

AN INTRODUCTION TO REAL ANALYSIS

MARK CRADDOCK

ABSTRACT. These are the lecture notes for the subject 35007 Real Analysis, taught at the University of Technology, Sydney. They are based on a number of sources, including notes by Gordon Mclelland. They are not for profit.

CONTENTS

1. Introductory Real Analysis	3
1.1. Countability	4
1.2. Sets, Real numbers	6
1.3. The Least Upper Bound Axiom	6
1.4. Limits and Sequences	7
1.5. Infinite Series and Convergence Tests.	12
2. Continuous Functions and their Properties.	17
2.1. Limits	17
2.2. The Definition of Continuity	18
2.3. Uniform Continuity	20
2.4. Maxima, Minima and the Intermediate Value Property	21
3. Differentiation and its Applications	24
3.1. The Derivative	24
3.2. Maxima and Minima.	28
3.3. Mean Value Theorems	29
3.4. Inverse Functions	31
3.5. Convex Functions	33
3.6. Power Series and Taylor Expansions	36
4. The Riemann Integral	42
4.1. Calculating Integrals By Riemann Sums	43
4.2. The Fundamental Theorem of Calculus	46
4.3. Integration Rules	48
4.4. Improper Riemann Integrals	49
5. Sequences of Functions	51
5.1. Pointwise and Uniform Convergence	51
5.2. The Weierstrass M-test	55
5.3. Swapping Limits and Integrals	57
5.4. Swapping Limits and Derivatives	59

5.5. The Weierstrass Approximation Theorem

1. INTRODUCTORY REAL ANALYSIS

Let us review a few important ideas introduced in a typical first course in analysis. Analysis is largely concerned with the behaviour of functions, which we usually want to be continuous. However in order to develop a useful theory of the behaviour of functions, we first need to study sequences. An investigation of sequences and functions leads to the major tools of elementary calculus, namely the derivative and integral.

We begin with one of the most basic concepts in mathematics.

1.0.1. *Sets.* Particularly in the theory of measure, we are required to manipulate sets. So here we review some elementary facts.

Definition 1.1. Let A and B be sets.

- (i) The union of A and B is denoted $A \cup B$ and is given by

$$A \cup B = \{x : x \in A \text{ or } B\}.$$

The union of a collection of sets $A_i, i \in \mathbb{N}$ is defined inductively and denoted $\cup_{i \in \mathbb{N}} A_i$.

- (ii) The intersection of A and B is denoted $A \cap B$ and is given by

$$A \cap B = \{x : x \in A \text{ and } B\}.$$

The intersection of a collection of sets $A_i, i \in \mathbb{N}$ is defined inductively and denoted $\cap_{i \in \mathbb{N}} A_i$.

- (iii) The set difference of A and B is denoted $A - B$ and is given by

$$A - B = \{x \in A, x \notin B\}.$$

- (iii) We say that A is a subset of B and write $A \subset B$ if every element of A is contained in B . If A is contained in and may be equal to B we write $A \subseteq B$.

Throughout these notes \emptyset will denote the empty set. That is, the set containing no elements.

Definition 1.2. Two sets A, B are said to be disjoint if $A \cap B = \emptyset$.

Unions complements and differences satisfy certain laws. The proof of the next result is a simple exercise.

Proposition 1.3. Let A, B, C be sets. The following relations hold.

(i) $(A \cup B) \cap C = (A \cap C) \cup (B \cap C).$

(ii) $(A \cap B) \cup C = (A \cup C) \cap (B \cup C).$

(iii) $(A \cup B) - C = (A - C) \cup (B - C).$

(iv) $(A \cap B) - C = (A - C) \cap (B - C).$

The most important rules for sets are deMorgan's laws.

Definition 1.4. Let $A \subset X$. Then $A^c = \{x \in X : x \notin A\}$. We call A^c the complement of A . We define $X^c = \emptyset$.

Theorem 1.5 (deMorgan). *Let $A_i, i \in \mathbb{N}$ be a collection of sets. Then*

$$\left(\bigcup_{i=1}^{\infty} A_i\right)^c = \bigcap_{i=1}^{\infty} A_i^c, \quad \left(\bigcap_{i=1}^{\infty} A_i\right)^c = \bigcup_{i=1}^{\infty} A_i^c.$$

Two other useful relationships are $A - B = A \cap B^c$ and $A \subseteq B$ if and only if $B^c \subseteq A^c$.

1.1. Countability. It is often important to distinguish between different kinds of infinite sets. For example, the rational numbers and the real numbers are both infinite sets, but there is a sense in which the real numbers is a bigger set than the rational numbers. To make this precise we introduce the notion of countability.

Definition 1.6. A set X is countable if there is a one to one function $f : X \rightarrow \mathbb{N}$. A countable set is also said to be denumerable. A set which is not countable is said to be uncountable. If f is also onto then we say that X is countably infinite.

Some authors prefer to use the term countable only when the set is countably infinite. A finite set might then be termed finitely countable. However this distinction is unimportant. There are equivalent formulations which are useful in establishing the countability of certain sets.

Theorem 1.7. *Let A be an infinite set. The following are equivalent.*

- (i) *A is countable.*
- (ii) *There exists a subset B of \mathbb{N} and a function $f : B \rightarrow A$ which is onto.*
- (iii) *There exists a function $g : A \rightarrow \mathbb{N}$ that is one to one.*

Proof. These are all straightforward. For example (iii) follows from the fact that there is a one to one and onto function $f : A \rightarrow \mathbb{N}$, so f is invertible. The others are exercises. \square

An important fact about countable sets follows.

Theorem 1.8. *Let $X_i, i = 1, 2, 3, \dots$ be countable sets. Then the union $X = \bigcup_{i=1}^{\infty} X_i$ is also countable.*

Proof. We let $X_i = \{x_1^i, x_2^i, x_3^i, \dots\}$. Let $B = \{2^k 3^n : k, n \in \mathbb{N}\}$. Now define $f : B \rightarrow A$ by $f(2^k 3^n) = x_k^n$. Then f maps B onto A , so A is countable by Theorem 1.7. \square

The proof of the next result is also an easy consequence of Theorem 1.7 and is an exercise.

Theorem 1.9. *Suppose that $X_i, i = 1, \dots, n, n < \infty$. are countable sets. Then $X_1 \times \dots \times X_n$ is countable.*

Example 1.1. The set $\{a, b, c, d\}$ is countable. For example we might have $f(a) = 1, f(b) = 2, f(c) = 3, f(d) = 4$.

Example 1.2. The natural numbers are countable. Just take $f(n) = n$.

Example 1.3 (Cantor). The rational numbers \mathbb{Q} are countable. Clearly $\mathbb{Q} = \mathbb{Q}^+ \cup \mathbb{Q}^-$, where the superscripts denote the negative and non-negative rationals respectively. So it is enough to show that the positive rationals are countable. Define $f : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{Q}^+$ by $f(m, n) = m/n$. Clearly f is onto, so by Theorem 1.7 the rationals are countable.

Example 1.4 (Cantor). The real numbers \mathbb{R} are uncountable. A very nice proof of this result uses the Baire Category Theorem and will be presented later. Cantor presented at least two proofs of this result. The second is the most famous and is known as his diagonal argument. Suppose that we restrict attention to the interval $[0, 1]$ and represent every number in $[0, 1]$ in binary, that is as a possibly infinite sequence of zeroes and ones. We then make a list of all the elements in some order. So for example if we create a list of sequences from the binary expansion of the numbers in $[0, 1]$, the list might look like this:

$$\begin{aligned} s_1 &= (0, 0, 0, 0, 0, 0, 0, \dots) \\ s_2 &= (0, 1, 1, 1, 0, 0, 1, \dots) \\ s_3 &= (1, 1, 1, 1, 0, 0, 0, \dots) \\ s_4 &= (0, 0, 1, 1, 0, 1, 0, \dots) \\ s_5 &= (0, 1, 1, 0, 0, 1, 1, \dots) \\ &\vdots \end{aligned}$$

We claim that no possible list can contain every possible sequence of zeroes and ones. To show this we construct an element s_0 which is not in the given list. We do so by looking down the diagonal of the array of numbers given above. That is, we look at the element s_{ii} and choose element number i of s_0 to not equal s_{ii} . So from the list here we would define

$$s_0 = (1, 0, 0, 0, 1, \dots)$$

Notice the first element of s_1 is 0, so we choose the first element of s_0 to be 1. The second element of s_2 is 1, so the second element of s_0 is 0. The third element of s_3 is 1, so the third element of s_0 is 0 and so on.

The sequence s_0 is not in the above list. Suppose otherwise. Then there is an integer N such that s_N is in the above list and $s_0 = s_N$. In particular the N th term of the sequence s_0 is the N th term of the sequence s_N . But this is a contradiction, because we constructed s_0 by choosing $s_{0N} \neq s_{NN}$. So s_0 is not in the above list. This is true for any

possible countable list. So no countable list of sequences of zeroes and ones can contain every sequence of zeroes and ones. Hence the interval $[0, 1]$ is not countable and hence \mathbb{R} is not countable.

1.2. Sets, Real numbers. Analysis makes use of an axiom and properties of the real numbers. We start with the humble triangle inequality, which is easily the most important inequality in mathematics.

Lemma 1.10. *For any real numbers a, b , $|a + b| \leq |a| + |b|$, where the absolute value is defined by $|x| = \sqrt{x^2}$ for $x \in \mathbb{R}$.*

Proof. This is elementary.

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

We used the fact that $ab \leq |ab|$. Now take square roots. \square

A rigorous treatment of analysis must begin with the axiom which is the foundation of the subject. If we consider sets of real numbers, then we can ask various questions of them. For example, are they bounded?

Definition 1.11. A finite number u is an upper bound for a set $A \subseteq \mathbb{R}$ if for every $x \in A$, $x \leq u$. Similarly, l is a lower bound for A if for every $x \in A$, $x \geq l$.

Now suppose that $A \subset \mathbb{R}$ is non-empty and that there is an upper bound? We can ask whether or not there is a least upper bound?

Definition 1.12. If \bar{u} is an upper bound of a set $A \subseteq \mathbb{R}$ with the property that $\bar{u} \leq u$, for all other upper bounds u , then \bar{u} is called the least upper bound or supremum of A . We write $\bar{u} = \sup A$. Similarly, a lower bound \bar{l} of a set $A \subset \mathbb{R}$ is the greatest lower bound or infimum, if for every lower bound of A we have $\bar{l} \geq l$. We write $\bar{l} = \inf A$.

1.3. The Least Upper Bound Axiom. Consideration of elementary examples would suggest that every non empty set, bounded above, does indeed have a least upper bound. Indeed it is impossible to write down a counter example. Many examples are straightforward. Take the set $[0, 1)$. The least upper bound is obviously 1. This is easy, but it turns out that there is no way that one can prove that every nonempty set of real numbers which is bounded above has a least upper bound. Instead we make it an axiom.

Axiom 1: The Least Upper Bound Axiom: Every non empty set of real numbers which is bounded above has a least upper bound.

Example 1.5. This example is important and uses ideas that we will develop below. Let $f : [0, 1] \rightarrow \mathbb{R}$ be a continuous function. Let

$\mathcal{P} = \{x_0, x_1, \dots, x_n\}$ be a partition of $[0, 1]$. Let $m_i = \max_{x \in [x_{i-1}, x_i]} f(x)$. Define $L(f, \mathcal{P}) = \sum_{i=1}^n m_i(x_i - x_{i-1})$. Then by our axiom,

$$I = \sup_{\mathcal{P}} \{L(f, \mathcal{P})\} \quad (1.1)$$

exists. The point is that we could not otherwise prove the existence of this supremum.

Example 1.6. Does the set of rational numbers in $A = [0, \pi]$ have a least upper bound in A ? Suppose that there is a rational number $0 < \epsilon < \pi$, then by the continuum property of the real numbers, there is a rational number δ , with $\epsilon < \delta < \pi$. So this set has no least upper bound contained in A . The least upper bound is obviously π . Note the supremum does not have to be in the set itself.

Example 1.7. Consider the set $A = \{x > 0, x \in \mathbb{Q}, x^2 \leq 2\}$. Since $\sqrt{2}$ is irrational, the supremum of A (i.e. $\sqrt{2}$) is not in A .

From this axiom all of analysis is derived. We start by proving that non-empty sets bounded below have greatest lower bounds.

Theorem 1.13. *A non-empty set of real numbers bounded below has a greatest lower bound.*

Proof. Suppose that A is nonempty and bounded below. Now consider the set $-A = \{-x : x \in A\}$. This set is nonempty and bounded above: If l is a lower bound of A , then $-l$ is an upper bound of $-A$. To see this, notice that if $x \in A$, then $l \leq x$. So $-l \geq -x$. Hence $-l$ is an upper bound. By **Axiom 1**, $-A$ has a least upper bound \bar{u} . Then it follows that $\bar{l} = -\bar{u}$ is the greatest lower bound for A . \square

An important result we use extensively follows.

Theorem 1.14. *Assume that $\sup A$ exists, where $A \subset \mathbb{R}$. Then for every $\epsilon > 0$ we can find an x such that*

$$\sup A - \epsilon < x \leq \sup A.$$

Proof. Suppose that for every $x \in A$, $x \leq \sup A - \epsilon$. Then $\sup A - \epsilon$ is an upper bound for A , less than $\sup A$, which is a contradiction. So there must be some x in A with $x > \sup A - \epsilon$. Clearly $x \leq \sup A$. \square

The Archimedean Property This is an obvious property, which nevertheless is fundamental: Given any two positive real numbers x, y , we can find a natural number n such that $nx > y$. Equivalently, there is no largest natural number.

1.4. Limits and Sequences. We now introduce the concept of a sequence.

Definition 1.15. A sequence in \mathbb{R} is a function $f : \mathbb{N} \rightarrow \mathbb{R}$. We usually write $f(n) = x_n$ and denote the sequence by $\{x_n\}_{n=1}^{\infty}$ or just $\{x_n\}$.

Sequences are what analysis is made of. Many practical problems have solutions which are given by constructing sequences which “converge” to a solution. That is, which gets closer and closer to the solution as n increases. We can formally define convergence as follows.

Definition 1.16. A sequence of real numbers $\{x_n\}_{n=1}^{\infty}$ is said to be convergent with limit x , if for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that $n \geq N$ implies $|x_n - x| < \epsilon$. We write $x_n \rightarrow x$. A sequence which does not converge is said to diverge.

Example 1.8. Let $a_n = \frac{1}{n}$. Let $\epsilon > 0$. We show that the limit is zero. We require $|\frac{1}{n} - 0| < \epsilon$. Clearly if we let $N \in \mathbb{N}$ with $N > 1/\epsilon$, then for all $n \geq N$ we have $|\frac{1}{n} - 0| < \frac{1}{N} < \epsilon$. So the sequence converges to 0.

Example 1.9. Let $a_n = \frac{n^2-4}{n^2+4}$. Then the sequence converges to 1. To show this, let $\epsilon > 0$. Then $|a_n - 1| < \epsilon$ implies

$$\begin{aligned} \left| \frac{n^2-4}{n^2+4} - 1 \right| &= \left| \frac{n^2-4-(n^2+4)}{n^2+4} \right| \\ &= \left| \frac{-8}{n^2+4} \right| < \epsilon. \end{aligned}$$

Now $n^2 + 4 > n$ so that $\frac{8}{n^2+4} < \frac{8}{n}$. Thus if $N > \frac{1}{8\epsilon}$ and $n \geq N$, then $|a_n - 1| < \epsilon$. We do not need to find the *best possible* value of N . Just one that works.

It is easy to establish some basic facts about convergent sequences. From here on we will see just how essential the triangle inequality is to analysis. The subject could not exist without it.

Theorem 1.17. *The limit of a convergent sequence is unique.*

Proof. Suppose that $x_n \rightarrow x$ and $x_n \rightarrow y$. Then for all n

$$|x - y| = |x - x_n + x_n - y| \leq |x_n - x| + |x_n - y|.$$

But $|x_n - x| \rightarrow 0$ and $|x_n - y| \rightarrow 0$, so $|x - y| = 0$. □

Theorem 1.18. *Every convergent sequence is bounded.*

Proof. Let $\{x_n\}_{n=1}^{\infty}$ be a convergent sequence with limit x and choose N such that $n \geq N$ implies $|x_n - x| < 1$. Now let

$$M = \max\{|x_1|, \dots, |x_{N-1}|, 1 + |x|\}.$$

Clearly if $1 \leq n \leq N - 1$ then $|x_n| \leq M$. Conversely, if $n \geq N$, Then

$$|x_n| = |x_n - x + x| \leq |x_n - x| + |x| < 1 + |x| \leq M.$$

So for all n , $|x_n| \leq M$, and hence the sequence is bounded. □

Convergent sequences behave as you would expect under addition, multiplication and division.

Theorem 1.19. *Let a, b be constants and suppose that $\{x_n\}$ and $\{y_n\}$ are convergent sequences, with limits x, y respectively. Then*

$$(1) \quad ax_n \rightarrow ax.$$

$$(2) \quad ax_n + by_n \rightarrow ax + by$$

$$(3) \quad x_n y_n \rightarrow xy$$

$$(4) \quad \text{If } y_n \text{ is never zero and } y \neq 0, x_n/y_n \rightarrow x/y.$$

Proof. These are easy to prove. For example, let $\epsilon > 0$. Choose M such that $n \geq N$ implies $|x_n - x| < \epsilon/(2|a|)$ and K such that $n \geq K$ implies $|y_n - y| < \epsilon/(2|b|)$. Then let $N = \max\{M, K\}$. Then for $n \geq N$,

$$\begin{aligned} |ax_n + by_n - ax - by| &\leq |a||x_n - x| + |b||y_n - y| \\ &< |a|\epsilon/(2|a|) + |b|\epsilon/(2|b|) = \epsilon. \end{aligned}$$

Proofs of the other results are exercises. \square

We also have the essential result that increasing bounded sequences are convergent.

Theorem 1.20. *Every monotone increasing sequence which is bounded above has a limit. Every monotone decreasing sequence bounded below has a limit.*

Proof. Let $\{x_n\}_{n=1}^{\infty}$ be a bounded, increasing sequence. Then for all n , $x_{n+1} \geq x_n$. Consider the set $A = \{x_1, x_2, x_3, \dots\}$. This set is non-empty and bounded above, so it has a least upper bound, which we denote by x . Now pick $\epsilon > 0$ and choose $N \in \mathbb{N}$ such that $x_N > x - \epsilon$. Since x is the least upper bound we can do this, otherwise it would not be the least upper bound. Since $\{x_n\}_{n=1}^{\infty}$ is increasing we have for $n \geq N$, $|x_n - x| = x - x_n < \epsilon$. Hence $x_n \rightarrow x$. The case of a decreasing sequence is similar. \square

Subsequences play an essential role in analysis.

