > 0$ be given, and take $N \in \mathbb{N}$ such that $\frac{1}{N} < \epsilon$. Then construct $S = \left\{x \in [0,1] : f(x) \geq \frac{1}{N}\right\}$ - this is evidently a finite set, with cardinality R_N. (This is analogous to the statement that the set of 'large' $\frac{1}{q}$ is 'small', i.e. finite.)

Hence we can choose some partition $\mathcal{D}_\epsilon$ such that the $t_j \in S$ belong to intervals $[x_{k-1}, x_k]$ with length less than $\frac{\epsilon}{R}$.

Then $S(f, \mathcal{D}) < R \times \frac{\epsilon}{R} + \epsilon = 2\epsilon$, so we are done.

A similar argument to the above is used in the proof of the statement below about the alteration of finitely many points in $[a,b]$ in an integrable function has no effects on integrability, or indeed the value of the integral.

Proposition 5.11. *In the below, we take f, g to be bounded, integrable functions on $[a,b]$.*

(i) *If $f \leq g$ on $[a,b]$, $\int_a^b f \leq \int_a^b g$.*

(ii) *$f + g$ is integrable on $[a,b]$ and $\int (f+g) = \int f + \int g$. (We take all unlabeled integrals here to be from f to g.)*

(iii) *For any constant k, kf is integrable, and $\int kf = k \int f$.*

(iv) *$|f|$ is integrable, and $\left|\int f\right| \leq \int |f|$.*

(v) *The product fg is integrable.*

(vi) *If $f(x) = h(x)$ on $[a,b]$ except at finitely many points, then h is also integrable and $\int f = \int h$.*

(vii) *If $a < c < b$, then f is integrable on $[a,c]$ and $[c,b]$ and $\int_a^b f = \int_a^c f + \int_c^b f$.*

Proof.

(i) If $f \leq g$, then $\int_a^b f = I^\star(f) \leq S(f, \mathcal{D}) \leq S(g, \mathcal{D})$ for all partitions $\mathcal{D}$. Now taking the infimum, we obtain $\int f \leq I^\star(g) = \int g$.

(ii) Observe that the supremum $\sup(f+g)(x) \leq \sup f(x) + \sup g(x)$ over any interval; hence $S(f+g, \mathcal{D}) \leq S(f, \mathcal{D}) + S(g, \mathcal{D})$. Similarly, for any two partitions,

$$\begin{aligned} I^\star(f+g) &\leq S(f+g, \mathcal{D}_1 \cup \mathcal{D}_2) \\ &\leq S(f, \mathcal{D}_1 \cup \mathcal{D}_2) + S(g, \mathcal{D}_1 \cup \mathcal{D}_2) \\ &\leq S(f, \mathcal{D}_1) + S(g, \mathcal{D}_2) \end{aligned}$$

But then, taking the infimum again, $I^{\star}(f+g) \leq I^{\star}(f) + I^{\star}(g)$. A similar argument gives $I^{\star}(f+g) \geq I^{\star}(f) + I^{\star}(g)$, and we are done.

(iii) Left as an exercise.

(iv) Let $f_{+}(x) = \max\{f(x), 0\}$. We aim to show this is integrable.

Note $\sup f_{+} - \inf f_{+} \leq \sup f - \inf f$. Thus for all $\mathcal{D}$, $S(f_{+}, \mathcal{D}) - s(f_{+}, \mathcal{D}) \leq S(f, \mathcal{D}) - s(f, \mathcal{D})$. Hence, by Riemann's criterion, we are done.

Similarly, $f_{-}(x) = \min\{f(x), 0\}$ is integrable.

But now $|f(x)| = f_{+}(x) - f_{-}(x)$, and hence by the above properties, $|f(x)|$ is integrable. (This decomposition is often useful.)

