MA131B

Analysis II
Revision Guide

Written by David McCormick

Contents

Introduction

This revision guide for MA131B Analysis II has been designed as an aid to revision, not a substitute for it. As can be seen from the length of this revision guide, Analysis II is a fairly long course with lots of confusing definitions and big theorems; hopefully this guide will help you make sense of how everything fits together. Only some proofs are included; the inclusion of a particular proof is no indication of how likely it is to appear on the exam. However, the exam format doesn't change much from year to year, so the best way of revising is to do past exam papers, as well as assignment questions, using this guide as a reference.

For further practice, the questions in R. P. Burn's *Numbers and Functions* are invaluable, and indeed one of the principal sources of this revision guide. For the more brave of heart, Rudin's *Principles of Mathematical Analysis* is a classic, unparalleled textbook, but is pitched somewhat above the level of the course, with some of the most challenging analysis questions known to man.

Disclaimer: Use at your own risk. No guarantee is made that this revision guide is accurate or complete, or that it will improve your exam performance. Use of this guide *will* increase entropy, contributing to the heat death of the universe. Contains no GM ingredients. Your mileage may vary. All your base are belong to us.

Authors

Written by D. S. McCormick (d.s.mccormick@warwick.ac.uk)
Based upon lectures given by David Elworthy at the University of Warwick, 2005 and 2006.
Minor corrections Stephen Loke & Guy Barwell, 2012.
Any corrections or improvements should be entered into our feedback form at http://tinyurl.com/WMSGuides (alternatively email revision.guides@warwickmaths.org).

History

First Edition: 2007.
Current Edition: January 18, 2020.

1 Review of MA131A Analysis I

Analysis is a linear subject, in the sense that most results depend critically on previous results. As such, a sound knowledge of MA131A Analysis I is vital for success in the Analysis II exam, and one question on the Analysis II exam will be exclusively on material from Analysis I. We state some of the key definitions and theorems for reference:

Definitions 1.1. A (real) *sequence* is a list of (real) numbers in a definite order (or a function $\mathbb{N} \to \mathbb{R}$), written $(a_n)_{n=1}^{\infty} = (a_1, a_2, \dots)$.

A sequence (a_n) is *increasing* if $a_n \le a_{n+1}$ for all n; it is *strictly increasing* if $a_n < a_{n+1}$ for all n; similarly for decreasing. A sequence is *monotonic* if it is either increasing or decreasing.

A sequence (a_n) is *bounded above* if $\exists U$ s.t. $a_n \le U$ for all n; similarly for bounded below. A sequence is *bounded* if it is bounded above and below.

A sequence (a_n) *converges* to a if $\forall \varepsilon > 0$, $\exists N \in \mathbb{N}$ s.t. $n \ge N \implies |a_n - a| < \varepsilon$; write $(a_n) \to a$.

A sequence (a_n) is *Cauchy* if $\forall \varepsilon > 0$, $\exists N \in \mathbb{N}$ s.t. $n, m \ge N \implies |a_n - a_m| < \varepsilon$.

A sequence (a_n) *tends to infinity* if $\forall C > 0$, $\exists N \in \mathbb{N}$ s.t. $n \ge N \implies a_n > C$; write $(a_n) \to \infty$.

A *subsequence* of (a_n) is a sequence of the form (a_{n_i}), where (n_i) is a strictly increasing sequence of natural numbers.

Proposition 1.2. Let $a, b \in \mathbb{R}$. Suppose $(a_n) \to a$ and $(b_n) \to b$. Then for any $c, d \in \mathbb{R}$, $(ca_n + db_n) \to ca + db$, $(a_n b_n) \to ab$, and if[1] $b \neq 0$ then $\left(\frac{a_n}{b_n}\right) \to \frac{a}{b}$.

Theorem 1.3 (Closed Interval Rule)**.** Suppose $(a_n) \to a$. If (eventually[2]) $A \le a_n \le B$, then $A \le a \le B$.

Completeness Axiom. Every non-empty subset $A \subseteq \mathbb{R}$ that is bounded above (resp. below) has a least upper bound, $\sup A$ (resp. greatest lower bound, $\inf A$). Equivalently:

Every bounded monotonic sequence is convergent.

Every bounded sequence has a convergent subsequence (this is the Bolzano–Weierstrass theorem).

Every Cauchy sequence is convergent.

Definitions 1.4. A *series* is an expression of the form $\sum_{n=1}^{\infty} a_n = a_1 + a_2 + \cdots$. Consider the series $\sum_{n=1}^{\infty} a_n$ with partial sums (s_n), where $s_n = \sum_{i=1}^{n} a_i$. We say:

$\sum a_n$ *converges* if (s_n) converges. If $s_n \to S$ then we write $\sum_{n=1}^{\infty} a_n = S$.

$\sum a_n$ *diverges* if (s_n) does not converge.

$\sum a_n$ *diverges to infinity* if (s_n) tends to infinity.

$\sum a_n$ *diverges to minus infinity* if (s_n) tends to minus infinity.