Definition 1.21. Let $\{x_n\}_{n=1}^{\infty}$ be a sequence. A subsequence $\{x_{n_K}\}_{K=1}^{\infty}$ is a sequence contained in $\{x_n\}_{n=1}^{\infty}$, where n_K is an increasing sequence of integers. So $n_K \rightarrow \infty$ as $K \rightarrow \infty$.

The most important results about sequences on the real line stem from the Bolzano-Weierstrass Theorem. To prove this, we need some preliminaries.

Theorem 1.22. *Every sequence has a monotone subsequence.*

Proof. We sketch the proof. We suppose that the sequence is not constant after some term x_N . If $x_n = a$ for all $n \geq N$, then the result is trivial. So suppose this is not the case. The basic idea is to construct the sequence. We pick the first element, say $y_1 = x_{n_1}$, then we move

along the sequence till we find another element x_{n_1} which is larger than x_n and then take $y_2 = x_{n_1}$. Now move along the sequence till we come to a larger element, make that the third element of the subsequence. Continuing we construct a monotone increasing sequence. Similarly for the case of a monotone decreasing sequence. \square

Theorem 1.23 (Bolzano-Weierstrass). *Every bounded sequence of real numbers has a convergent subsequence.*

Proof. A bounded sequence has a bounded monotone subsequence. Bounded monotone sequences are convergent. So every bounded sequence has a convergent subsequence. \square

The reason why this result is so important is that we often need to deal with sequences of real numbers on bounded intervals and in many proofs we pick a convergent subsequence to work with. Texts on elementary real analysis will deal with this and we will see some examples of this in practice.

Definition 1.24. A sequence of real numbers $\{x_n\}_{n=1}^{\infty}$ is said to be a Cauchy sequence if for every $\epsilon > 0$ we can find an $N \in \mathbb{N}$ such that $n, m \geq N$ implies $|x_n - x_m| < \epsilon$.

Every Cauchy sequence is convergent. This fact underpins a lot of what follows. Cauchy sequences and convergent sequences are basically the same. The point is that Cauchy's criterion gives us a different way of determining convergence, which is particularly useful when we do not have the limit available to us. First we show an easy result.

Proposition 1.25. *Every convergent sequence is a Cauchy sequence.*

Proof. We pick $N \in \mathbb{N}$ such that $n \geq N$ implies $|x_n - x| < \epsilon/2$. Then we have for $n, m \geq N$

$$|x_n - x_m| = |x_n - x + x - x_m| \leq |x_n - x| + |x_m - x| < \epsilon/2 + \epsilon/2 = \epsilon,$$

so a convergent sequence is Cauchy. \square

Proving the converse is more difficult. We have to show that a Cauchy sequence is bounded. The proof of this result is much the same as the proof that a convergent sequence is bounded. Then we establish the following result:

Proposition 1.26. *If a Cauchy sequence has a convergent subsequence with limit x , then the Cauchy sequence converges to x .*

Proof. To see this, suppose that $\{x_n\}_{n=1}^{\infty}$ is Cauchy and that there is a subsequence $\{x_{n_K}\}_{K=1}^{\infty}$ which converges to x . So that $\lim_{K \rightarrow \infty} x_{n_K} = x$. We then choose N large enough to make $|x_n - x_m| < \epsilon/2$ for all $n, m \geq N$ and pick K large enough to make $n_K > N$. Then by the triangle

inequality

$$\begin{aligned} |x_n - x| &= |x_n - x_{n_K} + x_{n_K} - x| \\ &\leq |x_n - x_{n_K}| + |x_{n_K} - x| \\ &< \epsilon/2 + \epsilon/2 = \epsilon. \end{aligned}$$

So $x_n \rightarrow x$. □

The next step is easy.

Theorem 1.27. *Every Cauchy sequence is convergent.*

Proof. Every Cauchy sequence is bounded. By the Bolzano-Weierstrass Theorem, it follows that every Cauchy sequence has a convergent subsequence. Consequently, every Cauchy sequence converges. □

Cauchy sequences are important because they allow us to establish convergence without knowing what the limit is. In most cases, we cannot compute the limit exactly, so we cannot prove convergence by establishing that $|x_n - x| \rightarrow 0$ since x is unknown. We can however often prove that $|x_n - x_m| \rightarrow 0$ as $n, m \rightarrow \infty$.

It is important to understand that a sequence with the property that $|x_{n+k} - x_n| \rightarrow 0$, as $n \rightarrow \infty$, for fixed k , is not necessarily Cauchy. We insist that $|x_n - x_m| \rightarrow 0$ as both $n, m \rightarrow \infty$. For example, the harmonic sequence

$$x_n = \sum_{k=1}^n \frac{1}{k},$$

diverges. The proof of this result is quite ancient and is often attributed to Nicolas Oresme (born between 1320-25, died 1382). However it may well have been established in India even earlier. It is based on the observation that

$$\begin{aligned} \sum_{k=N}^{2N} \frac{1}{k} &= \frac{1}{N} + \cdots + \frac{1}{2N} \\ &\geq N \times \frac{1}{2N} = \frac{1}{2} \end{aligned}$$

Similarly $\sum_{k=2N+1}^{4N} \frac{1}{k} \geq \frac{1}{2}$, $\sum_{k=4N+1}^{8N} \frac{1}{k} \geq \frac{1}{2}$ etc. So that

$$\sum_{k=1}^{\infty} \frac{1}{k} \geq 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \cdots$$

and so $x_n \rightarrow \infty$. However $|x_{n+1} - x_n| = \frac{1}{n} \rightarrow 0$ as $n \rightarrow \infty$.

The use of convergent subsequences is one of the most common techniques in analysis. Notice that we have also used the triangle inequality extensively. These are the fundamental tools of Analysis and they are used over and over again. When we extend Analysis to different settings, the first thing that we require is that we have a measure of the length

or size of some mathematical object for which the triangle inequality holds. It is the most important inequality in mathematics.

1.5. Infinite Series and Convergence Tests. As the example of the harmonic sequence shows, we can handle series by treating them as sequences.

Definition 1.28. A series $S = \sum_{n=1}^{\infty} a_n$ is said to be convergent with limit S if the sequence of partial sums $\{S_N\}_{N=1}^{\infty}$ with $S_N = \sum_{n=1}^N a_n$ is convergent with limit S . If the series is not convergent, we say it diverges.

The next result is obvious.

Theorem 1.29. *If $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ are convergent with sums S and T respectively, then $\sum_{n=1}^{\infty} (a_n + b_n) = S + T$. Further $\sum_{n=1}^{\infty} ca_n = cS$ for all $c \in \mathbb{R}$.*

Proof. This follows from previously established properties of sequences applied to $S_N = \sum_{n=1}^N a_n$ and $T_N = \sum_{n=1}^N b_n$. \square

Note it is not true that if $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ both converge then $\sum_{n=1}^{\infty} a_n b_n$ is convergent. We require absolute convergence for this.

Definition 1.30. A series is said to be absolutely convergent if the series $\sum_{n=1}^{\infty} |a_n|$ converges. If $\sum_{n=1}^{\infty} a_n$ converges, but $\sum_{n=1}^{\infty} |a_n|$ diverges, the series is said to be conditionally convergent.

Example 1.10. The harmonic series $\sum_{n=1}^{\infty} \frac{1}{n}$ is divergent. However the alternating series $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} = \ln 2$. Hence this series is conditionally convergent.

There are many simple but useful properties possessed by convergent series.

Lemma 1.31. *If $\sum_{n=1}^{\infty} a_n$ is convergent, then $\lim_{n \rightarrow \infty} a_n = 0$.*

Proof. We have $S_n = \sum_{k=1}^n a_k$ and $\{S_n\}_{n=1}^{\infty}$ is convergent with limit S . Then $S_{n+1} - S_n \rightarrow 0$ but $S_{n+1} - S_n = a_n$. \square

Another useful fact is the following. The proof is an exercise.

Proposition 1.32. *Let $\sum_{n=1}^{\infty} a_n$ be convergent. Then as, $N, M \rightarrow \infty$, $\sum_{n=M+1}^N a_n \rightarrow 0$.*

Let us see what is needed to guarantee that $\sum_{n=1}^{\infty} a_n b_n$ converges.

Proposition 1.33. *If $\sum_{n=1}^{\infty} a_n$ is convergent and $\sum_{n=1}^{\infty} b_n$ is absolutely convergent, then $\sum_{n=1}^{\infty} a_n b_n$ is convergent.*

Proof. Consider the sequence $S_N = \sum_{n=1}^N a_n b_n$. Now $\sum_{n=1}^{\infty} a_n$ is convergent, hence the sequence $a_n \rightarrow 0$ and so is bounded. Suppose that $|a_n| \leq K$. Then if $N > M$

$$\begin{aligned} |S_N - S_M| &= \left| \sum_{n=M+1}^N a_n b_n \right| \leq \sum_{n=M+1}^N |a_n b_n| \\ &\leq K \sum_{n=M+1}^N |b_n| \rightarrow 0, \end{aligned}$$

as $N, M \rightarrow \infty$ since $\sum_{n=1}^{\infty} |b_n|$ is convergent. Thus $\{S_N\}_{N=1}^{\infty}$ is Cauchy and hence it converges. \square

There are various tests for convergence. Most rely on the comparison test.

Theorem 1.34 (Comparison Test). *Suppose that $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ are series of positive terms. If there is an $N \in \mathbb{N}$ such that $n \geq N$ implies $a_n \leq b_n$ and $\sum_{n=1}^{\infty} b_n$ is convergent, then $\sum_{n=1}^{\infty} a_n$ is also convergent. Conversely if $\sum_{n=1}^{\infty} a_n$ is divergent, then $\sum_{n=1}^{\infty} b_n$ is also divergent.*

Proof. Let $T = \sum_{k=1}^{\infty} b_k$ and $S_n = \sum_{k=1}^n a_k$. Since the a_n are positive S_N is increasing. We show that it is bounded above.

$$\begin{aligned} S_n &= a_1 + a_2 + \cdots + a_{N-1} + a_N + \cdots + a_n \\ &\leq a_1 + \cdots + a_{N-1} + b_N + \cdots + b_n \\ &= b_1 + \cdots + b_{N-1} + b_N + \cdots + b_n \\ &\quad + (a_1 - b_1) + \cdots + (a_{N-1} - b_{N-1}) \\ &= T_n + \sum_{k=1}^{N-1} (a_k - b_k) \\ &\leq T + \sum_{k=1}^{N-1} (a_k - b_k). \end{aligned}$$

So $\{S_n\}_{n=1}^{\infty}$ is increasing and bounded above and hence converges.

The proof of the second part is similar. \square

A variant of this is the limit comparison test.

Theorem 1.35 (The Limit Comparison Test). *Let $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ be series of strictly positive terms. Suppose that*

$$\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = l \neq 0.$$

Then either both series converge or both series diverge.

Proof. We can assume that $a_n \rightarrow 0$ and $b_n \rightarrow 0$, since the series will diverge otherwise. Since $a_n/b_n \rightarrow l$, the sequence $\{a_n/b_n\}$ is bounded by some number K . From which it follows that $a_n \leq Kb_n$. So that if $\sum_{n=1}^{\infty} b_n$ converges, then $\sum_{n=1}^{\infty} a_n$ converges by the comparison test. Conversely, if $\sum_{n=1}^{\infty} a_n$ diverges, then so does $\sum_{n=1}^{\infty} b_n$.

Now we can find $N \in \mathbb{N}$ such that $n \geq N$ implies $a_n > \frac{1}{2}lb_n$, so that if $\sum_{n=1}^{\infty} a_n$ converges, then so does $\sum_{n=1}^{\infty} b_n$. The divergence of $\sum_{n=1}^{\infty} b_n$ implies the divergence of $\sum_{n=1}^{\infty} a_n$. \square

Example 1.11. Consider $S_n = \sum_{n=1}^{\infty} \frac{1}{n^2}$. Then

$$\begin{aligned} S_n &= 1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \cdots \\ &< 1 + \frac{1}{2} + \frac{1}{2 \cdot 3} + \frac{1}{3 \cdot 4} + \frac{1}{n(n-1)} \\ &= 1 + \left(1 - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \cdots + \left(\frac{1}{n-1} - \frac{1}{n}\right) \\ &= 2 - \frac{1}{n} < 2. \end{aligned}$$

So this series is convergent. (It actually equals $\pi^2/6$). Now consider the series $\sum_{n=1}^{\infty} \frac{n+1}{2n^3+n+3}$. We apply the limit comparison test and compute

$$\lim_{n \rightarrow \infty} \left(\frac{n+1}{2n^3+n+3} \right) / \left(\frac{1}{n^2} \right) = 1/2 \neq 0.$$

So the second series also converges.

A powerful convergence test follows.

Theorem 1.36 (The Ratio Test). *Let $\sum_{n=1}^{\infty} a_n$ be a series of strictly positive terms. Let $\lim_{n \rightarrow \infty} \frac{a_{n+1}}{a_n} = L$. Then the series converges if $L < 1$ and diverges if $L > 1$. If $L = 1$ then the test is inconclusive.*

Proof. Suppose that $L < 1$. Pick $r \in \mathbb{R}$ such that $L < r < 1$. We can choose $n \in \mathbb{N}$ such that $n \geq N$ implies

$$\left| \frac{a_{n+1}}{a_n} - L \right| < r - L$$

or

$$-(r - L) < \frac{a_{n+1}}{a_n} - L < r - L.$$

So $a_{n+1} < ra_n$. Also $a_n < ra_{n-1} < r^2a_{n-2} < r^3a_{n-3}$ etc. Indeed $a_n < r^k a_{n-k}$. Now let $k = n - N$. Then

$$a_n \leq r^{n-N} a_N = \left(\frac{a_N}{r^N} \right) r^n.$$

The series $\sum_{n=1}^{\infty} \left(\frac{a_N}{r^N}\right) r^n$ is a geometric series with common ratio $r < 1$ and so converges. By the comparison test $\sum_{n=1}^{\infty} a_n$ also converges.

For the case $L > 1$ the proof is similar, with the final inequalities reversed and $r > 1$, giving a divergent geometric series. Finally, for the series $\sum_{n=1}^{\infty} \frac{1}{n}$, $L = 1$ and the series diverges. For the series $\sum_{n=1}^{\infty} \frac{1}{n^2}$, $L = 1$ and the series converges. So the ratio test is inconclusive if $L = 1$. \square

Remark 1.37. We can apply the ratio test to series of nonpositive terms. We instead consider $\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| = L$ and the conclusions are the same as in the given result.

To state the n th root test, we introduce the idea of a limsup.

Definition 1.38. If $\{x_n\}$ is a bounded sequence, then the largest subsequential limit \bar{l} is

$$\bar{l} = \limsup x_n \quad (1.2)$$

and the smallest subsequential limit s is

$$s = \liminf x_n. \quad (1.3)$$

Some authors also write $\lim_{n \rightarrow \infty} \sup x_n$ and $\lim_{n \rightarrow \infty} \inf x_n$.

Example 1.12. If $x_n = (-1)^n$, then $\limsup x_n = 1$ and $\liminf x_n = -1$.

The convergence of a sequence can be given in terms of its limsup and liminf. The proof of the next result is an exercise.

Proposition 1.39. A sequence $\{x_n\}_{n=1}^{\infty}$ converges if and only if

$$\limsup x_n = \liminf x_n.$$

We now give yet another convergence test.

Theorem 1.40 (The n th root test). Let $\sum_{n=1}^{\infty} a_n$ be a series and suppose that

$$\limsup |a_n|^{1/n} = L. \quad (1.4)$$

If $L < 1$ the series converges. If $L > 1$ the series diverges. If $L = 1$ the series may converge or diverge.

Proof. This is another application of the comparison test. If $L < 1$, then there is an r such that $L < r < 1$ and for n large enough $|a_n| < r^n$. Convergence follows from the comparison test with a geometric series. The proof of the second case is similar. Finally, we can exhibit series which converge when $L = 1$ and diverge when $L = 1$. (This is an exercise). \square

There are a number of other, lesser known tests for convergence which can be very useful. We present one next.

Theorem 1.41 (Cauchy Condensation Test). *Suppose that the sequence a_n is positive and non-increasing. Then the series $\sum_{n=1}^{\infty} a_n$ converges, if and only if the series $\sum_{n=0}^{\infty} 2^n a_{2^n}$ converges. Moreover we have the estimate*

$$\sum_{n=1}^{\infty} a_n \leq \sum_{n=0}^{\infty} 2^n a_{2^n} \leq 2 \sum_{n=1}^{\infty} a_n.$$

Proof. Since the sequence is non-decreasing, we have $a_2 + a_3 \leq 2a_2$, $a_4 + a_5 + a_6 + a_7 \leq 4a_4$ etc. So that

$$\sum_{n=1}^{\infty} a_n \leq a_1 + 2a_2 + 4a_4 + 8a_8 + \cdots = \sum_{n=0}^{\infty} 2^n a_{2^n}.$$

Thus if $\sum_{n=0}^{\infty} 2^n a_{2^n}$ converges, then so does $\sum_{n=1}^{\infty} a_n$. Similarly $a_1 + a_2 \leq 2a_2$, $a_2 + a_4 + a_4 + a_4 \leq 2a_2 + 2a_3$, etc. So that

$$\begin{aligned} \sum_{n=0}^{\infty} 2^n a_{2^n} &= a_1 + 2a_2 + 4a_4 + \cdots \leq 2a_1 + 2a_2 + 2a_3 + \cdots \\ &= 2 \sum_{n=1}^{\infty} a_n. \end{aligned}$$

So by the comparison test the series $\sum_{n=0}^{\infty} 2^n a_{2^n}$ converges if $\sum_{n=1}^{\infty} a_n$ converges. The estimate follows from the above. \square

Example 1.13. We consider the convergence of $\sum_{n=1}^{\infty} \frac{1}{n^p}$. Now

$$\sum_{n=0}^{\infty} 2^n a_{2^n} = \sum_{n=0}^{\infty} 2^n \frac{1}{(2^n)^p} = \sum_{n=0}^{\infty} \frac{1}{2^{np-n}}.$$

This is a geometric series that will converge for $np - n > 1$ and diverge otherwise. Hence the original series converges for $p > 1$ and diverges for $p \leq 1$.

Finally we mention a test for alternating series.