In addition, note $-|f| \leq f \leq |f|$, so by the property above, $-\int |f| \leq \int f \leq \int |f|$, and hence

$$\left|\int f\right| \leq \left|\int |f|\right| = \int |f|$$

(v) First we show that f^2 is integrable; suppose $f \geq 0$. But f is integrable, so for $\epsilon > 0$ we have $\mathcal{D}$ with $S(f, \mathcal{D}) - s(f, \mathcal{D}) < \epsilon$. Let

$$M_j = \sup_{x \in [x_{j-1}, x_j]} f(x)$$

and let

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

But $\sup_{I_j} f^2(x) - \inf_{I_j} f^2(x) = M_j^2 - m_j^2$ so if $f \leq K$ is the bound on f,

$$\begin{aligned} S\left(f^2, \mathcal{D}\right) - s\left(f^2, \mathcal{D}\right) &= \sum \left(M_j^2 - m_j^2\right)(x_j - x_{j-1}) \\ &= \sum (M_j + m_j)(M_j - m_j)(x_j - x_{j-1}) \\ &\leq 2K \sum (M_j - m_j)(x_j - x_{j-1}) \\ &< 2K\epsilon \end{aligned}$$

so we have established that if $f \geq 0$, f^2 is integrable.

But now if f is any integrable function, we know $|f| \geq 0$ is integrable, and hence $|f|^2 = f^2$ is integrable.

Now to see fg is integrable, simple note

$$(f+g)^2 - (f-g)^2 = 4fg$$

and by the above results we are done.

(vi) Let $h = f - F$. Then $h(x) = 0$ everywhere except at some finitely many x. Then by the definition of the Riemann integral, using a simple argument, we have $\int h = 0$; then $\int h = \int (f - F) = 0$ so $\int f = \int F$ by the above.

(**vii**) Left as an exercise; this is readily seen by placing boundary points in the dissections at c.

□

5.2 Computation of Integrals

Now that we have some idea of the integrability of many functions, it becomes necessary to develop efficient ways to compute the actual value of the integral.

We assume $f : [a, b] \to \mathbb{R}$ is a bounded, integrable function, and write $F(x) = \int_a^x f(t)\,\mathrm{d}t$, for $x \in [a, b]$.

Theorem 5.12. *F is continuous.*

Proof. Let $x \in [a, b]$; then $F(x+h) - F(x) = \int_x^{x+h} f(t)\,dt$.

Therefore, letting $|f|$ be bounded by K,

$$|F(x+h) - F(x)| \leq \int_x^{x+h} |f(t)|\,\mathrm{d}t \leq K|h|$$

Then letting $h \to 0$, $F(x+h) \to F(x)$. □

Theorem 5.13 (The Fundamental Theorem of Calculus)**.** *If f is a continuous function, then F is differentiable, and $F\prime(x) = f(x)$ for all $x \in [a, b]$.*

Proof. This is actually a reasonably straightforward computation:

$$\begin{aligned}
\left|\frac{F(x+h) - F(x)}{h} - f(x)\right| &= \frac{1}{|h|}|F(x+h) - F(x) - hf(x)| \\
&= \frac{1}{|h|}\left|\int_x^{x+h} f(t)\,\mathrm{d}t - hf(x)\right| \\
&= \frac{1}{|h|}\left|\int_x^{x+h} [f(t) - f(x)]\,\mathrm{d}t\right| \\
&\leq \frac{1}{|h|}\left|\int_x^{x+h} |f(t) - f(x)|\,\mathrm{d}t\right| \\
&\leq \frac{1}{|h|}\left|\int_x^{x+h} \max_{\theta\in[0,1]} |f(x+\theta x) - f(x)|\,\mathrm{d}t\right| \\
&\leq \max_{\theta\in[0,1]} |f(x+\theta x) - f(x)|
\end{aligned}$$

Now let $h \to 0$; by the continuity of f, the right-hand side tends to 0, and hence $F'(x) = f(x)$. □

Corollary 5.14 (Integration is anti-differentiation). *If $f = g'$ is continuous on $[a, b]$, then*

$$\int_a^x f(t)\, dt = g(x) - g(a)$$

for all $x \in [a, b]$.

Proof. From the above, $F'(x) = f(x)$; hence $(F - g)'(x) = F'(x) - g'(x) = 0$. Hence, by the continuity of F and g, $F - g$ is a constant.

Hence $F(x) - g(x) = F(a) - g(a) = 0 - g(a)$.