The series $\sum a_n$ is *absolutely convergent* if $\sum |a_n|$ is convergent.

Lemma 1.5. Every absolutely convergent series is convergent.

Proposition 1.6. Suppose $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ are convergent series. Then, for all $c, d \in \mathbb{R}$, $\sum_{n=1}^{\infty}(ca_n + db_n)$ is a convergent series and $\sum_{n=1}^{\infty}(ca_n + db_n) = c\sum_{n=1}^{\infty} a_n + d\sum_{n=1}^{\infty} b_n$.

Theorem 1.7 (Null Sequence Test)**.** If (a_n) does not tend to zero, then $\sum_{n=1}^{\infty} a_n$ diverges.

Theorem 1.8 (Comparison Test)**.** Suppose $0 \le a_n \le b_n$ for all natural numbers n. If $\sum b_n$ converges then $\sum a_n$ converges and $\sum_{n=1}^{\infty} a_n \le \sum_{n=1}^{\infty} b_n$.

Lemma 1.9 (Geometric Series)**.** The series $\sum_{n=0}^{\infty} x^n$ is convergent if $|x| < 1$ and the sum is $\frac{1}{1-x}$. It is divergent if $|x| \ge 1$.

Theorem 1.10 (Ratio Test)**.** Suppose $a_n \neq 0$. If $\left|\frac{a_{n+1}}{a_n}\right| \to k$, then $\sum a_n$ converges absolutely if $0 \le k < 1$ and diverges if $k > 1$. If $\left|\frac{a_{n+1}}{a_n}\right| \to \infty$, then $\sum a_n$ diverges.

In analysis II we have a better way (in most cases) of proving that the series converges, namely Cauchy's n^{th} Root Test. This also provides radius of convergence (covered in the "Power Series" section).

[1] $b \neq 0 \implies b_n \neq 0$ eventually; for the finitely many n where $b_n = 0$ we can ignore the terms a_n/b_n.

[2] Something holds *eventually* if there exists an $N \in \mathbb{N}$ such that it holds for all $n \ge N$.

2 Continuity

Fundamental to all of mathematics is the notion of a function:

Definition 2.1. Given two sets A and B, a *function* $f\colon A \to B$ is a pairing of each element of A with an element of B. The element of B which is paired with $a \in A$ is denoted by $f(a)$.

Here the set A is called the *domain* of the function, while the set B is called the *co-domain* of the function. The *range* of the function is the set of points in B which is mapped to by some point in A, i.e. $\{f(a) \mid a \in A\} \subseteq B$. Throughout this module we only consider functions $f\colon A \to \mathbb{R}$, where $A \subseteq \mathbb{R}$; that is, real-valued functions of one real variable.

The general definition of function simply requires pairing each value x in the domain with a value $f(x)$ in the co-domain, even when something like a fresh formula is needed for each x, and so does not require the value of $f(x)$ to be anywhere close to the value of $f(y)$, even when x and y are arbitrarily close to one another. We are thus moved to define various classes of functions which are well-behaved; most important are *continuous functions*, which makes rigorous the idea that if x and y are close to each other, then $f(x)$ and $f(y)$ should also be close to each other.

2.1 Defining Continuity

In Analysis I, we used the idea of limits of sequences to get "close" to a number using a sequence tending to that number. So, our first definition of continuity (which we will call *sequential continuity* for the time being) is based on these ideas:

Definition 2.2. Given a function $f\colon A \to \mathbb{R}$, where $A \subseteq \mathbb{R}$, and given $a \in A$, then f is *(neighbourhood) continuous* at $x = a$ if, for each $\varepsilon > 0$, we can find a $\delta > 0$ such that

$$x \in A \quad \text{and} \quad |x - a| < \delta \implies |f(x) - f(a)| < \varepsilon$$

Definition 2.3. Given a function $f\colon A \to \mathbb{R}$, where $A \subseteq \mathbb{R}$, and given $a \in A$, then f is *sequentially continuous* at $x = a$ if, for every sequence (a_n) converging to a, where $a_n \in A$ for all n, the sequence $\big(f(a_n)\big)$ converges to $f(a)$.

This is commonly expressed as $f(x) \to f(a)$ as $x \to a$. But another way to describe nearness is to consider *neighbourhoods* of a point a, i.e. a set $\{x \mid a - \delta < x < a + \delta\}$ for some $\delta > 0$.

For a given function f, our choice of δ depends on the point a and the challenge ε; we sometimes write $\delta = \delta_a(\varepsilon)$ to emphasise this.

Having two different definitions of continuity would be disastrous if they led to different consequences. Thankfully however, we show now that the two definitions are in fact equivalent.

Theorem 2.4. Given a function $f\colon A \to \mathbb{R}$, where $A \subseteq \mathbb{R}$, and given $a \in A$, then f is (neighbourhood) continuous at $x = a$ if and only if f is sequentially continuous at $x = a$.