Theorem 1.42. *Let $\{a_n\}_{n=1}^{\infty}$ be a sequence of positive terms with $a_n \rightarrow 0$. Then the series $\sum_{n=1}^{\infty} (-1)^{n+1} a_n$ is convergent.*

Proof. We will show that the sequence of partial sums is Cauchy. Let $\epsilon > 0$ and note that since $a_n \rightarrow 0$ we can find an $N \in \mathbb{N}$ such that $a_n < \epsilon$ for all $n \geq N$. Next observe that since a_n is monotone decreasing, $a_n - a_{n+1} \geq 0$. Consequently

$$a_{m+1} - a_{m+2} + a_{m+3} - a_{m+4} + \cdots + (-1)^{n+1} a_n \leq a_{m+1}.$$

Now if $S_n = \sum_{k=1}^{\infty} (-1)^{k+1} a_k$, then pick $n > m \geq N$.

$$\begin{aligned} |S_n - S_m| &= |(a_1 - a_2 + a_3 - \cdots + (-1)^{n+1} a_n) - (a_1 - a_2 + a_3 \\ &\quad - \cdots + (-1)^{m+1} a_m)| \\ &= |a_{m+1} - a_{m+2} + a_{m+3} - a_{m+4} + \cdots + (-1)^{n+1} a_n| \\ &\leq |a_{m+1}| < \epsilon. \end{aligned}$$

So $\{S_n\}_{n=1}^{\infty}$ is Cauchy and so the series converges. \square

2. CONTINUOUS FUNCTIONS AND THEIR PROPERTIES.

Continuous functions are familiar from elementary calculus. To define them properly we need to know what continuity actually means.

2.1. Limits. It turns out that there are several different ways of defining continuity. These turn out to be equivalent. Which means that if a function is continuous by one definition, it is continuous by the others.

To proceed, we must extend the definition of a limit to functions. We will first define continuity in terms of sequences.

Definition 2.1. We define limit points and limits of functions as follows.

- (1) A point x is a limit point of a set $X \subseteq \mathbb{R}$ if there is a sequence $\{x_n\}_{n=1}^{\infty} \subset X$ such that $x_n \rightarrow x$. If there is no such sequence, then x is an isolated point.
- (2) Let $X \subseteq \mathbb{R}$, $f : X \rightarrow \mathbb{R}$ and x_0 a limit point of X . Then L is the limit of f as $x \rightarrow x_0$ if and only if, given $\epsilon > 0$, there exists $\delta > 0$ such that $x \in X$, $|x - x_0| < \delta$ implies $|f(x) - L| < \epsilon$.

Limits of functions satisfy the usual arithmetic properties.

Theorem 2.2. Let $f, g : X \rightarrow \mathbb{R}$ be functions and c a constant. If x_0 is a limit point of X and $\lim_{x \rightarrow x_0} f(x) = L$ and $\lim_{x \rightarrow x_0} g(x) = M$, then

$$\lim_{x \rightarrow x_0} cf(x) = cL \tag{2.1}$$

$$\lim_{x \rightarrow x_0} (f(x) + g(x)) = L + M \tag{2.2}$$

$$\lim_{x \rightarrow x_0} f(x)g(x) = LM \tag{2.3}$$

$$\lim_{x \rightarrow x_0} f(x)/g(x) = L/M, \tag{2.4}$$

provided $M \neq 0$ and g is nonzero.

Proofs of these results are exercises with the triangle inequality and are left to the reader. We can define right and left limits for functions.

Definition 2.3. Let $f : X \rightarrow \mathbb{R}$, where $X \subseteq \mathbb{R}$. We say that

$$\lim_{x \rightarrow a^+} f(x) = L,$$

if for every $\epsilon > 0$, there exists $\delta > 0$ such that $a < x < a + \delta$ implies $|f(x) - L| < \epsilon$. Similarly we say

$$\lim_{x \rightarrow a^-} f(x) = L,$$

if for every $\epsilon > 0$, there exists $\delta > 0$ such that $a - \delta < x < a$ implies $|f(x) - L| < \epsilon$.

An easy result follows.

Proposition 2.4. Let $f : X \rightarrow \mathbb{R}$, where $X \subseteq \mathbb{R}$. Then $\lim_{x \rightarrow a} f(x) = L$ if and only if $\lim_{x \rightarrow a^+} f(x) = \lim_{x \rightarrow a^-} f(x) = L$.

The proof is an exercise. Finally we define the limit at infinity.

Definition 2.5. Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be a function. Then $\lim_{x \rightarrow \infty} f(x) = L$ if for every $\epsilon > 0$ there exists an $M > 0$ such that $x \geq M$ implies $|f(x) - L| < \epsilon$. Similarly we say that $\lim_{x \rightarrow -\infty} f(x) = L$ if for every $\epsilon > 0$ there exists $M < 0$ such that $x \leq M$ implies $|f(x) - L| < \epsilon$.

2.1.1. *The composition of functions.* Given two continuous functions, f and g , there are many ways which we can combine them. One of the most important is the composition process. The composition of f and g is the new function

$$\begin{aligned} h(x) &= (f \circ g)(x) \\ &= f(g(x)). \end{aligned}$$

Example 2.1. Let $f(x) = x^2$, $g(x) = x^3$. Then $(f \circ g)(x) = f(g(x)) = (x^3)^2 = x^6$.

Example 2.2. Let $f(x) = \sin x$ and let $g(x) = \sqrt{x^2 + 1}$. Then

$$(g \circ f)(x) = g(f(x)) = \sqrt{\sin^2 x + 1}.$$

The composition of functions is an essential process. It turns out that if both f and g are continuous, then the composition of the two is also continuous. We will prove this later. Now we have to actually define continuity.

2.2. The Definition of Continuity. Having established the essentials about limits of functions, we introduce the crucial idea of continuity.

Definition 2.6. A function $f : X \rightarrow \mathbb{R}$ is said to be continuous at x if for any sequence $\{x_n\}_{n=1}^{\infty} \subset X$ which converges to x , we have

$$\lim_{n \rightarrow \infty} f(x_n) = f(x).$$

This can be recast in the following form.

Definition 2.7. A function $f : X \rightarrow \mathbb{R}$ is continuous at $x \in X$ if for any $\epsilon > 0$, we can find a $\delta_x > 0$ such that $|x - y| < \delta_x$ implies $|f(x) - f(y)| < \epsilon$.

We write δ_x to emphasise the dependence on the point x . So for each x we will require a different δ . If a function is continuous at every point in its domain, we say that it is continuous. The two definitions are clearly equivalent.

Theorem 2.8. *The two definitions of continuity stated above are equivalent.*

Proof. First suppose that f satisfies Definition 2.7. Let $\{x_n\}_{n=1}^\infty$ be a sequence in X with limit x . Pick $\epsilon > 0$ and $\delta > 0$ such that $|x - x_0| < \delta$ implies $|f(x) - f(x_0)| < \epsilon$. Since $x_n \rightarrow x$ we may find an $N \in \mathbb{N}$ such that $n \geq N$ implies $|x_n - x| < \delta$. Then $|f(x_n) - f(x)| < \epsilon$, but this means that $f(x_n) \rightarrow f(x)$, so f is continuous according to Definition 2.6.

Suppose that f does not satisfy Definition 2.7. Then we can find $\epsilon > 0$ such that for every $\delta > 0$ with $|x - x_0| < \delta$ we have $|f(x) - f(x_0)| \geq \epsilon$. Now choose a sequence $\{x_n\}_{n=1}^\infty$ in X with limit $x \in X$. Then given $\delta > 0$ we may find an $N \in \mathbb{N}$ such that $|x_n - x| < \delta$, but $|f(x_n) - f(x)| \geq \epsilon$. So $\{f(x_n)\}_{n=1}^\infty$ does not converge to $f(x)$ and thus f is not continuous by Definition 2.6. \square

There are other notions of continuity. The most important requires us to know what an open set is.

Definition 2.9. A set A of real numbers is *open* if for any $a \in A$ we can find an $\epsilon > 0$ such that the set $B_\epsilon(a) = \{x \in \mathbb{R}, |x - a| < \epsilon\}$ is contained entirely in A .

Another way of seeing this is that A does not contain its boundary. For example $(0, 1)$ is an open set because it does not contain its end points. The complement A^c of an open set is *closed*. A set which contains its boundary points is closed. So $[0, 1]$ is closed.

Using the notion of an open set we can define continuity as follows.

Definition 2.10. A function $f : X \subseteq \mathbb{R} \rightarrow Y \subseteq \mathbb{R}$ is continuous if for every open set $B \in Y$, the inverse image $f^{-1}(B) = \{x \in X : f(x) \in B\}$ is itself an open set.

This is equivalent to our previous two definitions, but we will not prove it. It is important because it allows us to define continuity in any settings where we have sets which can be seen as open in some sense. This is the basis of a branch of mathematics called *topology*.

Another reason why having equivalent definitions for the same concept is because it may be easier to use different definitions in different settings. Here is an example of that. Proving that the composition of

continuous functions is also continuous is a little tricky using our first two definitions. However using the definition of continuity in terms of open sets makes it easy.

Theorem 2.11. *Suppose that $g : X \subseteq \mathbb{R} \rightarrow Y \subseteq \mathbb{R}$ and $f : Y \rightarrow Z \subseteq \mathbb{R}$ are both continuous. Then $f \circ g : X \rightarrow Z$ is also continuous.*

Proof. This is a matter of carefully working out what the inverses are and where they are mapped to. Think about matrix inverses as a guide. Suppose that A and B are invertible matrices of the same size. Then $(AB)^{-1} = B^{-1}A^{-1}$, since $AB B^{-1}A^{-1} = A I A^{-1} = I$. This suggests that when we consider $f \circ g^{-1}(A)$ we have to flip the order of f and g .

In fact it is easy to see that $(f \circ g)^{-1}(A) = g^{-1}(f^{-1}(A))$. This is what we need. The proof is now easy. We let A be an open set. Since f is continuous, $f^{-1}(A) = B$ is an open set. However g is also continuous, so $g^{-1}(B)$ is again an open set. Thus $(f \circ g)^{-1}(A) = g^{-1}(B)$ is open and so $f \circ g$ is continuous. \square

It is also possible to consider partial forms of continuity. The most common forms are left and right continuity.

Definition 2.12. We say that f is right continuous at x_0 if $\lim_{x \rightarrow x_0^+} f(x)$ exists. If $\lim_{x \rightarrow x_0^-} f(x)$ exists, then we say that f is left continuous.

Example 2.3. The function $f(x) = \sqrt{x}$ is continuous on $(0, \infty)$, left continuous on $[0, \infty)$, but not continuous on \mathbb{R} .

2.3. Uniform Continuity. The most important form of continuity for the Riemann integral is uniform continuity.

Definition 2.13. A function $f : X \rightarrow \mathbb{R}$ is said to be uniformly continuous if given $\epsilon > 0$ we can find a $\delta > 0$ such that whenever $|x - y| < \delta$ we have $|f(x) - f(y)| < \epsilon$.

The point here is that unlike ordinary continuity, δ does not depend on x or y . Only on how far apart they are. Uniform continuity implies continuity, but the converse is false. Most functions on the real line are not uniformly continuous, but they are on compact intervals. In order to prove this we introduce an equivalent idea.

Definition 2.14. A function $f : X \subseteq \mathbb{R} \rightarrow \mathbb{R}$ is sequentially uniformly continuous if given $x_n, y_n \in X$, $y_n - x_n \rightarrow 0$ implies $f(y_n) - f(x_n) \rightarrow 0$.

The proof of the following is straightforward and we omit it.

Theorem 2.15. *A function $f : X \rightarrow \mathbb{R}$ is sequentially uniformly continuous if and only if it is uniformly continuous.*

Now we will prove a major result.

Theorem 2.16. *A continuous function on a closed bounded interval $[a, b]$ is uniformly continuous.*

Proof. Suppose that f is not uniformly continuous. It therefore cannot be sequentially uniformly continuous. Choose $r \geq 0$ such that for every $\delta > 0$ there exists $x, y \in [a, b]$ such that $|x - y| < \delta$ and $|f(x) - f(y)| > r$.

For each $N \in \mathbb{N}$, choose $x_n, y_n \in [a, b]$ such that

$$|x_n - y_n| < \frac{1}{n}, \text{ and } |f(x_n) - f(y_n)| \geq r.$$

By the Bolzano-Weierstrass Theorem, $\{x_n\}_{n=1}^\infty$ has a convergent subsequence $\{x_{n_K}\}_{K=1}^\infty$. Suppose that $x_{n_K} \rightarrow x$. Since $\{x_{n_K} - y_{n_K}\}_{K=1}^\infty$ is a subsequence of $\{x_n - y_n\}_{n=1}^\infty$ and $x_n - y_n \rightarrow 0$, so $x_{n_K} - y_{n_K} \rightarrow 0$. So we have

$$y_{n_K} = x_{n_K} - (x_{n_K} - y_{n_K}) \rightarrow x - 0 = x.$$

But f is continuous on $[a, b]$ and hence at x . So $f(x_{n_K}) \rightarrow f(x)$. and $f(y_{n_K}) \rightarrow f(x)$ and so $f(x_{n_K}) - f(y_{n_K}) \rightarrow 0$. But we have assumed that

$$|f(x_{n_K}) - f(y_{n_K})| \geq r > 0, \quad (2.5)$$

for all $K > 0$. We have a contradiction. So f is sequentially uniformly continuous and hence uniformly continuous. \square

The fact that continuous functions on closed and bounded intervals are uniformly continuous is essential to many other results. For example, the proof of Riemann's theorem that every continuous function is Riemann integrable requires it. So does the proof of the Fundamental Theorem of Calculus.

Another widely used type of continuity is Lipschitz continuity.

Definition 2.17. A function $f : X \rightarrow \mathbb{R}$ is Lipschitz continuous if there exists a constant M such that

$$\text{for every } x, y \in X, |f(x) - f(y)| \leq M|x - y|.$$

Lipschitz continuous functions are obviously continuous and in fact uniformly continuous.

2.4. Maxima, Minima and the Intermediate Value Property.

We now turn to another of the big results about continuous functions. This is about maxima and minima.

Theorem 2.18. *A continuous function on a closed, bounded interval $[a, b]$ is bounded. Moreover it attains its maximum and minimum values on $[a, b]$.*

Proof. Suppose that f is unbounded. Then given $n \in \mathbb{N}$, n is not a bound for f and thus there exists $x_n \in [a, b]$ such that $|f(x_n)| > n$. However, we know that $[a, b]$ is closed and bounded, and so the sequence $\{x_n\}_{n=1}^\infty$ has a convergent subsequence $\{x_{n_K}\}_{K=1}^\infty$. Suppose that $x_{n_K} \rightarrow x$ as $K \rightarrow \infty$. By continuity of f , $f(x_{n_K}) \rightarrow f(x)$. But this is impossible, since $f(x_{n_K}) > n_K$ for each K and $n_K \rightarrow \infty$, so the

sequence $\{f(x_{n_K})\}_{K=1}^\infty$ is not convergent, and hence f is not continuous at x . This is a contradiction and we therefore conclude that f is bounded.

Now suppose that $M = \sup_{x \in [a,b]} f(x)$. For each $n \in \mathbb{N}$ choose $x_n \in [a, b]$ such that $f(x_n) > M - 1/n$. Then $f(x_n) \rightarrow M$. $\{x_n\}_{n=1}^\infty$ is contained in $[a, b]$, so is bounded and hence has a convergent subsequence $\{x_{n_K}\}_{K=1}^\infty$. Suppose $x_{n_K} \rightarrow c \in [a, b]$. By continuity, $f(x_{n_K}) \rightarrow f(c)$. But the sequence $\{f(x_n)\}_{n=1}^\infty$ is convergent, so the sequence $\{f(x_{n_K})\}_{K=1}^\infty$ has the same limit. Thus $f(c) = M$, so f reaches its maximum. The case for the minimum is similar. \square

We also need to mention the intermediate value property. This is the result which tells us that we can solve certain equations.

Theorem 2.19. *Suppose $f : [a, b] \rightarrow \mathbb{R}$ is continuous on $[a, b]$ and $f(a)f(b) < 0$. Then there is a $c \in [a, b]$ such that $f(c) = 0$.*

Proof. Without loss of generality, we suppose that $f(a) < 0, f(b) > 0$. Let $A = \{x \in [a, b] : f(x) < 0\}$. Then $a \in A$ and so A is nonempty and bounded above. It therefore has a least upper bound, which we we call c . Choose x_n such that $c - 1/n < x_n \leq c$. Then $f(x_n) < 0$. By continuity, $f(c) = \lim_{n \rightarrow \infty} f(x_n) \leq 0$. Now take $y_n = c + (b - c)/n$. Then $y_n \rightarrow c$ and by continuity $f(c) = \lim_{n \rightarrow \infty} f(y_n) \geq 0$. Hence $f(c) = 0$. The case $f(a) > 0$ and $f(b) < 0$ is similar. \square

Corollary 2.20. *Let f be continuous on $[a, b]$. Suppose that $f(a) \neq f(b)$ and that M lies between $f(a)$ and $f(b)$. Then there is a $c \in [a, b]$ such that $f(c) = M$.*

Proof. Apply Theorem 2.19 to the function $g(x) = f(x) - M$. \square

Definition 2.21. A function is said to be monotone increasing if for each $x \geq y$ we have $f(x) \geq f(y)$. We say that f is monotone decreasing if $f(y) \leq f(x)$.

An important question that arose at the end of the nineteenth century was which functions are differentiable? Weierstrass had constructed a nowhere differentiable function, an event that was a considerable shock to mathematicians. Lebesgue proved that every monotone function is differentiable ‘almost everywhere’. As a first step we can show that monotone functions are continuous except possibly on a countable set of points.

Theorem 2.22. *Suppose that f is monotone on (a, b) . Then f is continuous except possibly on a countable set of points in (a, b) .*

Proof. Without loss of generality we can assume that f is increasing. If f is decreasing we can just multiply by minus one to obtain an increasing function. We can also assume that (a, b) is bounded. Otherwise we can write it as a countable union of open, bounded subintervals and

the discontinuities of f will be a countable union of the discontinuities on each subinterval. Now let $x_0 \in (a, b)$ and then by the least upper bound axiom

$$f(x_0^-) = \sup\{f(x) : a < x < x_0\}, \quad (2.6)$$

and

$$f(x_0^+) = \inf\{f(x) : x_0 < x < b\}, \quad (2.7)$$

both exist and $f(x_0^-) \leq f(x_0^+)$. The only way that f can be discontinuous at x_0 is if there is a jump at x_0 . We define the jump interval at x_0 by $J(x_0) = \{y \in (f(x_0^-), f(x_0^+))\}$. Clearly $J(x_0) \subseteq (a, b)$ and so it is bounded. The jump intervals for f are also obviously disjoint. So for every $n \in \mathbb{N}$, there are only a finite number of jump intervals of length greater than $1/n$. Hence the set of points of discontinuity of f is a countable union of finite sets and is therefore countable.

□

3. DIFFERENTIATION AND ITS APPLICATIONS

Now we can begin to develop calculus. The notion of the derivative should already be familiar from a first course in calculus. We will develop the properties of the derivative systematically with an emphasis on rigour. Naturally we begin with the definition.

3.1. The Derivative. The derivative is one of the two major tools of calculus. It is the limit of the Newton quotient.

Definition 3.1. A function $f : X \rightarrow \mathbb{R}$, where X is open, is said to be differentiable at x if

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}, \quad (3.1)$$

exists. We say that $f'(x)$ is the derivative of f at x . We also write $\frac{df}{dx}$ for f' .