Thus $\int_a^x f = g(x) - g(a)$. □

Therefore, if we know a *primitive* of *anti-derivative* for f (i.e. a function such that $g' = f$) we can compute $\int_a^x f$. By the above, we know that primitives of continuous functions always exist, and they differ only up to a constant.

The following are all immediate corollaries of the above results, but prove incredibly useful in practice.

Theorem 5.15 (Integration by parts). *Suppose f' and g' exist and are continuous on $[a, b]$. Then*

$$\int_a^b f'g = f(b)\, g(b) - f(a)\, g(a) - \int_a^b fg'$$

Proof. By the product rule, $(fg)' = f'g + fg'$. Thus $\int_a^b (f'g + fg') = [fg]_a^b$, so by linearity we get the required result. □

Theorem 5.16 (Integration by substitution). *Let $g : [\alpha, \beta] \to [a, b]$ be some map such that $g(\alpha) = a$ and $g(\beta) = b$, with g' being a continuous function on $[\alpha, \beta]$. Then for a continuous function $f : [a, b] \to \mathbb{R}$. Then*

$$\int_a^b f(x)\, dx = \int_\alpha^\beta f(g(t))\, g'(t)\, dt$$

Proof. This is a consequence of the chain rule, in much the same way as the above is a consequence of the product rule.

Let $F(x) = \int_a^x f(t)\, dt$, and $h(t) = F(g(t))$ - note this is well-defined as $g(t) \in [a, b]$, the domain of F. Then

$$h'(t) = g'(t)\, F'(g(t)) = g'(t)\, f(g(t))$$

and hence by the above

$$\begin{aligned}
\int_{\alpha}^{\beta} h' &= \int_{\alpha}^{\beta} f\left(g\left(t\right)\right) g'\left(t\right) \mathrm{d}t \\
&= \left[h\right]_{\alpha}^{\beta} \\
&= F\left(g\left(\beta\right)\right) - F\left(g\left(\alpha\right)\right) \\
&= F\left(b\right) - F\left(a\right) \\
&= \int_{a}^{b} f\left(t\right) \mathrm{d}t
\end{aligned}$$

□

Remark. Note that there are no restrictions on g other than the need for a continuous derivative and the fact that its range is contained within the domain that f is known to be 'well-behaved' upon.

5.3 Mean Values and Taylor's Theorem Again, Improper Integrals and the Integral Test

5.3.1 The mean value theorem and Taylor's theorem

We can derive a third form of the remainder in Taylor's Theorem using the tools developed up to this point, with the stronger assumption that $f^{(n)}$ be continuous (as opposed to only existing) on the given interval.

Theorem 5.17 (Taylor's Theorem with Integral Remainder)**.** *Let $f^{(n)}$ be continuous for $x \in [0, h]$. Then*

$$f\left(h\right) = f\left(0\right) + f'\left(0\right) h + \cdots + \frac{f^{(n-1)}\left(0\right) h^{n-1}}{(n-1)!} + R_n$$

where

$$R_n = \frac{h^n}{(n-1)!} \int_0^1 \left(1-t\right)^{n-1} f^{(n)}\left(th\right) dt$$

Proof. Make the substitution $u = th$, and then proceed as follows:

$$\begin{aligned}
R_n &= \frac{h^n}{(n-1)!} \int_0^h \left(1 - \frac{u}{h}\right)^{n-1} f^{(n)}\left(u\right) \frac{\mathrm{d}u}{h} \\
&= \frac{1}{(n-1)!} \int_0^h \left(h-u\right)^{n-1} f^{(n)}\left(u\right) \mathrm{d}u \\
&= \frac{1}{(n-1)!} \left[\left[\left(h-u\right)^{n-1} f^{(n-1)}\left(u\right)\right]_0^h - \int_0^h -\left(n-1\right)\left(h-u\right)^{n-2} f^{(n-1)}\left(u\right) \mathrm{d}u\right] \\
&= \frac{-h^{n-1}}{(n-1)!} f^{(n-1)}\left(0\right) + \frac{1}{(n-2)!} \int_0^h \left(h-u\right)^{n-2} f^{(n-1)}\left(u\right) \mathrm{d}u \\
&= \frac{-h^{n-1}}{(n-1)!} f^{(n-1)}\left(0\right) + R_{n-1}
\end{aligned}$$

Then we recursively obtain the desired result,

$$\begin{aligned} R_n &= -\frac{h^{n-1}}{(n-1)!} f^{(n-1)}(0) - \cdots - hf'(0) + \int_0^h f'(u)\,\mathrm{d}u \\ &= -\frac{h^{n-1}}{(n-1)!} f^{(n-1)}(0) - \cdots - hf'(0) + f(h) - f(0) \end{aligned}$$

□

This integral form of the remainder (for continuous $f^{(n)}$) also gives the Cauchy and Lagrange forms, as shown below.

Proposition 5.18. *For any continuous function f defined on $[a, b]$, there is a point $c \in (a, b)$ such that*

$$\int_a^b f(x)\, dx = f(c)[b - a]$$

Proof. This is immediate from the mean value theorem applied to $F(x) = \int_a^x f(t)\,\mathrm{d}t$. □

Corollary 5.19. *Cauchy's form of the remainder in Taylor's theorem (for continuous f).*

Proof. Applying the mean value theorem we immediately get $R_n = \frac{h^n}{(n-1)!}(1-\theta)^{n-1} f^{(n)}(\theta h)[1-0]$, $\theta \in (0, 1)$, which is Cauchy's form. □

Theorem 5.20 (Mean Value Theorem for Integrals)**.** *If $f, g : [a, b] \to \mathbb{R}$ are continuous functions with $g(x) \geq 0$ everywhere in the interval, then*

$$\int_a^b f(x)\, g(x)\, dx = f(c) \int_a^b g(x)\, dx$$

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

Proof. Since f is continuous, it has finite bounds on the interval $[a, b]$, say $m \leq f(x) \leq M$ for all x there. Hence

$$m \int_a^b g(x)\,\mathrm{d}x \leq \int_a^b f(x)\, g(x)\,\mathrm{d}x \leq M \int_a^b g(x)\,\mathrm{d}x$$

since, for example, $f(x)\, g(x) - mg(x) = [f(x) - m]\, g(x) \geq 0$.

Therefore, letting $I = \int_a^b g(x)\,\mathrm{d}x$, we have

$$m \leq \frac{1}{I} \int_a^b f(x)\, g(x)\,\mathrm{d}x \leq M$$

But now since f attains its bounds, i.e. f is a surjective map $f : [a, b] \to [m, M]$, we can apply the intermediate value theorem to obtain that there is some $c \in (a, b)$ with

$$f(c) = \frac{1}{I} \int_a^b f(x) g(x) \, \mathrm{d}x$$

as required. $\square$

Corollary 5.21. *Lagrange's form of the remainder in Taylor's theorem (for continuous f).*

Proof. We have

$$R_n = \frac{h^n}{(n-1)!} \int_0^1 (1-t)^{n-1} f^{(n)}(th) \, \mathrm{d}t$$

so therefore there is some $\theta \in (0, 1)$ such that

$$\begin{aligned} R_n &= \frac{h^n}{(n-1)!} f^{(n)}(\theta h) \int_0^1 (1-t)^{n-1} \, \mathrm{d}t \\ &= \frac{h^n}{(n-1)!} f^{(n)}(\theta h) \frac{1}{n} \\ &= \frac{h^n}{n!} f^{(n)}(\theta h) \end{aligned}$$

as required. $\square$

5.3.2 Infinite (improper) integrals and the integral test

Suppose $f : (a, \infty] \to \mathbb{R}$ is bounded and integrable on $[a, R]$ $\forall R \geq a$.

Definition 5.22. If $\lim_{R\to\infty} \int_a^R f(x) \, \mathrm{d}x = l$ exists, we say that the *improper integral* (*of the first kind)*

$$\int_a^\infty f(x) \, \mathrm{d}x = l$$

exists, or *converges*. If there is no such limit, we say the integral *diverges*.

Example 5.23. Consider $\int_1^\infty \frac{1}{x^k} \mathrm{d}x$, for $k \neq 1$.