An equivalent formulation is

$$f'(x_0) = \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0}. \quad (3.2)$$

Derivatives are defined on open sets. So one talks about a function being differentiable on an open interval (a, b) rather than on $[a, b]$, because the limit in the definition is not necessarily defined at the end points of an interval.

Example 3.1. Let $f(x) = x^n$. By the Binomial Theorem $(x+h)^n = x^n + nx^{n-1}h + n(n-1)x^{n-2}h^2 + \dots$. So

$$\begin{aligned} f'(x) &= \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \\ &= \lim_{h \rightarrow 0} \frac{x^n + nx^{n-1}h + n(n-1)x^{n-2}h^2 + \dots - x^n}{h} \\ &= \lim_{h \rightarrow 0} (nx^{n-1} + hn(n-1)x^{n-2} + \dots) \\ &= nx^{n-1}. \end{aligned}$$

This familiar rule is the most important derivative in calculus.

In order to treat the exponential function, we need to introduce the natural logarithm. We have not introduced the integral yet, however we will assume here some familiarity with integration, since it makes our discussion much easier at this point.

Example 3.2. We define the natural logarithm of $y > 0$ to be

$$\ln y = \int_1^y \frac{1}{t} dt. \quad (3.3)$$

Then the exponential is the inverse of the natural logarithm. So that $\ln(e^y) = y$. The Fundamental Theorem of Calculus (which we prove later) shows that

$$\frac{d}{dy} \ln y = \frac{1}{y}, \quad y > 0.$$

Example 3.3. Let $f(x) = e^x$. Then it is not hard to show that

$$\begin{aligned} f'(x) &= \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \\ &= \lim_{h \rightarrow 0} \frac{e^{x+h} - e^x}{h} \\ &= e^x \lim_{h \rightarrow 0} \frac{e^h - 1}{h}. \end{aligned}$$

In Tutorial Two we prove that $\lim_{n \rightarrow \infty} (1 + \frac{1}{n})^n = e$. In fact it is also true that $\lim_{n \rightarrow \infty} (1 + \frac{h}{n})^n = e^h$. Now by the Binomial Theorem

$$\frac{1}{h} \left(\left(1 + \frac{h}{n}\right)^n - 1 \right) = 1 + \sum_{k=2}^n \binom{n}{k} \frac{h^{k-1}}{n^k}.$$

Using the ratio test one can prove that the series $\sum_{k=2}^n \binom{n}{k} \frac{h^{k-1}}{n^k}$ is convergent for all h . So let

$$g(h) = \lim_{n \rightarrow \infty} \sum_{k=2}^n \binom{n}{k} \frac{h^{k-1}}{n^k}.$$

Then we have

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{e^h - 1}{h} &= \lim_{h \rightarrow 0} \lim_{n \rightarrow \infty} \left(1 + \sum_{k=2}^n \binom{n}{k} \frac{h^{k-1}}{n^k} \right) \\ &= \lim_{h \rightarrow 0} (1 + hg(h)) = 1. \end{aligned}$$

Thus $f'(x) = e^x$, so the exponential function is its own derivative.

Example 3.4. We compute the derivative of $f(x) = \sin x$. To do this we need the expansion formula for sine. We have

$$\begin{aligned} f'(x) &= \lim_{h \rightarrow 0} \frac{\sin(x+h) - \sin x}{h} \\ &= \lim_{h \rightarrow 0} \frac{\sin x \cos h + \sin h \cos x - \sin x}{h} \\ &= \lim_{h \rightarrow 0} \left(\frac{\sin x (\cos h - 1)}{h} + \frac{\sin h \cos x}{h} \right). \end{aligned}$$

Now $\lim_{h \rightarrow 0} \frac{\sin h}{h} = 1$. (This is a well known result. It is also a tutorial exercise). For small h , $\cos h \approx 1$. So we have $\frac{\cos h - 1}{h} \approx 0$ for small h . Indeed $\frac{\cos h - 1}{h} \rightarrow 0$ as $h \rightarrow 0$. Thus $f'(x) = \cos x$.

The basic rules of differentiation are well known. We state them here for convenience.

Theorem 3.2. *Let c be constant and f, g be differentiable at x_0 . Then*

$$(cf)'(x_0) = cf'(x_0) \quad (3.4)$$

$$(f + g)'(x_0) = f'(x_0) + g'(x_0) \quad (3.5)$$

$$(fg)'(x_0) = f'(x_0)g(x_0) + f(x_0)g'(x_0) \quad (3.6)$$

Proof. Once more this an exercise manipulating limits. For example, the product rule is proved as follows.

$$\begin{aligned} (fg)'(x_0) &= \lim_{x \rightarrow x_0} \frac{(fg)(x) - (fg)(x_0)}{x - x_0} \\ &= \lim_{x \rightarrow x_0} \left[\frac{f(x)g(x) - f(x)g(x_0) + f(x)g(x_0) - f(x_0)g(x_0)}{x - x_0} \right] \\ &= \lim_{x \rightarrow x_0} f(x) \frac{g(x) - g(x_0)}{x - x_0} + \lim_{x \rightarrow x_0} g(x_0) \frac{f(x) - f(x_0)}{x - x_0} \\ &= f(x_0)g'(x_0) + f'(x_0)g(x_0). \end{aligned}$$

The remaining proofs are exercises. \square

The next result is easy to prove and will be used in the proof of the chain rule.

Theorem 3.3. *If f is differentiable at a point x , then it is continuous at x .*

Proof. We can write

$$f(x) = (x - x_0) \left(\frac{f(x) - f(x_0)}{x - x_0} \right) + f(x_0) \quad (3.7)$$

Since f is differentiable at x_0 , we have

$$\begin{aligned} \lim_{x \rightarrow x_0} f(x) &= \lim_{x \rightarrow x_0} \left((x - x_0) \left(\frac{f(x) - f(x_0)}{x - x_0} \right) + f(x_0) \right) \\ &= \lim_{x \rightarrow x_0} (x - x_0) \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} + f(x_0) \\ &= 0 \times f'(x_0) + f(x_0) = f(x_0). \end{aligned}$$

So f is continuous at x_0 . \square

Again the converse of this result is false. The function $f(x) = |x|$ is continuous at zero, but is not differentiable there. Indeed Karl Weierstrass (1815-1897) proved that there are functions which are continuous everywhere, but differentiable nowhere. This discovery astonished nineteenth century mathematicians and led to a deep investigation of the properties of continuous functions.

The most important result is the chain rule. This is the result that is used more often than any other in the whole of calculus. One cannot

understand calculus without understanding the chain rule. It tells us how to differentiate functions of functions.

Theorem 3.4 (The Chain Rule). *Suppose that g is differentiable at x and f is differentiable at $y = g(x)$. Then*

$$(f \circ g)'(x) = f'(y)g'(x). \quad (3.8)$$

Proof. Write $k = g(x + h) - g(x)$. Since g is differentiable at x , it is continuous there and so as $h \rightarrow 0$, $k \rightarrow 0$. Now

$$\begin{aligned} \frac{f(g(x + h)) - f(g(x))}{h} &= \frac{f(g(x + h)) - f(g(x))}{g(x + h) - g(x)} \frac{g(x + h) - g(x)}{h} \\ &= \frac{f(y + k) - f(y)}{k} \frac{g(x + h) - g(x)}{h}. \end{aligned}$$

Suppose that at no value of h does $k = 0$. Then taking the limit as $h \rightarrow 0$ gives the result. To take care of the case $k = 0$ we let

$$F(k) = \begin{cases} \frac{f(y+k)-f(y)}{k} & k \neq 0 \\ f'(y) & k = 0. \end{cases} \quad (3.9)$$

By differentiability of f , as $k \rightarrow 0$ $F(k) \rightarrow f'(y)$ and so F is continuous at 0. Thus as $h \rightarrow 0$, $F(k) \rightarrow f'(y)$. So for $k \neq 0$

$$\frac{f(g(x + h)) - f(g(x))}{h} = F(k) \frac{g(x + k) - g(x)}{h}. \quad (3.10)$$

This also holds when $k = 0$ since both sides will be zero. Consequently

$$\frac{f(g(x + h)) - f(g(x))}{h} \rightarrow f'(y)g'(x) \quad (3.11)$$

as $h \rightarrow 0$. □

Example 3.5. Let us compute the derivative of a reciprocal. We have $f(x) = 1/g(x) = h(g(x))$, where $h(u) = 1/u$. Hence

$$\frac{d}{dx} f(x) = g'(x)h'(u) = -\frac{g'(x)}{(g(x))^2}.$$

Example 3.6. The quotient rule is obtained by combining the chain rule and the product rule:

$$\begin{aligned} \frac{d}{dx} \frac{f(x)}{g(x)} &= \frac{f'(x)}{g(x)} + f(x) \frac{d}{dx} \frac{1}{g(x)} \\ &= \frac{f'(x)}{g(x)} - \frac{f(x)g'(x)}{g(x)^2} \\ &= \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}. \end{aligned}$$

3.2. Maxima and Minima. The first application of differentiation that we see is usually to the problem of obtaining maxima and minima.

Definition 3.5. A function $f : X \rightarrow \mathbb{R}$ has a local maximum at $c \in X$ if there is a subset $Y \subseteq X$ such that $c \in Y$ and $f(c) \geq f(x)$ for all $x \in Y$. A point c is a local minimum for f if there is a subset $Y \subseteq X$ such that $c \in Y$ and $f(c) \leq f(x)$ for all $x \in Y$. If f has a local maximum at c , then c is called a maximiser. If f has a local minimum at c , then c is called a minimiser. In general c is called an extreme point.

Theorem 3.6. Let I be an open interval in \mathbb{R} , $f : I \rightarrow \mathbb{R}$ be differentiable at $c \in I$. If f attains a local maximum or minimum at c , then $f'(c) = 0$.

Proof. There are two cases to consider, which turn out to be very similar. So we only prove the case for a local maximum. The proof proceeds by contradiction, so we assume that c is a point where f attains a local maximum and that $f'(c) > 0$. Choose $\delta > 0$ such that for $x \in I$ and $0 < |x - c| < \delta$ we have

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

Pick an $x > c$ with $|x - c| < \delta$. Then we have

$$-f'(c) < \frac{f(x) - f(c)}{x - c} - f'(c) < f'(c).$$

Which implies

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

and hence $f(x) > f(c)$, which is a contradiction. Thus $f'(c) \leq 0$.

Suppose then that $f'(c) < 0$. Pick a $\delta > 0$ such that for $x \in I$ and $0 < |x - c| < \delta$ we have

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

Pick an $x < c$ with $|x - c| < \delta$. Then

$$f'(c) < \frac{f(x) - f(c)}{x - c} - f'(c) < -f'(c).$$

Which implies

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

and hence $f(x) < f(c)$, since $x - c < 0$, which is a contradiction once more. Thus $f'(c) = 0$. The proof for a local minimum is essentially the same. \square

3.3. Mean Value Theorems. A useful corollary of the previous result is called Rolle's Theorem. It is named after Michel Rolle, (1652-1719) a French mathematician. He actually invented Gaussian elimination in Linear Algebra, well before Gauss. Rolle's Theorem was published in 1691, but was rediscovered a number of times before having Rolle's name attached to it.

Theorem 3.7 (Rolle's Theorem). *Let $[a, b]$ be a closed interval in \mathbb{R} and suppose that f is continuous on $[a, b]$ and differentiable on (a, b) . If $f(a) = f(b) = 0$ then there is a point $c \in (a, b)$ such that $f'(c) = 0$.*

Proof. Continuous functions attain their maximum and minimum values on closed bounded intervals. If $c \in (a, b)$ is an extreme point, then we already know that $f'(c) = 0$. Suppose now that both the maximum and minimum values occur at the end points. Then since $f(a) = f(b)$, it follows that f is constant and so $f'(x) = 0$ for all $x \in (a, b)$. \square

This simple fact is surprisingly useful. The main applications of Rolle's Theorem are to prove the Mean Value Theorem and Taylor's Theorem, which are two of the most useful results in the whole of analysis. We start with the Mean Value Theorem.

Theorem 3.8 (Mean Value Theorem). *Let $[a, b]$ be a closed and bounded interval on \mathbb{R} and $f : [a, b] \rightarrow \mathbb{R}$ a continuous function which is differentiable on (a, b) . Then there is a point $c \in (a, b)$ such that*

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

Proof. As we noted above, the proof is just an application of Rolle's Theorem. The key is to choose the right function. We let

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

Then $g(a) = g(b) = 0$ and

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

By Rolle's Theorem there is a $c \in (a, b)$ with $g'(c) = 0$, which proves the result. \square

The MVT is one of the most powerful results in calculus. Let us consider some simple applications. Later we will see it used to prove a result about the behaviour of limits for sequences of derivatives. One can also use it to prove the Fundamental Theorem of Calculus. It is quite ubiquitous.

Corollary 3.9. *If $[a, b]$ is a closed and bounded interval in \mathbb{R} and f is continuous on $[a, b]$ and differentiable on (a, b) , then f is Lipschitz continuous on $[a, b]$.*

Proof. For any $x, y \in (a, b)$ the MVT gives, $|f(x) - f(y)| \leq |f'(c)||x - y|$ for some $c \in (x, y)$. \square

The following result is well known from high school calculus, but usually is not given a rigorous proof.

Corollary 3.10. *If f is continuous on $[a, b]$ and differentiable on (a, b) , and $f'(x) = 0$ for all $x \in (a, b)$, then f is constant on $[a, b]$.*

Proof. For any $x, y \in (a, b)$, $f(x) - f(y) = f'(c)(x - y) = 0$. Hence $f(x) = f(y)$ for all x, y and so f is constant on (a, b) . By continuity it is also constant on $[a, b]$. \square

As a third application, we use it to prove uniqueness for the solution of a simple differential equation.

Proposition 3.11. *The equation $y' = ky, y(0) = y_0$ has a unique solution.*

Proof. We let $y(x) = y_0 e^{kx}$. Then this is clearly a solution of the differential equation. To see this, just substitute the function into the equation. Now suppose that f is any solution of the equation. Consider $h(x) = f(x)e^{-kx}$. Then

$$h'(x) = f'(x)e^{-kx} - ke^{-kx}f(x) = e^{-kx}(f'(x) - kf(x)) = 0.$$

Thus h is constant. Hence $f(x) = Ce^{kx}$. The condition that $f(0) = y_0$ proves the result. \square

There is a more general version of the MVT. It is due to Cauchy and is often called the Cauchy Mean Value Theorem. We call it the Generalised MVT.

Theorem 3.12 (Generalised Mean Value Theorem). *Suppose that f and g are continuous functions on $[a, b]$, which are differentiable on (a, b) and suppose that $g'(x) \neq 0$ for all $x \in (a, b)$. Then there exists a point $c \in (a, b)$ such that*

$$\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(c)}{g'(c)}. \quad (3.12)$$

Proof. This again relies upon Rolle's Theorem. First, observe that if $g(b) - g(a) = 0$, then the Mean Value Theorem tells us that there exists a point $c \in (a, b)$ such that $g'(c) = 0$. However we have assumed that g' is nonzero, so $g(b) - g(a) \neq 0$. Next introduce the function

$$h(x) = f(x)[g(b) - g(a)] - g(x)[f(b) - f(a)].$$

Then

$$\begin{aligned} h(a) &= f(a)[g(b) - g(a)] - g(a)[f(b) - f(a)] \\ &= f(a)g(b) - f(b)g(a) = h(b). \end{aligned}$$

Rolle's Theorem then tells us that there is a $c \in (a, b)$ such that $h'(c) = 0$. Which means that

$$f'(c)[g(b) - g(a)] - g'(c)[f(b) - f(a)] = 0. \quad (3.13)$$

Rearranging gives the result. \square

As an application of this result we prove L'Hôpital's rule.

Theorem 3.13. *Suppose that f and g are differentiable on (a, b) and that $g(x) \neq 0$ and $g'(x) \neq 0$ for all $x \in (a, b)$. Suppose further that $\lim_{x \rightarrow a^+} f(x) = \lim_{x \rightarrow a^+} g(x) = 0$. Then,*

$$\lim_{x \rightarrow a^+} \frac{f(x)}{g(x)} = \lim_{x \rightarrow a^+} \frac{f'(x)}{g'(x)}, \quad (3.14)$$

provided the right side exists.

Proof. Suppose that

$$\lim_{x \rightarrow a^+} \frac{f'(x)}{g'(x)} = L.$$

Then given $\epsilon > 0$ we can find $\delta > 0$ such that if $c \in (a, a + \delta)$ then

$$\left| \frac{f'(c)}{g'(c)} - L \right| < \epsilon.$$

However, by the generalised MVT, if $x \in (a, a + \delta)$ then

$$\left| \frac{f(x)}{g(x)} - L \right| = \left| \frac{f(x) - f(a)}{g(x) - g(a)} - L \right| < \epsilon.$$

\square

The extension of this result to the case when

$$\lim_{x \rightarrow a} f(x) = \lim_{x \rightarrow a} g(x) = \infty,$$

can also be established using the generalised MVT.

Remark 3.14. L'Hôpital's rule was actually discovered by the Swiss mathematician Johann Bernoulli, who taught Euler and worked for L'Hôpital. L'Hôpital published the rule in his textbook on calculus, and it became known by his name. Many results in mathematics are not named after the people who originally discovered them.

3.4. Inverse Functions. We first state our definitions.

Definition 3.15. A function $f : X \rightarrow Y$ is said to be one to one if for each $y \in Y$ there is at most one $x \in X$ such that $f(x) = y$. We also say that such an f is a bijection. If $f : X \rightarrow Y$ is one to one then it has an inverse function $f^{-1} : Y \rightarrow X$ which satisfies

$$f(f^{-1}(f)) = f^{-1}(f(x)) = x,$$

for all $x \in X$.

Suppose that $f : X \subseteq \mathbb{R} \rightarrow \mathbb{R}$ is strictly increasing (or decreasing). Then f is clearly one to one, and hence it has an inverse. If f is continuous, then the inverse function will also be continuous.

Theorem 3.16. *Suppose that $f : X \subseteq \mathbb{R} \rightarrow Y$ is a strictly increasing (or decreasing) continuous function. Then the inverse function f^{-1} exists and is continuous and increasing (or decreasing) on $f(X)$.*

Proof. We only deal with the case when f is increasing. We show that f^{-1} is increasing. Assume not. Then we can find $y_1, y_2 \in Y$ with $y_2 > y_1$ and $f^{-1}(y_2) < f^{-1}(y_1)$. But f is increasing, so

$$f(f^{-1}(y_2)) < f(f^{-1}(y_1)),$$

so that $y_2 < y_1$ which is a contradiction.