Note that $\int_1^R \frac{1}{x^k} \mathrm{d}x = \frac{R^{1-k}-1}{1-k}$ if $k \neq 1$. Then as $R \to \infty$, this converges iff $k > 1$ - in the case it does converge, we have

$$\int_1^\infty \frac{1}{x^k} \mathrm{d}x = \frac{1}{k-1}$$

(Of course, if $k = 1$ the integral is $\ln R \to \infty$, so the integral diverges in this case also.)

It is perhaps suggestive that this integral shares properties with the infinite series $\sum \frac{1}{x^k}$, and this seems natural since summing and integration are closely related processes. We shall in fact establish a result to confirm this relationship below.

Proposition 5.24. *If $f \geq 0$ and $g \geq 0$ for $x \geq a$ are both integrable functions, and $f(x) \leq kg(x)$ for some constant k, then:*

- *if $\int_a^\infty g$ converges, so does $\int_a^\infty f$; and in this case*
- *$\int_a^\infty f \leq \int_a^\infty g$.*

Proof. If $f(x) \leq kg(x)$, then

$$\int_a^R f(x)\,\mathrm{d}x \leq k\int_a^R g(x)\,\mathrm{d}x$$

.

Since the map $R \mapsto \int_a^R f(x)\,\mathrm{d}x$ is increasing, and the right hand side converges, the left hand side must also converge, and to a limit not exceeding the limit of the right-hand side. □

Note that whilst the above analogue of the comparison test holds for integration, it is *not* true that the convergence of $\int_a^\infty f(x)\,\mathrm{d}x$ implies that $f \to 0$ as $x \to \infty$, as the following example shows:

Example 5.25. Let f be the function such that if $|x-n| \leq \frac{1/2}{(n+1)^2}$ for some integer n, then $f(x) = 1$; otherwise, $f(x) = 0$.

Then since the sum $\sum \frac{1}{(n+1)^2}$ converges, the infinite integral $\int_1^\infty f$ must also converge. But clearly $f(n) = 1$ for all integer n, so $f \not\to 0$.

Note that this function consists of discontinuous, narrowing rectangles at the integers; one could also easily construct a continuous version by making these triangles.

Theorem 5.26 (The integral test). *Let $f : (1,\infty] \to \mathbb{R}_{\geq 0}$ be a positive, decreasing function. Then*

(i) *$\int_1^\infty f(x)\,dx$ and $\sum_1^\infty f(n)$ both converge or diverge; and*

(ii) *As $n \to \infty$, $\sum_1^n f(k) - \int_1^n f(x)\,dx$ tends to some limit l, with $0 \leq l \leq f(1)$.*

Proof. Note that since f is decreasing, it is Riemann integrable on every interval $[1,R]$.

Now if $x \in [n-1,n]$, $f(n-1) \geq f(x) \geq f(n)$, and hence

$$f(n-1) \geq \int_{n-1}^n f(x)\,\mathrm{d}x \geq f(n)$$

Summing, we obtain

$$\sum_{k=1}^{n-1} f(k) \geq \int_1^n f(x)\,\mathrm{d}x \geq \sum_{k=2}^{n} f(k)$$

(i) Suppose $\int_1^\infty f(x)\,\mathrm{d}x$ converges; then $\int_1^n f(x)\,\mathrm{d}x$ is bounded in n, so by the above $\sum f(k)$ is bounded, and since it is increasing it must converge.

Similarly, if the sum $\sum f(k)$ converges, its partial sums are bounded, so the increasing function $\int_1^x f(t)\,\mathrm{d}t$ is bounded and hence converges.

(ii) Let $\phi(n) = \sum_{k=1}^{n} f(r) - \int_1^n f(x)\,\mathrm{d}x$ be the discrepancy between these two quantities. Then

$$\phi(n) - \phi(n-1) = f(n) - \int_{n-1}^{n} f(x)\,\mathrm{d}x \leq 0$$

as f is decreasing, as noted above, and so $\phi(n)$ is *decreasing*. But then from the double inequality, we have $0 \leq \phi(n) \leq f(1)$. So we have established $\phi(n)$ is bounded and decreasing, and hence it converges to some limit $l \in [0, f(1)]$.