To prove continuity, take $y_0 \in f(X)$. Then there exists $x_0 \in X$ with $f(x_0) = y_0$. We suppose that y_0 is not an endpoint, so x_0 is not an endpoint and we may find $\epsilon_0 > 0$ such that the interval

$$(f^{-1}(y_0) - \epsilon_0, f^{-1}(y_0) + \epsilon_0) \subset X.$$

Pick $\epsilon < \epsilon_0$. Then there exist $y_1, y_2 \in f(X)$ such that $f^{-1}(y_1) = f^{-1}(y_0) - \epsilon$ and $f^{-1}(y_2) = f^{-1}(y_0) + \epsilon$. Because f is increasing $y_1 < y_0 < y_2$ and the inverse is increasing so for all $y \in (y_1, y_2)$ we have the inequality

$$f^{-1}(y_0) - \epsilon < f^{-1}(y) < f^{-1}(y_0) + \epsilon.$$

Consequently, if $\delta = \min\{y_2 - y_0, y_0 - y_1\}$, then

$$|f^{-1}(y_0) - f^{-1}(y)| < \epsilon$$

whenever $|y_0 - y| < \delta$. So f^{-1} is continuous at y_0 .

We can also prove that if y_0 is a left (or right) endpoint, then f^{-1} is left (or right) continuous at y_0 . \square

The most important result about inverse functions relates the derivative of f and that of f^{-1} .

Theorem 3.17 (The Inverse Function Theorem). *Suppose that f is differentiable and one to one on an open interval I . If $f'(a) \neq 0$, $a \in I$, then f^{-1} exists and is differentiable at $f(a)$ and*

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

Proof. Since f' is nonzero on I , it follows that f is either increasing or decreasing on I , and hence f is invertible. The inverse is continuous. Since f is decreasing or increasing, for $x \neq a$ it follows that $f(x) \neq f(a)$.

Now

$$\begin{aligned} \lim_{y \rightarrow f(a)} \frac{f^{-1}(y) - f^{-1}(f(a))}{y - f(a)} &= \lim_{f(x) \rightarrow f(a)} \frac{f^{-1}(f(x)) - f^{-1}(f(a))}{f(x) - f(a)} \\ &= \lim_{x \rightarrow a} \left(\frac{x - a}{f(x) - f(a)} \right)^{-1} \\ &= \frac{1}{f'(a)}. \end{aligned}$$

□

3.5. Convex Functions. An interesting and important class of functions for which we can establish some very general results about differentiability are convex functions. We begin with the definition.

Definition 3.18. A function $f : I \rightarrow \mathbb{R}$ is said to be convex on an open interval I if for all $x \in I$ and $a > 0, b > 0$ with $a + b = 1$, we have

$$f(ax + by) \leq af(a) + bf(b).$$

If

$$f(ax + by) \geq af(a) + bf(b),$$

then f is said to be concave.

An equivalent formulation of convexity is that for all $x_1, x_2, x_3 \in I$ with $x_1 < x_2 < x_3$ we have

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} \leq \frac{f(x_3) - f(x_2)}{x_3 - x_2}. \quad (3.15)$$

Convex functions are automatically continuous. To prove this we have a preliminary result.

Proposition 3.19. *If f is convex on an open interval $I \subset \mathbb{R}$, then the left and right hand derivatives, defined respectively by,*

$$f'(x^+) = \lim_{h \rightarrow 0^+} \frac{f(x + h) - f(x)}{h},$$

and

$$f'(x^-) = \lim_{h \rightarrow 0^-} \frac{f(x + h) - f(x)}{h},$$

both exist for each $x \in I$. Moreover, if $x, y \in (a, b)$ and $y > x$, then

$$f'(x^-) \leq f'(x^+) \leq \frac{f(y) - f(x)}{y - x} \leq f'(y^-) \leq f'(y^+). \quad (3.16)$$

Proof. We let $0 < h_1 < h_2$, then observe that

$$\frac{f(x + h_1) - f(x_1)}{h_1} \leq \frac{f(x + h_2) - f(x_1)}{h_2}.$$

Hence

$$F(h) = \frac{f(x + h) - f(x_1)}{h},$$

is an increasing function on some interval $(0, \delta)$ and hence $\lim_{h \rightarrow 0^+} F(h)$ exists. Similarly for the second limit. The inequality follows from (3.15) and is an easy exercise. \square

Note, this result does not mean that f is differentiable at x . We have not established equality of the limits and in fact, this may not hold. However, it is a remarkable fact that convex functions are differentiable except possibly on a countable set of points. We will prove this below.

An application of the mean value theorem allows us to establish the following test for convexity.

Theorem 3.20. *Suppose that f is differentiable on an open interval I . Then f is convex on I if and only if f' increases on I .*

Proof. Suppose that f' is increasing on I and pick three points $x_1 < x_2 < x_3 \in I$. Then by the mean value theorem there exists points $a, b \in I$ with $b > a$ such that

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} = f'(a), \quad (3.17)$$

and

$$\frac{f(x_3) - f(x_2)}{x_3 - x_2} = f'(b). \quad (3.18)$$

Now f' is increasing, hence $f'(a) \leq f'(b)$ and thus f is convex.

Conversely, suppose that f is convex. Then for points $x_1 < x_2 < x_3 < x_4$ we have

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} \leq \frac{f(x_3) - f(x_2)}{x_3 - x_2} \leq \frac{f(x_4) - f(x_3)}{x_4 - x_3}. \quad (3.19)$$

Letting $x_2 \rightarrow x_1^+$ and $x_2 \rightarrow x_4^-$ shows that $f'(x_3) \leq f'(x_4)$ and so f is increasing. \square

For twice differentiable functions we have a simple result.

Theorem 3.21. *Let f be twice differentiable on an open interval I . Then f is convex on I if and only if $f''(x) \geq 0$ for all $x \in I$.*

Now we prove continuity for convex functions.

Theorem 3.22. *Suppose that f is convex on an open interval I . Then f is continuous on I .*

Proof. We let $x \in I$. Then

$$\lim_{h \rightarrow 0^+} (f(x+h) - f(x)) = \lim_{h \rightarrow 0^+} \left(\frac{f(x+h) - f(x)}{h} \right) h = 0,$$

and

$$\lim_{h \rightarrow 0^-} (f(x+h) - f(x)) = \lim_{h \rightarrow 0^-} \left(\frac{f(x+h) - f(x)}{h} \right) h = 0.$$

Thus both limits exist and are equal, so f is continuous at x . \square

Actually, convex functions are not just continuous.

Proposition 3.23. *Let f be a convex function on (a, b) . Then f is Lipschitz continuous on each closed bounded subinterval $[c, d]$ of (a, b) .*

Proof. This follows from the inequality

$$f'(c^+) \leq f'(u^+) \leq \frac{f(v) - f(u)}{v - u} \leq f'(v^-) \leq f'(d^-), \quad (3.20)$$

valid for $c \leq u \leq v \leq d$. So that for all $u, v \in [c, d]$ with $M = \max\{|f'(c^+)|, |f'(d^-)|\}$ we have

$$|f(u) - f(v)| \leq M|u - v|.$$

□

Indeed we can show something even stronger.

Theorem 3.24. *A convex function f on an interval (a, b) is differentiable except at most on a countable set of points. Moreover, the derivative is an increasing function.*

Proof. We already know that the left and right derivatives at a point exist. They are also increasing, and so they are continuous except at most on a countable set of points, which we denote by \mathcal{D} . Take a point $x \in (a, b) - \mathcal{D}$ and let $x_n \rightarrow x^+$ and apply the inequality (3.20). This gives

$$f'(x^-) \leq f'(x^+) \leq f'(x^-)$$

so that $f'(x^+) = f'(x^-)$ and hence f is differentiable at x . That f is increasing also follows from (3.20). □

Lebesgue proved a deeper result about differentiability. Namely that any monotone function is differentiable almost everywhere. This is an extremely difficult result to prove. “Almost everywhere” has a specific meaning and the proof of Lebesgue’s Theorem usually uses what are known as *Dini Derivatives*.

We can define higher derivatives in the obvious way. So

$$\frac{d^2 f}{dx^2} = \frac{d}{dx} \left(\frac{df}{dx} \right),$$

or $f''(x) = (f')'(x)$ etc.

Definition 3.25. A function $f : X \rightarrow \mathbb{R}$ for which the n th derivative $f^{(n)}$ exists for all $n \in \mathbb{N}$ is said to be infinitely differentiable, or smooth. We write $f \in C^\infty(X)$. (Pronounced C infinity on X). If f is n times differentiable for finite n we write $f \in C^n(X)$.

3.6. Power Series and Taylor Expansions. A power series about a point x_0 is an expression of the form

$$f(x) = \sum_{n=0}^{\infty} a_n(x - x_0)^n.$$

Power series are one of the most important tools in Analysis. We will see that many functions can be written as power series.

For example it turns out that the sine function can be expressed as

$$\sin x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}.$$

This is an example of a *Taylor* series, which will be the culmination of this section.

By the ratio test a power series will converge if

$$\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}(x - x_0)^{n+1}}{a_n(x - x_0)^n} \right| = L < 1.$$

Upon rewriting this becomes

$$|x - x_0| \lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| = L < 1. \quad (3.21)$$

We can think of this as determining the values of x for which the series converges.

Definition 3.26. Suppose that for the series $\sum_{n=0}^{\infty} a_n(x - x_0)^n$

$$|x - x_0| \lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| = L < 1 \quad (3.22)$$

for all $|x - x_0| < R$. We call R the radius of convergence of the power series.

Note a power series with radius of convergence R may converge or diverge when $|x - x_0| = R$. One has to check convergence at the end points individually.

Example 3.7. The series $1 + x + x^2 + \cdots = \sum_{n=0}^{\infty} x^n$ is convergent for all $|x| < 1$. Hence the radius of convergence is 1.

For simplicity we will take $x_0 = 0$ in what follows. All results can be transferred to the more general case by making the replacement $x \rightarrow x - x_0$.

Power series have very nice properties. In particular they converge absolutely within their radius of convergence.

Theorem 3.27. Let $\sum_{n=0}^{\infty} a_n x^n$ be a power series with radius of convergence R . Then the series converges absolutely for $|x| < R$ and diverges for $|x| > R$.

Proof. Let $t \in (-R, R)$, then $\sum_{n=0}^{\infty} a_n t^n$ converges and the sequence $a_n t^n \rightarrow 0$ and is thus bounded. Let M be a bound. Now pick x with $|x| < |t|$, then

$$|a_n x^n| = |a_n t^n| \left| \frac{x}{t} \right|^n \leq M r^n$$

where $r = |x/t| < 1$. But $\sum_{n=0}^{\infty} M r^n$ is a convergent geometric series, and so $\sum_{n=0}^{\infty} |a_n x^n|$ converges by the comparison test. The other result is similar. \square

Power series actually converge uniformly, a result we prove later. An important fact is that we can differentiate power series term by term and this does not change the radius of convergence.

Theorem 3.28. *Let $\sum_{n=0}^{\infty} a_n x^n$ have radius of convergence R . Then the power series $\sum_{n=1}^{\infty} n a_n x^{n-1}$ has radius of convergence R .*

Proof. Suppose that the series $\sum_{n=1}^{\infty} n a_n x^{n-1}$ has radius of convergence $R_d < R$. Choose r, s so that $R_d < r < s < R$. Clearly $\sum_{n=0}^{\infty} a_n s^n$ converges which shows that $a_n s^n \rightarrow 0$ and so is bounded by a constant M . Then

$$\begin{aligned} |n a_n r^{n-1}| &= n |a_n| s^{n-1} \left(\frac{r}{s} \right)^{n-1} \\ &\leq \frac{M}{s} n \left(\frac{r}{s} \right)^{n-1}. \end{aligned}$$

Now

$$\lim_{n \rightarrow \infty} \frac{(M/s)(n+1)(r/s)^n}{(M/s)n(r/s)^{n-1}} = \frac{r}{s} < 1.$$

Thus the series $\sum_{n=1}^{\infty} \frac{M}{s} n \left(\frac{r}{s} \right)^{n-1}$ is convergent by the ratio test. Thus $\sum_{n=1}^{\infty} n a_n r^{n-1}$ is absolutely convergent, which is a contradiction since $r > R_d$. Hence $R \leq R_d$. Similarly we show that $R_d > R$ leads to a contradiction. (Exercise). Hence $R = R_d$. \square

From this we can establish an important corollary.

Theorem 3.29. *Let $\sum_{n=0}^{\infty} a_n x^n$ be a power series with radius of convergence $R > 0$. Let $f : (-R, R) \rightarrow \mathbb{R}$ be defined by*

$$f(x) = \sum_{n=0}^{\infty} a_n x^n.$$

Then f is differentiable on $(-R, R)$ and

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

To prove this result we actually need some more information about the convergence of series. The key is that the series for f and f' both

converge uniformly. We will discuss uniform convergence later. So we will defer our proof till then.

The most commonly encountered power series are functions given by Taylor series expansions. These are named after Brook Taylor (1685-1731), an English mathematician. Perhaps you will not be surprised to learn that they were actually known before Taylor.

Definition 3.30. Let f be smooth in a neighbourhood X of a point a . We let the Taylor series for f at a be given by

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

If the series is convergent for all $x \in X$ and $|T_f(x) - f(x)| = 0$ for all $x \in X$, we say that f is analytic at a . If we truncate the Taylor expansion after n terms, the resulting expression is known as the n th Taylor polynomial. The quantities $\frac{1}{n!}f^{(n)}(a)$ are called the Taylor coefficients.

Even if the Taylor series does not converge, smooth functions can be approximated by Taylor polynomials.

Theorem 3.31 (Taylor's Theorem). *Let I be an open interval in \mathbb{R} , $n \in \mathbb{N}$ and $f \in C^{n+1}(I)$. Let $a \in I$ and $x \in I$, with $x \neq a$. Then there is a point ξ between a and x such that*

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

Proof. The proof uses Rolle's Theorem and is conceptually similar to the proof of the MVT. We define a function

$$\begin{aligned} F(t) = & f(x) - f(t) - f'(t)(x-t) - \frac{1}{2!}f''(t)(x-t)^2 - \cdots \\ & - \frac{f^{(n)}(t)}{n!}(x-t)^n. \end{aligned} \tag{3.23}$$

Plainly $F(x) = 0$. Since $f \in C^{(n+1)}(I)$ we see that F is differentiable. Now

$$\begin{aligned} F'(t) = & -f'(t) - f''(t)(x-t) + f'(t) - \frac{f'''(t)}{2!}(x-t)^2 + 2\frac{f''(t)}{2!}(x-t) \\ & - \cdots - \frac{f^{(n+1)}(t)}{n!}(x-t)^n + n\frac{f^{(n)}(t)}{n!}(x-t)^{n-1} \\ = & -\frac{f^{(n+1)}(t)}{n!}(x-t)^n. \end{aligned}$$

Next we introduce the function

$$G(t) = F(t) - \left(\frac{x-t}{x-a}\right)^{n+1} F(a).$$

Obviously $G(a) = 0$ and $G(x) = F(x) = 0$. Then

$$\begin{aligned} G'(t) &= F'(t) \\ &= -\frac{f^{(n+1)}(t)}{n!}(x-t)^n + (n+1)\frac{(x-t)^n}{(x-a)^{n+1}}F(a). \end{aligned}$$

By Rolle's Theorem there is a point ξ between x and a such that $G'(\xi) = 0$. That is

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

Rearranging we get

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

If we substitute this into (3.23) we have Taylor's Theorem. □

Example 3.8. Let us take the function $f(x) = e^x$. We let $a = 0$. So we are *expanding our function around zero*.¹ Now we have $f'(x) = e^x$, $f''(x) = e^x$. Indeed $f^{(n)}(x) = e^x$ for all n . So $f^{(n)}(0) = 1$. So by the formula for the Taylor coefficients we have the Taylor expansion

$$\begin{aligned} e^x &= 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \\ &= \sum_{n=0}^{\infty} \frac{x^n}{n!}. \end{aligned} \tag{3.24}$$

This in fact converges for all $x \in \mathbb{R}$. This is a tutorial exercise. This is one of the most important of all Taylor series.

We can use the previous example to define the number e . There are two equivalent definitions of e . Here we set

$$e = 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots = \sum_{n=0}^{\infty} \frac{1}{n!}. \tag{3.25}$$

Using this we can prove that e is an irrational number. This proof is due to Joseph Fourier (1768-1830). Fourier was one of the most important mathematicians in history. An entire branch of mathematics is named after him. This is called *Fourier Analysis*.²

¹Series where we expand about zero are also called MacLaurin series. These are named after the Scottish mathematician Colin MacLaurin (1698-1746).

²In an indirect way we owe our understanding of Egyptian hieroglyphics to Fourier. After Fourier visited Egypt with Napoleon in 1798, he brought back a number of artefacts with Ancient Egyptian writing on them. Then one day back in France, some visitors came to see him. They brought a young boy with them. The boy was incredibly gifted and Fourier promised to ensure that he got the education that he needed. The boy was also fascinated by the items from Egypt and asked what the strange symbols on them meant? Fourier told him that nobody knew.

Theorem 3.32. *The number e defined by (3.25) is irrational.*

Proof. We give a proof by contradiction. So we suppose that e is rational. This means that there are whole numbers a and b such that $e = \frac{a}{b}$. Since e is not an integer, $b > 1$. This means that

$$e = \frac{a}{b} = \sum_{n=0}^{\infty} \frac{1}{n!}.$$

Now suppose that N is an integer such that $N > b$. Then $N!e$ is an integer. This means that

$$N!e = \sum_{n=0}^N \frac{N!}{n!} + \sum_{n=N+1}^{\infty} \frac{N!}{n!}.$$

Now $n!$ divides $N!$ for all $n \leq N$. Hence $\sum_{n=0}^N \frac{N!}{n!}$ is an integer. Thus

$$N!e - \sum_{n=0}^N \frac{N!}{n!} = \sum_{n=N+1}^{\infty} \frac{N!}{n!},$$

is an integer. Now

$$\begin{aligned} \sum_{n=N+1}^{\infty} \frac{N!}{n!} &= \frac{1}{N+1} + \frac{1}{(N+1)(N+2)} + \frac{1}{(N+1)(N+2)(N+3)} + \cdots \\ &\leq \frac{1}{N+1} + \frac{1}{(N+1)^2} + \frac{1}{(N+1)^3} + \cdots \\ &= \frac{\frac{1}{(N+1)}}{1 - \frac{1}{N+1}} \\ &= \frac{1}{N} < 1. \end{aligned}$$

This is a contradiction since there are no whole number between 0 and 1. Thus e is not rational. \square

Example 3.9. We now find the Taylor series about zero of $\sin x$. We let $f(x) = \sin x$. Then $f'(x) = \cos x$. $f''(x) = -\sin x$. $f'''(x) = -\cos x$ and $f^{(iv)}(x) = \sin x$. Then the pattern repeats.

Now $f(0) = 0$, $f'(0) = 1$, $f''(0) = 0$, $f'''(0) = -1$ and $f^{(iv)}(0) = 0$.

The secret to reading hieroglyphics had been lost for nearly fifteen hundred years. So the boy, whose name was Jean-François Champollion, decided that he would work out what they meant. He did. He became one of history's greatest linguists and became fluent in numerous languages. Using the Rossetta Stone, a large tablet written in both Ancient Greek and Ancient Egyptian, he was able to crack the language. He also used the insights of the English physicist Thomas Young, who realised that hieroglyphics are largely phonetic, not pictographic. (e.g. A square represents the sound 'p'.) In 1822 Champollion began publishing his decipherment of the Ancient Egyptian writing system. Today his decipherment is universally accepted. Fourier also redesigned the French postal system so that it actually worked and discovered the greenhouse effect in his study of heat.

Hence the Taylor expansion about zero of the sine function is

$$\begin{aligned}\sin x &= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots \\ &= \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}.\end{aligned}\tag{3.26}$$

As with the Taylor series for e^x , this converges for every x . The proof of this is a tutorial exercise. Now try and establish the result

$$\begin{aligned}\cos x &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots \\ &= \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!}.\end{aligned}$$

This converges for all $x \in \mathbb{R}$.

Notice that $\sin x$ is an odd function and the series only involves odd powers of x . Whereas the $\cos x$ is an even function and it only involves even powers of x . This is true in general. An interesting question is whether an even function can have an expansion involving only odd powers and vice versa. If $a = 0$, then the answer is no. What if $a \neq 0$?

4. THE RIEMANN INTEGRAL

The other major tool in analysis is the integral. Although the Fundamental Theorem of Calculus was first stated by Newton and Leibnitz, the first rigorous theory of integration was developed by Cauchy, and extended by Riemann. We will develop the theory here.

We take an interval $[a, b]$ and partition it as

$$\mathcal{P} = \{x_0, x_1, \dots, x_n\},$$

where $x_0 = a$, $x_0 < x_1 < \dots < x_n$ and $x_n = b$.

Now let f be a bounded function on $[a, b]$ then define

$$M_i = \sup\{f(x) : x \in [x_{i-1}, x_i]\},$$

and

$$m_i = \inf\{f(x) : x \in [x_{i-1}, x_i]\}.$$

We then form the upper and lower Riemann sums

$$U(f, \mathcal{P}) = \sum_{i=1}^n M_i(x_i - x_{i-1}), \quad (4.1)$$

and

$$L(f, \mathcal{P}) = \sum_{i=1}^n m_i(x_i - x_{i-1}). \quad (4.2)$$

The least upper bound axiom establishes that the upper and lower integrals

$$\overline{\int_a^b} f = \inf\{U(f, \mathcal{P}) : \mathcal{P} \text{ a partition of } [a, b]\} \quad (4.3)$$

and

$$\underline{\int_a^b} f = \sup\{L(f, \mathcal{P}) : \mathcal{P} \text{ a partition of } [a, b]\} \quad (4.4)$$

both exist. We then say that f is Riemann integrable on $[a, b]$ if $\overline{\int_a^b} f = \underline{\int_a^b} f$. The Riemann integral is then equal to the upper (or lower) integral.

It is easy to prove the following results.

Proposition 4.1. *The Riemann integral has the following properties.*

(1) *If c is a constant, $\int_a^b c dx = c(b - a)$.*

(2) $\left| \int_a^b f(x) dx \right| \leq \int_a^b |f(x)| dx$.

The most important results about the Riemann integral are as follows.

Theorem 4.2 (Riemann's Criterion). *Let f be a bounded function on the closed interval $[a, b]$. Then f is Riemann integrable on $[a, b]$ if and only if, given any $\epsilon > 0$, there exists a partition \mathcal{P} of $[a, b]$ such that $U(f, \mathcal{P}) - L(f, \mathcal{P}) < \epsilon$.*

From this one establishes the first major result.

Theorem 4.3. *Every continuous function on a closed bounded interval $[a, b]$ is Riemann integrable.*

Proof. The function f is continuous on $[a, b]$ and so is bounded. Let $\epsilon > 0$. Since f is continuous it is uniformly continuous and so we can choose $\delta > 0$ such that $x, y \in [a, b]$ with $|x - y| < \delta$ implies $|f(x) - f(y)| < \epsilon/(b - a)$. Now choose $N \in \mathbb{N}$ such that $N > (b - a)/\delta$. For each $i = 0, 1, \dots, N$, let $x_i = a + (b - a)i/N$. Then $\mathcal{P} = \{x_0, x_1, \dots, x_N\}$ is a partition of $[a, b]$, with $|x_i - x_{i-1}| < \delta$. By continuity, f attains its maximum and minimum values on each closed subinterval $[x_{i-1}, x_i]$. Now let

$$f(c_i) = \inf\{f(x) : x \in [x_{i-1}, x_i]\}, \quad (4.5)$$

$$f(d_i) = \sup\{f(x) : x \in [x_{i-1}, x_i]\}. \quad (4.6)$$

Obviously $|d_i - c_i| < \delta$ and $f(d_i) \geq f(c_i)$. By uniform continuity

$$f(d_i) - f(c_i) < \frac{\epsilon}{(b - a)}.$$

So we have

$$\begin{aligned} U(f, \mathcal{P}) - L(f, \mathcal{P}) &= \sum_{i=1}^N f(d_i)(x_i - x_{i-1}) - \sum_{i=1}^N f(c_i)(x_i - x_{i-1}) \\ &= \sum_{i=1}^N (f(d_i) - f(c_i))(x_i - x_{i-1}) \\ &< \sum_{i=1}^N \frac{\epsilon}{b - a} (x_i - x_{i-1}) \\ &= \frac{\epsilon}{b - a} \sum_{i=1}^N (x_i - x_{i-1}) = \epsilon. \end{aligned}$$

Thus by Riemann's criterion, f is integrable on $[a, b]$. □

4.1. Calculating Integrals By Riemann Sums. It is possible to explicitly compute a surprisingly large class of integrals by evaluating Riemann sums. For monotone functions, the construction of upper and lower sums is straightforward. One simply picks sample points at the ends of each subinterval. We restrict our attention to $[0, 1]$. We can extend to the interval $[a, b]$ by a linear change of variable.

Example 4.1. We integrate $f(x) = x^2$ on $[0, 1]$ Since f is increasing we can take $\mathcal{P} = \{0, 1/n, 2/n, \dots, n/n\}$ and note that

$$\sum_{i=1}^n i^2 = \frac{1}{6}n(n+1)(2n+1). \quad (4.7)$$

Now we observe that

$$\begin{aligned} m_i(f, \mathcal{P}) &= \inf\{x^2 : x \in [\frac{i-1}{n}, \frac{i}{n}]\} \\ &= \frac{(i-1)^2}{n^2} \\ M_i(f, \mathcal{P}) &= \sup\{x^2 : x \in [\frac{i-1}{n}, \frac{i}{n}]\} \\ &= \frac{i^2}{n^2}. \end{aligned}$$

Then

$$\begin{aligned} L(f, \mathcal{P}) &= \sum_{i=1}^n \frac{(i-1)^2}{n^2} \left(\frac{i}{n} - \frac{(i-1)}{n} \right) \\ &= \frac{1}{n^3} \sum_{i=1}^n (i-1)^2. \end{aligned}$$

Also

$$\begin{aligned} U(f, \mathcal{P}) &= \sum_{i=1}^n \frac{i^2}{n^2} \left(\frac{i}{n} - \frac{(i-1)}{n} \right) \\ &= \frac{1}{n^3} \sum_{i=1}^n i^2. \end{aligned}$$

Using (4.7) we get

$$U(f, \mathcal{P}) - L(f, \mathcal{P}) = \frac{1}{6}n(n+1)(2n+1)\frac{1}{n^3} - \frac{(n-1)n(2n-1)}{6n^3} = \frac{1}{n}.$$

By Riemann's Criterion, f is Riemann integrable if for any $\epsilon > 0$ we can find a partition \mathcal{P} such that $U(f, \mathcal{P}) - L(f, \mathcal{P}) < \epsilon$. Clearly we can do this by taking $n > 1/\epsilon$. So f is Riemann integrable. Further

$$\begin{aligned} \int_0^1 f(x)dx &= \sup\{L(f, \mathcal{P}), \mathcal{P} \text{ a partition of } [0, 1]\} \\ &= \sup_{n \geq 1} \left\{ \frac{(n-1)n(2n-1)}{6n^3} \right\} \\ &= \sup_{n \geq 1} \left\{ \frac{1}{6n^2} - \frac{1}{2n} + \frac{1}{3} \right\} = \frac{1}{3}. \end{aligned}$$

Example 4.2. Let $a \neq 0$ and consider $f(x) = e^{ax}$ on $[0, 1]$. The function is monotone and we take the same partition as in the previous example. Then

$$m_k(f, \mathcal{P}) = \inf \left\{ e^{ax} : x \in \left[\frac{k-1}{n}, \frac{k}{n} \right) \right\} = e^{(k-1)a/n} \quad (4.8)$$

$$M_k(f, \mathcal{P}) = \sup \left\{ e^{ax} : x \in \left[\frac{k-1}{n}, \frac{k}{n} \right) \right\} = e^{ka/n} \quad (4.9)$$

Then

$$\begin{aligned} L(f, \mathcal{P}) &= \sum_{k=1}^n m_k(f, \mathcal{P})(x_k - x_{k-1}) \\ &= \frac{1}{n}(1 + e^{a/n} + \dots + e^{(n-1)a/n}) \end{aligned}$$

and

$$\begin{aligned} U(f, \mathcal{P}) &= \sum_{k=1}^n M_k(f, \mathcal{P})(x_k - x_{k-1}) \\ &= \frac{1}{n}(e^{a/n} + e^{2a/n} + \dots + e^{an/n}). \end{aligned}$$

So

$$U(f, \mathcal{P}) - L(f, \mathcal{P}) = \frac{e^a - 1}{n}.$$

This can be made smaller than ϵ by picking $n > \epsilon/(e^a - 1)$. Thus by Riemann's Criterion, f is Riemann integrable on $[0, 1]$. We can explicitly evaluate the upper and lower sums by noticing that they are sums of geometric progressions with common ratio e^a . Hence

$$\begin{aligned} L(f, \mathcal{P}) &= \frac{1}{n}(1 + e^{a/n} + \dots + e^{(n-1)a/n}) \\ &= \frac{1}{n} \frac{(1 - e^a)}{(1 - e^{a/n})}. \end{aligned}$$

So we have

$$\begin{aligned} \int_0^1 e^{ax} dx &= \sup_n \left\{ \frac{1}{n} \frac{(1 - e^a)}{(1 - e^{a/n})} \right\} \\ &= \lim_{u \rightarrow 0} \frac{u(1 - e^a)}{1 - e^{au}} \\ &= \frac{1}{a}(e^a - 1), \end{aligned}$$

where we put $u = 1/n$ and used L'Hôpital's rule to evaluate the limit.

We can actually prove that bounded monotone functions are Riemann integrable.

Theorem 4.4. *Suppose that $f : [0, 1] \rightarrow \mathbb{R}$ is monotone increasing and $f(1)$ is bounded. Then f is Riemann integrable on $[0, 1]$.*

Proof. With the previous partition of $[0, 1]$ we have

$$U(f, \mathcal{P}) - L(f, \mathcal{P}) = \frac{1}{n}(f(1) - f(0)). \quad (4.10)$$

Since f is monotone increasing, $f(0)$ must be finite and $f(1)$ is also finite, we can make this smaller than any $\epsilon > 0$ by suitable choice of n . So f is Riemann integrable. \square

4.2. The Fundamental Theorem of Calculus. It is possible to evaluate many integrals by means of Riemann sums- in particular, we can integrate any polynomial- but it is clearly a laborious procedure. Fortunately we have a far more powerful means of doing integration. The key is the following result, which is at the heart of modern science.

Theorem 4.5 (Fundamental Theorem of Calculus). *If f is a continuous function on $[a, b]$, then for all $x \in [a, b]$*

$$\frac{d}{dx} \int_a^x f(t) dt = f(x).$$

Proof. We define the function $F(x) = \int_a^x f(t) dt$. Since f is continuous, it is bounded. Thus there is an $M > 0$ such that $|f(t)| \leq M$ for all $t \in [a, b]$. Then

$$\begin{aligned} |F(x) - F(y)| &= \left| \int_y^x f(t) dt \right| \\ &\leq \int_y^x |f(t)| dt \\ &\leq M|x - y|. \end{aligned}$$

Consequently, F is Lipschitz continuous on $[a, b]$ and hence continuous. Now

$$\begin{aligned} \frac{F(x) - F(y)}{x - y} - f(y) &= \frac{1}{x - y} (F(x) - F(y) - (x - y)f(y)) \\ &= \frac{1}{x - y} \int_y^x (f(t) - f(y)) dt. \end{aligned}$$

By uniform continuity of f , given $\epsilon > 0$, we may find $\delta > 0$ such that $|x - y| < \delta$ implies $|f(x) - f(y)| < \epsilon$. We choose such an ϵ and δ to obtain

$$\begin{aligned} \left| \frac{F(x) - F(y)}{x - y} - f(y) \right| &\leq \frac{1}{|x - y|} \int_y^x |f(t) - f(y)| dt \\ &< \frac{1}{|x - y|} \epsilon(x - y) = \epsilon \end{aligned}$$

as $x > y$. Thus F is differentiable and $F' = f$. \square

In other words, integration is essentially the inverse of differentiation. From this we can establish the well known second form of the fundamental theorem.

Corollary 4.6 (The Fundamental Theorem of Calculus II). *Let f be a Riemann integrable function on $[a, b]$. Then if $F' = f$ on (a, b) the integral is given by*

$$\int_a^b f(x)dx = F(b) - F(a). \quad (4.11)$$

Proof. Suppose that $G(x) = \int_a^x f(t)dt$ and $F'(x) = f(x)$. It follows that $G - F$ is a constant, since $G' = f$. Hence $G(b) - F(b) = G(a) - F(a)$. But $G(a) = 0$. Hence $G(b) = \int_a^b f(x)dx = F(b) - F(a)$. \square

There is a mean value theorem for the Riemann integral which is often useful.

Theorem 4.7 (Mean Value Theorem for Integrals). *Suppose that f and g are continuous on $[a, b]$ and $g(x) \geq 0$, for all $x \in [a, b]$. Then there exists $c \in [a, b]$ such that*

$$\int_a^b f(x)g(x)dx = f(c) \int_a^b g(x)dx. \quad (4.12)$$

Proof. By continuity f is bounded. Suppose that for all $t \in [a, b]$ $m \leq f(t) \leq M$. Then

$$m \int_a^b g(x)dx \leq \int_a^b f(x)g(x)dx \leq M \int_a^b g(x)dx.$$

Let $F(t) = \int_a^t g(t)dt$. By the Intermediate Value Theorem, there is a $c \in [a, b]$ such that

$$F(c) = f(c) \int_a^b g(t)dt = \int_a^b f(x)g(x)dx.$$

\square

Notice that if $g = 1$ and $F' = f$ then we have the existence of a $c \in [a, b]$ such that

$$\int_a^b f(x)dx = F(b) - F(a) = F'(c)(b - a), \quad (4.13)$$

which is the mean value theorem. Actually the mean value theorem can be used to prove the fundamental theorem of calculus. This is an exercise.

4.3. Integration Rules. Integration is intrinsically more difficult than differentiation. Useful rules for evaluating integrals exist however. Integration by parts is simply the product rule of differentiation backwards. Specifically

$$(fg)'(x) = f'(x)g(x) + f(x)g'(x).$$

Integrating both sides gives the integration by parts rule

$$\int_a^b f(x)g'(x)dx = [f(x)g(x)]_a^b - \int_a^b f'(x)g(x)dx. \quad (4.14)$$

The most important technique for evaluating integrals is the use of substitutions. This is the chain rule in reverse. The chain rule says that $(f \circ g)'(x) = f'(g(x))g'(x)$. Thus letting $u = g(x)$ gives

$$\int_a^b f(g(x))g'(x)dx = \int_{g(a)}^{g(b)} f(u)du. \quad (4.15)$$

We can use integration by parts to show how Taylor's Theorem follows from the Fundamental Theorem of Calculus. Assume that f is continuously differentiable $n + 1$ times. We know that

$$f(x) - f(a) = \int_a^x f'(t)dt. \quad (4.16)$$

We are going to integrate by parts. Notice however that instead of using the obvious anti-derivative of 1, we are going to use $t - x$, which is also an anti-derivative of 1. So that

$$\begin{aligned} f(x) - f(a) &= [(t - x)f'(t)]_a^x - \int_a^x (t - x)f'(t)dt \\ &= (x - a)f'(a) + \int_a^x (x - t)f'(t)dt \\ &= (x - a)f'(a) + \frac{(x - a)^2}{2}f''(x) + \frac{1}{2} \int_a^x (x - t)^2 f''(t)dt. \end{aligned}$$

Repeating this n times gives

$$\begin{aligned} f(x) &= f(a) + (x - a)f'(a) + \frac{(x - a)^2}{2}f''(x) + \dots \\ &\quad + \frac{1}{n!}(x - a)^n f^{(n)}(a) + \frac{1}{n!} \int_a^x (x - t)^n f^{(n)}(t)dt. \end{aligned}$$

This gives us the useful form for the remainder in the Taylor series expansion

$$R_n(a, x) = \frac{1}{n!} \int_a^x (x - t)^n f^{(n)}(t)dt.$$

Using the mean value theorem for integrals we can show that this is the same as the derivative form we found earlier.

4.4. Improper Riemann Integrals. It is often the case that we wish to consider an integral of a function over a set where the function is discontinuous.

Definition 4.8. Let $f : [a, b] \rightarrow \mathbb{R}$ be continuous on $(a, b]$, but f is discontinuous at a . Then the improper Riemann integral of f over $[a, b]$ is define by

$$\int_a^b f(x)dx = \lim_{X \rightarrow a} \int_X^b f(x)dx, \quad (4.17)$$

provided the limit exists. Similarly, if the discontinuity is at $x = b$ then

$$\int_a^b f(x)dx = \lim_{X \rightarrow b} \int_a^X f(x)dx, \quad (4.18)$$

provided the limit exists.

Example 4.3. Consider $f(x) = 1/\sqrt{x}$ on $[0, 1]$. Then f is continuous on $(0, 1]$ with a discontinuity at 0. Thus the improper Riemann integral of f over $[0, 1]$ is

$$\begin{aligned} \int_0^1 f(x)dx &= \lim_{X \rightarrow 0} \int_X^1 \frac{dx}{\sqrt{x}} \\ &= \lim_{X \rightarrow 0} 2\sqrt{x} \Big|_X^1 \\ &= \lim_{X \rightarrow 0} (2\sqrt{1} - \sqrt{X}) = 2. \end{aligned}$$

For integrals on unbounded domains we can use the same idea.