□

We have already seen one example of this above, with $f(x) = \frac{1}{x^k}$. Another follows:

Example 5.27. $\sum_2^\infty \frac{1}{n \log n}$. (We have previously treated this via the Cauchy condensation test.)
Let $f(x) = \frac{1}{x \log x}$. Then $\int_2^R \frac{1}{x \log x}\mathrm{d}x = \log(\log R) - \log(\log 2) \to \infty$ as $R \to \infty$.

It also has application in the asymptotic analysis of infinite sums.

Corollary 5.28 (Euler's constant)**.** $1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} - \log n \to \gamma$ *as* $n \to \infty$*, with* $\gamma \in [0,1]$.

Proof. This is simply obtained from setting $f(x) = \frac{1}{x}$ in the integral test. □

Remark. The actual value of the Euler[–Mascheroni] constant is around 0.577, but it is not even known whether it is irrational.

We should note that the definition of improper integrals for $\int_a^\infty$ can be extended to both negative infinite lower bounds:

Definition 5.29. $\int_{-\infty}^a f = \lim_{R\to\infty} \int_{-R}^a f$ if this limit exists.

and also to integrals over the whole real line, as follows:

Definition 5.30. $\int_{-\infty}^\infty f$ *converges* if *both* $\int_{-\infty}^a f = l_-$ and $\int_a^\infty f = l_+$ converge; then

$$\int_{-\infty}^{\infty} f = l_- + l_+$$

Remark. The a taken is arbitrary.

The important this to note about this definition is that it is *not* equivalent to stating that $\lim_{R\to\infty} \int_{-R}^R f(x)\,\mathrm{d}x$ exists, since this would allow, for instance, all odd functions like $f(x) = x$ to be integrable in this manner.

Finally, we introduce one more exception to the rule, in allowing certain unbounded functions f to be said to be integrable:

Definition 5.31. If f is unbounded at b but is otherwise integrable on $[a,b]$ (i.e. on all intervals within $[a,b-\epsilon]$ where $\epsilon>0$ for $b-\epsilon\geq a$), then we define the *improper integral of the second kind*

$$\int_a^b f=\lim_{\epsilon\to0}\int_a^{b-\epsilon} f$$

if this limit exists. (Note that we can then extend this definition to multiple points of unboundedness at arbitrary positions within the interval.)

Example 5.32. If we attempt to calculate the integral $\int_0^1 f$ via anti-differentiation, where $f(x)=\frac{1}{\sqrt{x}}$, we will fail; however, if $\delta\in(0,1)$,

$$\int_\delta^1 f(x)\,\mathrm{d}x=2\left(\sqrt{1}-\sqrt{\delta}\right)$$

and as $\delta\to0$, $\int_\delta^1 f(x)\,\mathrm{d}x\to2$, so we write

$$\int_0^1 f(x)\,\mathrm{d}x=2$$

for short, using the above definition at the lower limit.

6 * Riemann Series Theorem

The interesting result alluded to above, namely that any real series which is conditionally convergent (and hence *not* absolutely convergent - see Theorem 1.23) can be re-arranged to give any real sum desired, is actually relatively straightforward to prove, and so the elementary proof follows:

Theorem 6.1 (Riemann Series Theorem). *If $\sum_{n=1}^{\infty} a_n$ is a conditionally convergent real series, then there is a permutation σ of the natural numbers such that $\sum_{n=1}^{\infty} a_{\sigma(n)}$*

(i) *converges to any real number M;*

(ii) *tends to positive or negative infinity;*

(iii) *fails to approach any limit, finite or infinite.*

Proof. Define the two complementary sequences $a_n^+ = \frac{a_n + |a_n|}{2} = \max\{a_n, 0\}$ and $a_n^- = \frac{a_n - |a_n|}{2} = \min\{a_n, 0\}$.

Since the original series is *conditionally* convergent, the two series $\sum a_n^+$ and $\sum a_n^-$ must diverge. (If both converge, the series is absolutely convergent; if one does, the series is not convergent at all.)

Now let $M > 0$ be a given real number. Take sufficiently many a_n^+ terms so that $\sum_{n=1}^{p} a_n^+ > M$ but $\sum_{n=1}^{p-1} a_n^+ \leq M$; this is possible because the series tends to positive infinity.

Ignoring 0 terms in a_n^+, leaving us with m_1 strictly positive terms from the original sequence, we write

$$\sum_{n=1}^{p} a_n^+ = a_{\sigma(1)} + \cdots + a_{\sigma(m_1)}$$

where $a_{\sigma(j)} > 0$ and $\sigma(1) < \cdots < \sigma(m_1) = p$. We can append values skipped where $a_n = 0$ at this point, since it is irrelevant where they are included in the series.

Now we repeat, adding just enough further terms (say q of them) found in a_n^- such that the sum is pushed back below M,

$$\sum_{n=1}^{p} a_n^+ + \sum_{n=1}^{q} a_n^- < M$$

whilst

$$\sum_{n=1}^{p} a_n^+ + \sum_{n=1}^{q-1} a_n^- \geq M$$

Then one has

$$\begin{aligned}\sum_{n=1}^{p} a_n^+ + \sum_{n=1}^{q} a_n^- \quad = \quad & \underbrace{a_{\sigma(1)} + \cdots + a_{\sigma(m_1)}}_{\text{positive terms}} + \underbrace{a_{\sigma(m_1+1)} + \cdots + a_{\sigma(m_2)}}_{\text{actual zero terms}} \\ & + \underbrace{a_{\sigma(m_2+1)} + \cdots + a_{\sigma(m_3)}}_{\text{negative terms}} + \underbrace{a_{\sigma(m_3+1)} + \cdots + a_{\sigma(m_4)}}_{\text{actual zero terms}}\end{aligned}$$

and so on. Note that σ is injective (a_1 is either equal to $a_{\sigma(1)}$, $a_{\sigma(m_1+1)}$ or $a_{\sigma(m_2+1)}$ according to whether $a_1 < 0$, $a_1 = 0$ or $a_1 > 0$, and so on similarly).

Now note that because we always stop our individual stages at 'just the right point', the difference of the partial sum and M is never more than $\left|a_{\sigma(n)}\right|$ where n is at least as recent as the last 'change in direction', and hence since the series $\sum a_n$ converges, and thus $\left|a_{\sigma(n)}\right| \to 0$, and we are done.

The same method can be used to construct series tending to 0 or negative limits.

To construct infinite limits, one need only change the stopping condition to that of having exceeded some increasing target.

To construct divergent series, one need only choose two different stopping conditions for the positive and negative parts. □

Corollary 6.2. *If $\sum^{\infty} a_n$ is a complex series, then either:*

(i) it converges absolutely;

(ii) it can be rearranged to sum to any value on a line $L = \{a + tb : t \in \mathbb{R}\}$ in the complex plane ($a, b \in \mathbb{C}$, $b \neq 0$), but no other finite values; or

(iii) it can be rearranged to sum to any value in the complex plane.

More generally, given a convergent series of vectors in a finite dimensional vector space, the set of sums of converging rearranged series is an (affine) subspace of the vector space. (Affine because it does not necessarily contain the origin.)

7 * Measure Theory

We have already shown a necessary and sufficient condition for integrability in the form of Riemann's criterion (see Theorem 5.7), but this is not particularly enlightening in that we have no real grasp of what Riemann's criterion gives. The following theorem gives a complete understanding of which functions are integrable:

Theorem 7.1. *A function $f : \mathbb{R} \to \mathbb{R}$ is integrable iff the set of discontinuities of f has measure zero.*

Obviously, we first need to introduce the following definition:

Definition 7.2. A set $S \subset \mathbb{R}$ has *measure zero* if, for any positive $\epsilon > 0$, there is a set of open intervals A_i covering S - that is $(\cup_i A_i) \cap S = S$ - such that $\sum_i \text{length}\,(A_i) < \epsilon$.

That is, if S has measure zero, we can trap the elements in an arbitrarily small collection of intervals. This seems to fit in well with the ideas explored above, where we saw that finite collections of discontinuities were acceptable precisely by trapping them in narrow bands.

Example 7.3. A countable union of measure zero sets is a measure zero set.

Proof. Coming soon! □