Definition 4.9. The improper Riemann integral of f over \mathbb{R} is defined by

$$\int_{-\infty}^{\infty} f(x)dx = \lim_{R \rightarrow \infty} \int_{-R}^0 f(x)dx + \lim_{T \rightarrow \infty} \int_0^T f(x)dx, \quad (4.19)$$

provided the limits exist.

One has to be careful to distinguish between Definition 4.9 and the Cauchy Principal value.

Definition 4.10. The quantity

$$\text{pv} \int_{-\infty}^{\infty} f(x)dx = \lim_{R \rightarrow \infty} \int_{-R}^R f(x)dx,$$

is known as the Cauchy Principal value of the integral, provided that the limit exists.

Example 4.4. The improper Riemann integral $\int_{-\infty}^{\infty} x dx$ does not exist, but

$$\begin{aligned} \text{pv} \int_{-\infty}^{\infty} x dx &= \lim_{R \rightarrow \infty} \left[\frac{x^2}{2} \right]_{-R}^R \\ &= \frac{1}{2}(R^2 - R^2) = 0. \end{aligned}$$

5. SEQUENCES OF FUNCTIONS

5.1. Pointwise and Uniform Convergence. The Riemann integral is a powerful tool, but it has limitations. The most important relates to the question of swapping integrals and limits. To see what this involves, let us introduce the notion of convergence of a sequence of functions.

Definition 5.1. We say that a sequence of functions $\{f_n\}_{n=1}^{\infty}$ converges pointwise to a function f on a set $X \subseteq \mathbb{R}$ if $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ for all $x \in X$.

Example 5.1. Let $f_n(x) = \frac{n^2 x^2}{n^2 + 1}$ and $f(x) = x^2$. Then $f_n(x) \rightarrow x^2$ for all $x \in \mathbb{R}$. To see this note that

$$f_n(x) = \frac{(n^2 + 1 - 1)x^2}{n^2 + 1} = x^2 - \frac{x^2}{n^2 + 1}.$$

For each $x \in \mathbb{R}$ we have

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{x^2}{n^2 + 1} &= x^2 \lim_{n \rightarrow \infty} \frac{1}{n^2 + 1} \\ &= 0. \end{aligned}$$

Thus for each x , $\lim_{n \rightarrow \infty} f_n(x) = x^2$. So $f_n \rightarrow f$ pointwise.

Given a sequence of pointwise convergent functions, we would like to be able to conclude that

$$\lim_{n \rightarrow \infty} \int_a^b f_n(x) dx = \int_a^b f(x) dx.$$

A common example is where we have a function defined as an infinite sum

$$f = \sum_{n=1}^{\infty} f_n,$$

and we would like to determine $\int f$ by term by term integration. So that we would like

$$\int_a^b f(x) dx = \sum_{n=1}^{\infty} \int_a^b f_n(x) dx.$$

Unfortunately this is not in general true. Finding counterexamples is not very hard. Consider the sequence

$$f_n(x) = nxe^{-nx^2}. \tag{5.1}$$

Then on $[0, 1]$, $f_n \rightarrow 0$ pointwise as $n \rightarrow \infty$. To show convergence we have use L'Hôpital's rule.

$$\begin{aligned}\lim_{n \rightarrow \infty} \frac{nx}{e^{nx^2}} &= \lim_{n \rightarrow \infty} \frac{x}{x^2 e^{nx^2}} \\ &= \lim_{n \rightarrow \infty} \frac{1}{x e^{nx^2}} = 0,\end{aligned}$$

for each $x \neq 0$. We differentiated with respect to n here. Hence

$$\int_0^1 \lim_{n \rightarrow \infty} f_n(x) dx = 0.$$

However

$$\int_0^1 n x e^{-nx^2} dx = \frac{1 - e^{-n}}{2}.$$

Hence

$$\lim_{n \rightarrow \infty} \int_0^1 f_n(x) dx = \frac{1}{2} \neq \int_0^1 \lim_{n \rightarrow \infty} f_n(x) dx.$$

So we cannot just take the limit inside the integral. Even for this simple example it does not work. However there are cases where this is possible. We will see that in order to safely swap limits and Riemann integrals, we need what is known as *uniform convergence*.

Definition 5.2. A sequence of functions $\{f_n\}_{n=1}^\infty$ on a set $X \subseteq \mathbb{R}$ converges uniformly to f on X if for any $\epsilon > 0$ we can find $N \in \mathbb{N}$ such that $n \geq N$ implies $|f_n(x) - f(x)| < \epsilon$ for all $x \in X$.

Example 5.2. The sequence $f_n(x) = \frac{n^2 x^2}{n^2 + 1}$ converges uniformly to $f(x) = x^2$ on the interval $[-1, 1]$. To see this we write

$$\begin{aligned}|f_n(x) - f(x)| &= \left| x^2 - \frac{x^2}{n^2 + 1} - x^2 \right| \\ &= |x^2| \left| \frac{1}{n^2 + 1} \right| \\ &\leq \frac{1}{n^2 + 1},\end{aligned}$$

since $|x| \leq 1$ on $[-1, 1]$. Now if $n \geq N > \sqrt{\frac{1}{\epsilon} - 1}$, $\epsilon < 1$, then for every $x \in [-1, 1]$ $|f_n(x) - f(x)| < \epsilon$. So $f_n \rightarrow f$ uniformly. In fact $f_n \rightarrow f$ uniformly on any bounded interval $[a, b]$.

Example 5.3. Consider the sequence of functions $f_n(x) = \sqrt{x^2 + 1/n^2}$. For all x , $f_n(x) \rightarrow |x|$ uniformly. To see this, observe that

$$\begin{aligned} \sqrt{x^2 + \frac{1}{n^2}} - |x| &= \left(\sqrt{x^2 + \frac{1}{n^2}} - |x| \right) \frac{\sqrt{x^2 + \frac{1}{n^2}} + |x|}{\sqrt{x^2 + \frac{1}{n^2}} + |x|} \\ &= \frac{1}{n^2 \left(\sqrt{x^2 + \frac{1}{n^2}} + |x| \right)} \\ &\leq \frac{1}{n}. \end{aligned}$$

Thus for any $\epsilon > 0$, if $n \geq N > \frac{1}{\epsilon}$, we have $|f_n(x) - |x|| < \epsilon$ for all x .

The key point in these examples is that the same value of N works for every value of x in the specified domain. That is what the word *uniform* is referring to in the definition. For pointwise convergence, we may need a different value of N for different values of x .

The first result is a trivial exercise.

Lemma 5.3. *If $f_n \rightarrow f$ uniformly on X , then $f_n \rightarrow f$ pointwise.*

The converse of this result is false. Pointwise convergent sequences usually do not converge uniformly. The example (5.1) converges pointwise but not uniformly. There is a result due to Egoroff that tells us that a sequence of functions converging pointwise on a closed and bounded interval $[a, b]$, converges uniformly on $[a, b] - E$, where E is a small set. However this is not enough to swap limits and Riemann integrals. We do not discuss Egoroff's Theorem in this subject. See however Lebesgue Integration and Fourier Analysis, which is a higher level subject.

Uniformly convergent sequences have nice properties. An important one is that they preserve continuity. Pointwise convergence does not. This is really the heart of the problem.

Theorem 5.4. *If $\{f_n\}_{n=1}^\infty$ is a uniformly convergent sequence of continuous functions on $X \subseteq \mathbb{R}$, with $f_n \rightarrow f$ then f is continuous on X .*

Proof. Since f_k is continuous at $x \in X$, given $\epsilon > 0$, we may choose $\delta > 0$ such that for all y satisfying $0 < |x - y| < \delta$ we have

$$|f_k(x) - f_k(y)| < \epsilon/3.$$

By uniform convergence, we may choose $N \in \mathbb{N}$ such that $k \geq N$ implies

$$|f(x) - f_k(x)| < \epsilon/3,$$

for all $x \in X$. Consequently, given $x \in X$, then for all $y \in X$ satisfying $0 < |x - y| < \delta$ we have

$$\begin{aligned} |f(x) - f(y)| &= |f(x) - f_k(x) + f_k(x) - f_k(y) + f_k(y) - f(y)| \\ &\leq |f(x) - f_k(x)| + |f_k(x) - f_k(y)| + |f_k(y) - f(y)| \\ &< \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon. \end{aligned}$$

Thus f is continuous at x . \square

This result is not true for pointwise convergence. For example, the pointwise convergent double sequence $f_{n,j}(x) = (\cos(n!\pi x))^{2j}$ does not have a continuous limit on $[0, 1]$. In fact it converges to the Dirichlet function

$$\mathbb{D}(x) = \begin{cases} 1 & x \in \mathbb{Q} \cap [0, 1] \\ 0 & x \notin \mathbb{Q} \cap [0, 1]. \end{cases} \quad (5.2)$$

To see this, observe that if x is rational, then $x = p/q$ for some integers p, q . Then for $k > q$ it follows that $\pi k!x = N\pi$ for some integer N . Now $\cos^{2j}(N\pi) = 1$ for all j . Thus if x is rational, $\lim_{j,k \rightarrow \infty} g_{k,j}(x) = 1$. Suppose however that x is irrational. Then $\pi k!x$ is never an integer multiple of π and so $-1 < \cos(\pi k!x) < 1$. Now if $|r| < 1$, $r^{2j} \rightarrow 0$ as $j \rightarrow \infty$. So for x irrational, $\lim_{j,k \rightarrow \infty} g_{k,j}(x) = 0$. Hence the limit of this sequence of functions is a function that is 1 if x is rational and 0 if x is irrational. This function is not Riemann integrable and it is not even continuous, despite the fact that every function in the sequence is not only continuous, but is analytic.

Uniform convergence is preserved under addition.

Proposition 5.5. *If $f_n \rightarrow f$ and $g_n \rightarrow g$ uniformly, then $af_n + bg_n \rightarrow af + bg$ uniformly, where a, b are constants.*

Proof. Suppose that $a, b \neq 0$. $f_n, g_n : X \subseteq \mathbb{R} \rightarrow \mathbb{R}$. Then given $\epsilon > 0$ there exists $N_1 \in \mathbb{N}$ such that

$$\sup_{x \in X} |f_n(x) - f(x)| < \epsilon/(2|a|)$$

and there exists $N_2 \in \mathbb{N}$ such that

$$\sup_{x \in X} |g_n(x) - g(x)| < \epsilon/(2|b|).$$

Let $N = \max N_1, N_2$. Then for $n \geq N$

$$\begin{aligned} \sup_{x \in X} |af_n(x) + bg_n(x) - af(x) - bg(x)| &\leq |a| \sup_{x \in X} |f_n(x) - f(x)| \\ &\quad + |b| \sup_{x \in X} |g_n(x) - g(x)| \\ &< \epsilon/2 + \epsilon/2 = \epsilon. \end{aligned}$$

\square

However, uniform convergence is *not* preserved under pointwise multiplication. That is, if $f_n \rightarrow f$ uniformly on $X \subseteq \mathbb{R}$ and $g_n \rightarrow g$ uniformly on X , it is not in general true that $f_n g_n \rightarrow fg$ uniformly on X . The best we can say is the following.

Theorem 5.6. *Suppose that $f_n \rightarrow f$ uniformly on the closed and bounded interval $[a, b]$ and $g_n \rightarrow g$ uniformly on $[a, b]$. Then $f_n g_n \rightarrow fg$ uniformly on $[a, b]$.*

Proof. Since $f_n \rightarrow f$ uniformly, it converges pointwise on $[a, b]$ and hence each sequence $\{f_n(x)\}_{n=1}^\infty$ is bounded, for all $x \in [a, b]$. Consequently f is also bounded. Similarly for $\{g_n\}_{n=1}^\infty$. Let

$$A = \sup_{x \in [a, b]} |f(x)|, \quad B = \sup_{x \in [a, b], n \geq 1} |g_n(x)|.$$

Choose an $\epsilon > 0$. We can find $N_1 \in \mathbb{N}$ such that $n \geq N_1$ implies $|f_n(x) - f(x)| < \epsilon/(2B)$ and an $N_2 \in \mathbb{N}$ such that $n \geq N_2$ implies that $|g_n(x) - g(x)| < \epsilon/(2A)$. Then take $N = \max\{N_1, N_2\}$ and for $n \geq N$

$$\begin{aligned} \sup_{x \in [a, b]} |f_n(x)g_n(x) - f(x)g(x)| &= \sup_{x \in [a, b]} |f_n(x)g_n(x) - f(x)g_n(x) \\ &\quad + f(x)g_n(x) - f(x)g(x)| \\ &= \sup_{x \in [a, b]} |g_n(x)||f_n(x) - f(x)| \\ &\quad + \sup_{x \in [a, b]} |f(x)||g_n(x) - g(x)| \\ &< \epsilon/(2A) + \epsilon/(2B) = \epsilon. \end{aligned}$$

□

5.2. The Weierstrass M-test. There are various tests for uniform convergence. For series we have the following powerful result.

Theorem 5.7 (Weierstrass M-Test). *Let $\{f_n\}_{n=1}^\infty$ be a sequence of functions on X such that $|f_n(x)| \leq M_n$ all $x \in X$ and $\sum_{n=1}^\infty M_n < \infty$. Then the series $\sum_{n=1}^\infty f_n(x)$ is uniformly convergent.*

Proof. Let $S_N(x) = \sum_{n=1}^N f_n(x)$ and suppose that $|f_n(x)| \leq M_n$. Then for all $N \geq M$

$$\begin{aligned} |S_N(x) - S_M(x)| &= \left| \sum_{n=M+1}^N f_n(x) \right| \\ &\leq \sum_{n=M+1}^N |f_n(x)| \\ &\leq \sum_{n=M+1}^N M_n \rightarrow 0, \end{aligned}$$

as $N, M \rightarrow \infty$. So the series S_N converges independently of x and hence is uniformly convergent. □

Example 5.4. The M test is generally easy to use. To illustrate, consider the series

$$f(x) = \sum_{n=1}^{\infty} \frac{\cos(nx)}{n^2 + 1}. \quad (5.3)$$

Letting $f_n(x) = \frac{\cos(nx)}{n^2 + 1}$, we immediately see that

$$|f_n(x)| \leq \frac{1}{n^2 + 1}, \quad (5.4)$$

and by the comparison test $\sum_{n=1}^{\infty} \frac{1}{n^2 + 1} < \infty$. Hence the series (5.3) is uniformly convergent and so f is a continuous function.

As an application we prove a result about power series. We mentioned this previously, but deferred the proof.

Theorem 5.8. *Let $\sum_{n=0}^{\infty} a_n x^n$ be a power series with radius of convergence R . Let $0 < r < R$. Then the series converges uniformly on $[-r, r]$.*

Proof. Let $\sum_{n=0}^{\infty} a_n x^n$ be convergent for $|x| < R$. Then it is absolutely convergent. Pick $x = x_0 > r$ and $x_0 < R$ and we have $\sum_{n=0}^{\infty} a_n x_0^n$ is convergent, hence $a_n x_0^n \rightarrow 0$. So there is an $M > 0$ such that $|a_n x_0^n| \leq M$. Then

$$\begin{aligned} |a_n x^n| &\leq |a_n| r^n \\ &= |a_n x_0^n| \left| \frac{r}{x_0} \right|^n \\ &\leq M \left| \frac{r}{x_0} \right|^n. \end{aligned}$$

Now $\sum_{n=0}^{\infty} M \left| \frac{r}{x_0} \right|^n$ converges and hence the power series converges uniformly by the Weierstrass M test. \square

It is important to note that this theorem does not say that a power series converges uniformly on $(-R, R)$. Indeed the series $\sum_{n=0}^{\infty} x^n$ converges on $(-1, 1)$ but the convergence is not uniform. It does converge uniformly on $[-r, r]$ for $r < 1$. The point is we cannot necessarily extend the uniform convergence to the entire interval of convergence.

Another useful test is due to Dini.

Theorem 5.9 (Dini). *Let $\{f_n\}_{n=1}^{\infty}$ be a sequence of continuous functions on $[a, b]$ which converges monotonically to a continuous function f on $[a, b]$. Then $f_n \rightarrow f$ uniformly on $[a, b]$.*

Proof. We can with no loss of generality suppose that $f = 0$ and that $f_n(x)$ decreases monotonically to 0 for all $x \in [a, b]$. If this is not the case then we can consider the functions $g_n = \pm(f_n - f)$ depending

on whether f_n increases or decreases. The sequence $\{g_n\}$ will then decrease monotonically to 0.

Now set

$$M_n = \sup\{f_n(x) : x \in [x, b]\}.$$

Since f_n decreases to 0, it follows that M_n is decreasing. We claim that $M_n \rightarrow 0$. This will be sufficient to establish that the convergence is uniform, since then given $\epsilon > 0$ we will be able to find N such that for all $n \geq N$ $M_n < \epsilon$ and so for $n \geq N$ we will have

$$\sup_{x \in [a, b]} |f_n(x) - f(x)| < \epsilon. \quad (5.5)$$

We proceed by contraction. Suppose otherwise. Then we can find $\delta > 0$ such that for every n , $M_n > \delta$. Consequently for every n there is a point x_n such that $f(x_n) > \delta$. The sequence $\{x_n\} \in [a, b]$ is bounded, so by the Bolzano-Weierstrass Theorem it contains a convergent subsequence $\{x_{n_k}\}_{k=1}^\infty$. Let $x_{n_k} \rightarrow \alpha$ as $k \rightarrow \infty$. By assumption, $f_n(\alpha) \rightarrow 0$ as $n \rightarrow \infty$. So there exists $N > 0$ such that for $n \geq N$, we have $|f_n(\alpha)| < \delta$. But each f_n is continuous, so that we can find $\epsilon > 0$ such that $|x - \alpha| < \epsilon$ implies $|f_n(x)| < \delta$ for all $n \geq N$. But this is a contradiction, because we can choose N such that for all $n_k \geq N$, $|x_{n_k} - \alpha| < \epsilon$ and $f_{n_k}(x_{n_k}) > \delta$. Thus our assumption must be false and hence $M_n \rightarrow 0$. \square

5.3. Swapping Limits and Integrals. Uniform convergence allows us to reverse the order of a limit and a Riemann integral. This is a tremendously useful property of uniformly convergent sequences. Swapping limits and integrals is a problem that arises frequently.

Theorem 5.10. *If $\{f_n\}_{n=1}^\infty$ is a sequence of Riemann integrable functions converging uniformly to f on $[a, b]$, then f is Riemann integrable on $[a, b]$ and*

$$\lim_{n \rightarrow \infty} \int_a^b f_n(x) dx = \int_a^b f(x) dx.$$

Proof. If the functions f_n are continuous then the proof is easy. By uniform convergence we can choose N such that $n \geq N$ implies

$$|f_n(x) - f(x)| < \epsilon/(b - a).$$

As f is continuous by Theorem 5.4, we have for $n \geq N$

$$\left| \int_a^b f_n(x) dx - \int_a^b f(x) dx \right| \leq \int_a^b |f_n(x) - f(x)| dx \quad (5.6)$$

$$< \int_a^b \frac{\epsilon}{(b - a)} dx = \epsilon. \quad (5.7)$$

If the functions $\{f_n\}$ are not assumed to be continuous, then we have to prove the limit is integrable. Each f_n is bounded, so the limit f is bounded. Pick $\epsilon > 0$ and by uniform convergence we can choose $N \in \mathbb{N}$

such that for all $x \in [a, b]$ $n \geq N$ implies $|f_n(x) - f(x)| < \frac{\epsilon}{3(b-a)}$. Since f_N is integrable, by Riemann's criterion we can choose a partition \mathcal{P} of $[a, b]$ such that

$$U(f_N, \mathcal{P}) - L(f_N, \mathcal{P}) < \frac{\epsilon}{3}.$$

Now $\sup_{x \in [a, b]} |f_N(x) - f(x)| < \frac{\epsilon}{3(b-a)}$, so we have

$$\begin{aligned} U(f, \mathcal{P}) - L(f, \mathcal{P}) &= U(f + f - f_N, \mathcal{P}) - L(f + f_N - f, \mathcal{P}) \\ &\leq U(f_N, \mathcal{P}) + U(f - f_N, \mathcal{P}) - L(f_N, \mathcal{P}) \\ &\quad - L(f - f_N, \mathcal{P}) \\ &= U(f_N, \mathcal{P}) - L(f_N, \mathcal{P}) + U(f - f_N, \mathcal{P}) \\ &\quad - L(f - f_N, \mathcal{P}) \\ &< \frac{\epsilon}{3} + 2 \frac{\epsilon}{3(b-a)}(b-a) = \epsilon. \end{aligned}$$

The inequalities used above can be verified by direct calculation. So f is Riemann integrable. The rest of the proof is as in the continuous case. \square

This swapping of limits does not work for pointwise convergence with the Riemann integral. The limit function may not even be Riemann integrable, as is the case with the double sequence $f_{n,k}$ above. Unfortunately when we have sequences of functions, they often do not converge uniformly. This leads to the question of how can we modify the integral in such a way as to be able to swap limits and integrals, even when we do not have uniform convergence? This problem led to a new theory of integration due to Henri Lebesgue.

Uniform convergence is equivalent to a sequence being uniformly Cauchy.

Definition 5.11. A sequence $f_n : X \subseteq \mathbb{R} \rightarrow \mathbb{R}$ is uniformly Cauchy if given $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $n, m \geq N$

$$\sup_{x \in I} |f_n(x) - f_m(x)| < \epsilon.$$

The next result connects uniformly Cauchy and uniformly convergent sequences.

Theorem 5.12. *Every uniformly convergent sequence of functions is uniformly Cauchy.*

Proof. Let $\{f_n\}_{n=1}^\infty$ be a uniformly convergent sequence of functions defined on $X \subseteq \mathbb{R}$ and suppose it has limit f . Let $\epsilon > 0$. Choose $N \in \mathbb{N}$ such that $x \in X, n \geq N$ implies

$$|f_n(x) - f(x)| < \frac{\epsilon}{2}.$$

Then if $x \in X, m, n \geq N$ implies

$$\begin{aligned} |f_m(x) - f_n(x)| &= |f_m(x) - f(x) + f(x) - f_n(x)| \\ &\leq |f_m(x) - f(x)| + |f(x) - f_n(x)| \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \end{aligned}$$

So $\{f_n\}_{n=1}^\infty$ is uniformly Cauchy. \square

The converse of this result is also true, but the proof is a little harder.

Theorem 5.13. *Every uniformly Cauchy sequence of functions is uniformly convergent.*

Proof. There are two parts. First we have to define the limit and then we have to prove that the convergence is uniform. The first part proceeds as follows.

Let $\{f_n\}_{n=1}^\infty$ be a uniformly Cauchy sequence on $X \subseteq \mathbb{R}$. Let $x_0 \in X$. Then $\{f_n(x_0)\}_{n=1}^\infty$ is a Cauchy sequence in \mathbb{R} and hence it is convergent. Let us denote the limit by $f(x_0)$. This process defines a limit function for each $x \in X$. By construction, $f_n \rightarrow f$ pointwise.

Now we prove that $f_n \rightarrow f$ uniformly. Let $\epsilon > 0$ and choose $N \in \mathbb{N}$ such that for all $x \in X, n, m \geq N$ we have

$$|f_m(x) - f_n(x)| < \frac{\epsilon}{2}.$$

Now $f_m \rightarrow f$ pointwise. So for all $x \in X, n \geq N$

$$|f(x) - f_n(x)| = \lim_{m \rightarrow \infty} |f_m(x) - f(x)| < \frac{\epsilon}{2}.$$

Thus $f_n \rightarrow f$ uniformly. \square

5.4. Swapping Limits and Derivatives. Interchanging limits and derivatives is actually harder than interchanging a limit and an integral. Thus we can bound the difference between the n th term of the sequence and the limit independently of x and so the convergence is uniform. So uniform convergence is not enough to guarantee that the limit function is differentiable.

Even if the limit is differentiable, it does not follow that $f'_n \rightarrow f'$. Consider again the sequence from Example 5.3. There $f_n(x) \rightarrow f(x) = |x|$ uniformly for all x , but although $f'_n(0)$ exists for all n , $f'(0)$ does not exist. Here is another example.

Example 5.5. Let $f_n(x) = \frac{x}{1 + nx^2}$. Now $f_n \rightarrow 0$ for all x as $n \rightarrow \infty$.

But

$$f'_n(x) = \frac{(1 - nx^2)}{(1 + nx^2)^2}$$

and so $f'_n(0) \rightarrow 1 \neq 0 = f'(0)$.

We actually require uniform convergence of the derivatives in order to swap differentiation and limits. The relevant result follows.

Theorem 5.14. *Let I be an open interval in \mathbb{R} , $f : I \rightarrow \mathbb{R}$ and let $\{f_n\}_{n=1}^{\infty}$ be a sequence of differentiable functions on I which converges pointwise to f on I . Let $g : I \rightarrow \mathbb{R}$ and let the sequence of derivatives $\{f'_n\}_{n=1}^{\infty}$ converge uniformly to g on I . Then f is differentiable on I and $f'(x) = g(x)$ for all $x \in I$.*

Proof. Let $\epsilon > 0$ and pick an $N_1 \in \mathbb{N}$ such that

$$\sup_{x \in I} |f_n(x) - f(x)| < \frac{\epsilon}{3}.$$

The sequence $\{f'_n\}_{n=1}^{\infty}$ is uniformly Cauchy on I . So we can find $N_2 \in \mathbb{N}$ such that for all $x \in I$, $n, m \geq N_2$

$$|f'_m(x) - f'_n(x)| < \frac{\epsilon}{3}.$$

Now let $N = \max N_1, N_2$. The function f_N is differentiable on I and so at any point $x_0 \in I$ there exists $\delta > 0$ such that for $x \in I$, $0 < |x - x_0| < \delta$ we have

$$\left| \frac{f_N(x) - f_N(x_0)}{x - x_0} - f'_N(x_0) \right| < \frac{\epsilon}{3}.$$

Now let $x \in I$, $x \neq x_0$ and choose $M \geq N$. $f_M - f_N$ is differentiable and so by the Mean Value Theorem we can find c between x and x_0 such that

$$\frac{(f_M - f_N)(x) - (f_M - f_N)(x_0)}{x - x_0} = (f_M - f_N)'(c).$$

which is the same as

$$(f_M - f_N)(x) - (f_M - f_N)(x_0) = (f_M - f_N)'(c)(x - x_0).$$

From this we deduce that

$$\begin{aligned} |f_M(x) - f_M(x_0) - (f_N(x) - f_N(x_0))| &= |f'_M(c) - f'_N(c)||x - x_0| \\ &< \frac{\epsilon}{3}|x - x_0|, \end{aligned} \tag{5.8}$$

where $|f'_M(c) - f'_N(c)| < \epsilon/3$ by the fact that the sequence is uniformly Cauchy and $M, N \geq N$ and $c \in I$. Taking limits as $M \rightarrow \infty$ in (5.8) we get

$$|f(x) - f(x_0) - (f_N(x) - f_N(x_0))| \leq \frac{\epsilon}{3}|x - x_0|,$$

which leads to the inequality

$$\left| \frac{f(x) - f(x_0)}{x - x_0} - \frac{f_N(x) - f_N(x_0)}{x - x_0} \right| \leq \frac{\epsilon}{3}.$$

Finally, we put this altogether and let $x \in I$, $0 < |x - x_0| < \delta$, then adding and subtracting appropriate terms, and using the triangle inequality, we can write

$$\begin{aligned} \left| \frac{f(x) - f(x_0)}{x - x_0} - g(x_0) \right| &\leq \left| \frac{f(x) - f(x_0)}{x - x_0} - \left(\frac{f_N(x) - f_N(x_0)}{x - x_0} \right) \right| \\ &\quad + \left| \frac{f_N(x) - f_N(x_0)}{x - x_0} - f'_N(x_0) \right| \\ &\quad + |f'_N(x_0) - g(x_0)| \\ &< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon. \end{aligned}$$

Hence f is differentiable at x_0 and $f'(x_0) = g(x_0)$. □

5.5. The Weierstrass Approximation Theorem. One of the most important results on uniform approximation of functions is due to Weierstrass. This says that any continuous function on a closed and bounded interval $[a, b]$ can be approximated uniformly by a polynomial. Equivalently, there is a sequence of polynomials converging uniformly to f . This can be proved in a number of ways. We can use Fourier series to establish the result, but Bernstein actually constructed a sequence of polynomials which approximates any continuous function uniformly. To present Bernstein's proof, we require a preliminary lemma.

Lemma 5.15. *For each fixed x , the following identities hold.*

$$\sum_{k=0}^n \binom{n}{k} x^k (1-x)^{n-k} = 1 \tag{5.9}$$

$$\sum_{k=1}^n k \binom{n}{k} x^k (1-x)^{n-k} = nx \tag{5.10}$$

$$\sum_{k=1}^n k^2 \binom{n}{k} x^k (1-x)^{n-k} = n(n-1)x^2 + nx. \tag{5.11}$$

Proof. For the first identity, observe that $1^n = (x + (1-x))^n$ and apply the Binomial Theorem to both sides. For the second,

$$\begin{aligned} k \binom{n}{k} &= \frac{kn!}{k!(n-k)!} = \frac{n(n-1)!}{(k-1)!(n-1-(k-1))!} \\ &= n \binom{n-1}{k-1}. \end{aligned}$$

So that

$$\begin{aligned} \sum_{k=1}^n k \binom{n}{k} x^k (1-x)^{n-k} &= \sum_{k=1}^n n \binom{n-1}{k-1} x^k (1-x)^{n-k} \\ &= n \sum_{j=0}^{n-1} \binom{n-1}{j} x^{j+1} (1-x)^{n-1-j} \\ &= nx(x + (1-x))^{n-1} = nx. \end{aligned}$$

For the final identity, notice that $k^2 = k(k-1) + k$, so that

$$\begin{aligned} \sum_{k=1}^n k^2 \binom{n}{k} x^k (1-x)^{n-k} &= \sum_{k=1}^n (k(k-1) + k) \binom{n}{k} x^k (1-x)^{n-k} \\ &= nx + \sum_{k=1}^n k(k-1) \binom{n}{k} x^k (1-x)^{n-k}. \end{aligned}$$

Now

$$k(k-1) \binom{n}{k} = n(n-1) \binom{n-2}{k-2},$$

so that

$$\begin{aligned} \sum_{k=1}^n k(k-1) \binom{n}{k} x^k (1-x)^{n-k} &= \sum_{k=1}^n n(n-1) \binom{n-2}{k-2} x^k (1-x)^{n-k} \\ &= n(n-1)x^2(x + 1-x)^{n-2} \\ &= n(n-1)x^2. \end{aligned}$$

This completes the proof. □

Now we come to the final result in the subject. This is Weierstrass' Theorem.

Theorem 5.16 (Weierstrass Approximation Theorem). *Let f be a continuous function on a closed and bounded interval $[a, b]$. Then given any $\epsilon > 0$ there is a polynomial P with the property that*

$$\sup_{x \in [a, b]} |f(x) - P(x)| < \epsilon.$$

Proof. This is Bernstein's proof. For simplicity we can restrict attention to the interval $[0, 1]$, since $[0, 1]$ can be mapped to $[a, b]$ by the function $\varphi(t) = a(1-t) + bt$, where $0 \leq t \leq 1$. It is not hard to show that if $P_n \rightarrow g$ uniformly on $[0, 1]$, then

$$P_n(\varphi(t)) \rightarrow g\left(\frac{t-a}{b-a}\right),$$

uniformly on $[a, b]$.

Now let f be continuous on $[0, 1]$ and define the polynomials

$$P_n(x) = \sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k}. \quad (5.12)$$

These are the Bernstein polynomials for f . We claim that $P_n \rightarrow f$ uniformly on $[0, 1]$. That is, given $\epsilon > 0$ we want to find an $n \in \mathbb{N}$ such that $n \geq N$ implies $\sup_{x \in [0,1]} |P_n(x) - f(x)| < \epsilon$. Notice that

$$\begin{aligned} P_n(x) - f(x) &= \sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k} \\ &\quad - f(x) \sum_{k=0}^n \binom{n}{k} x^k (1-x)^{n-k} \\ &= \sum_{k=0}^n \left[f\left(\frac{k}{n}\right) - f(x) \right] \binom{n}{k} x^k (1-x)^{n-k}. \end{aligned} \quad (5.13)$$

We want to use the uniform continuity of f to make this small. So let $\epsilon > 0$ and pick $\delta > 0$ such that $|x - y| < \epsilon/2$ implies

$$|f(x) - f(y)| < \epsilon/2.$$

We therefore want to consider values of k/n such that $|x - k/n| < \delta$. Since k, n are integers we need to use the integer part function $[x] =$ greatest integer $\leq x$. We split the sum as

$$\begin{aligned} P_n(x) - f(x) &= \sum_{[x-k/n] < \delta} \left[f\left(\frac{k}{n}\right) - f(x) \right] \binom{n}{k} x^k (1-x)^{n-k} \\ &\quad + \sum_{[x-k/n] \geq \delta} \left[f\left(\frac{k}{n}\right) - f(x) \right] \binom{n}{k} x^k (1-x)^{n-k}. \end{aligned}$$

Using the fact that $\sum_{k=0}^n \binom{n}{k} x^k (1-x)^{n-k} = 1$ and the continuity of f we get

$$\left| \sum_{[x-k/n] < \delta} \left[f\left(\frac{k}{n}\right) - f(x) \right] \binom{n}{k} x^k (1-x)^{n-k} \right| < \frac{\epsilon}{2},$$

since $|k/n - x| < \delta$. For the second sum

$$\begin{aligned} &\left| \sum_{[x-k/n] \geq \delta} \left[f\left(\frac{k}{n}\right) - f(x) \right] \binom{n}{k} x^k (1-x)^{n-k} \right| \\ &\leq 2 \sup_{x \in [0,1]} |f(x)| \left| \sum_{[x-k/n] \geq \delta} \binom{n}{k} x^k (1-x)^{n-k} \right|. \end{aligned}$$

Now $|x - k/n| \geq \delta$, so that $(x - k/n)^2/\delta^2 \geq 1$. We can then produce the estimate

$$\begin{aligned} \sum_{[x-k/n] \geq \delta} \binom{n}{k} x^k (1-x)^{n-k} &\leq \frac{1}{\delta^2} \sum_{k=0}^n \left(x - \frac{k}{n}\right)^2 \binom{n}{k} x^k (1-x)^{n-k} \\ &= \frac{1}{\delta^2} \sum_{k=0}^n \left(x^2 - \frac{2xk}{n} + \frac{x^2}{k^2}\right) \binom{n}{k} x^k (1-x)^{n-k} \\ &= \frac{1}{\delta^2} \left(x^2 - \frac{2x}{n} nx + \frac{1}{n^2} (n(n-1)x^2 + nx)\right) \\ &= \frac{x(1-x)}{n\delta^2} \leq \frac{1}{4n\delta^2}, \end{aligned}$$

since $x(1-x) \leq 1/4$ if $x \in [0, 1]$. We thus arrive at

$$\left| \sum_{[x-k/n] \geq \delta} \left[f\left(\frac{k}{n}\right) - f(x) \right] \binom{n}{k} x^k (1-x)^{n-k} \right| \leq \frac{1}{4n\delta^2} 2A, \quad (5.14)$$

where $A = \sup_{x \in [0,1]} |f(x)|$. Hence we need to choose $n \geq 2\epsilon A/\delta^2$. This will guarantee that $\sup_{x \in [0,1]} |P_n(x) - f(x)| < \epsilon$. \square

Weierstrass' Theorem was extended by Marshall Stone to algebras of functions on more abstract spaces and the result is known as the Stone-Weierstrass Theorem. It plays a central role in modern analysis.

Remark 5.17. A word of caution. Weierstrass' Theorem does not mean that the Taylor series of a function f will converge to f uniformly. The Taylor series may not exist, since we only assume continuity, not differentiability for f . Even when f is infinitely differentiable the Taylor series may still not converge.

The point of this result is that polynomials are easy to evaluate and work with. So for example, if we have a continuous function f on a closed interval $[a, b]$ and we want to integrate f over $[a, b]$, then given ϵ , we can find a polynomial P such that for all $x \in [a, b]$ we have $|f(x) - P(x)| < \epsilon$. Hence

$$\begin{aligned} \left| \int_a^b f(x) dx - \int_a^b P(x) dx \right| &\leq \int_a^b |f(x) - P(x)| dx \\ &< \int_a^b \epsilon dx = \epsilon(b-a). \end{aligned}$$

We can perform the last step because the approximation is uniform. Thus the inequality holds for all x . So by choosing ϵ small enough and the appropriate P we can approximate the integral of f on $[a, b]$ to high accuracy by integrating P , which is straightforward since it is a polynomial.

For many problems, the methods we have developed are sufficient to provide a solution. However analysis does not stop at this point. There are areas where more sophisticated techniques are needed. For example, do we really need uniform convergence to swap a limit and an integral, or can we do better? The answer is that yes we can, but it will require us to completely redefine what we mean by integration.

At the start of the twentieth century the French mathematician Henri Lebesgue invented a new integral that bears his name. Many of the limitations of the Riemann integral are overcome with this new form of integration. See the subject Lebesgue Integration and Fourier Analysis for the development of this new theory. Nevertheless, the Riemann integral remains a major tool in mathematics.

The tools we have developed here underpin much of modern mathematics. There are many directions that we can go in. We can extend our results to higher dimensions, vector valued functions and complex variables. We can also study analysis in more abstract settings. We can study analysis on vector spaces, which provides the mathematical underpinning for much of modern physics. Indeed Analysis is an active field, with new discoveries being made all the time.