Proof. ($\Longrightarrow$) Suppose f is neighbourhood continuous at $x = a$. Let $(a_n) \to a$ be a sequence in A. As f is neighbourhood continuous at a, we have that $\forall \varepsilon > 0$, $\exists\, \delta > 0$ such that $x \in A$ and $|x - a| < \delta \implies |f(x) - f(a)| < \varepsilon$. Now $a_n \in A$ for all n, so $|a_n - a| < \delta \implies |f(a_n) - f(a)| < \varepsilon$. But since $(a_n) \to a$, we have that $\forall\, \delta > 0$, $\exists\, N$ such that for $n \geq N$ we have $|a_n - a| < \delta$, and hence that $|f(a_n) - f(a)| < \varepsilon$. So $\big(f(a_n)\big) \to f(a)$.

($\Longleftarrow$) We use the contrapositive; suppose f is not neighbourhood continuous at $x = a$. Then $\exists\, \varepsilon > 0$ such that given any $\delta > 0$, there is at least one x for which $|x - a| < \delta$, but $|f(x) - f(a)| \geq \varepsilon$. For each $n \in \mathbb{N}$, let $\delta = \frac{1}{n}$ and choose an x satisfying this condition and label it a_n. Now for each n we have $|a_n - a| < \frac{1}{n}$, so $(a_n) \to a$. But for each n, $|f(a_n) - f(a)| \geq \varepsilon$ for some $\varepsilon > 0$. Hence $\big(f(a_n)\big)$ does not tend to $f(a)$, and f is not sequentially continuous at $x = a$. □

Definition 2.5. If a function $f\colon A \to \mathbb{R}$ is continuous at every point $a \in A$, then we say f is *continuous on A*, or simply a *continuous function*.

Lemma 2.6 (Preservation of sign)**.** If $f : E \to R$ is continuous at $c \in E$ and $f(c) > 0$ then there exists a $\delta > 0$ such that $f(x) > 0$ for all $x \in E$ with $|x - c| < \delta$. Similarly, if $f(c) < 0$ then there exists a $\delta > 0$ such that $f(x) < 0$ for all $x \in E$ with $|x - c| < \delta$.

Proof. If $f : E \to R$, $E \subseteq \mathbb{R}$ is continuous at $c \in E$ and $f(c) > 0$ then $\exists \delta > 0$ such that $|x - c| < \delta \implies |f(x) - f(c)| < f(c)$ so $f(x) > -f(c) + f(c) = 0$ by the same reasoning if $f(c) < 0$ then there exists $\delta > 0$ such that for $|x - c| < \delta \implies f(x) < 0$ □

Notice that if f is not continuous at $c \in E$ this does not have to hold, consider $f(x) = -1$ for $x \neq 0$ and $f(x) = 1$ for $x = 0$ is not continuous at 0 and preservation of sign does not hold. Notice also, we can change 0 in the inequalities $f(c) > 0$ or $f(c) < 0$ for any number p to give a similar result. For example, with a function continuous at c and satisfying the inequality $f(c) > p$, it holds that there is a neighbourhood $A \subseteq E$ around c such that for any $x \in A$ the inequality $f(x) > p$ holds, (similarly for $f(c) < p$). Can you prove this?

Most functions we are familiar with are continuous. For example, the functions $f, g \colon \mathbb{R} \to \mathbb{R}$ given by $f(x) = c$ (for any constant $c \in \mathbb{R}$) and $g(x) = x$ are continuous. Moreover, when we have continuous functions, we can create new continuous functions by adding, subtracting, multiplying or dividing them. The following result follows immediately from the sequential definition of continuity and the analogous results for sequences:

Proposition 2.7. Given $A \subseteq \mathbb{R}$, and given $a \in A$, suppose that $f, g \colon A \to \mathbb{R}$ are continuous at a. Then:

1. the function $f + g \colon A \to \mathbb{R}$ defined by $(f + g)(x) := f(x) + g(x)$ is continuous at a;
2. the function $f \cdot g \colon A \to \mathbb{R}$ defined by $(f \cdot g)(x) := f(x) \cdot g(x)$ is continuous at a; and
3. if also $g(x) \neq 0$ for all $x \in A$, then the function $\frac{f}{g} \colon A \to \mathbb{R}$ defined by $\left(\frac{f}{g}\right)(x) = \frac{f(x)}{g(x)}$ is continuous at a.

Since the function $x \mapsto x$ is continuous, we can multiply it by itself to get that $x \mapsto x^n$ is continuous for any $n \in \mathbb{N}$. Multiplying by any constant a_n shows that $x \mapsto a_n x^n$ is continuous. Hence the sum $x \mapsto a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$ is continuous, i.e.:

Proposition 2.8. Any (finite) polynomial function $p \colon \mathbb{R} \to \mathbb{R}$ given by $p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$ is continuous.

Corollary 2.9. Any rational function $q \colon A \to \mathbb{R}$ given by $q(x) = \frac{p_1(x)}{p_2(x)}$ where p_1, p_2 are (finite) polynomials is continuous provided that $p_2(x) \neq 0$ for all $x \in A$.

Theorem 2.10. If the function $f \colon A \to \mathbb{R}$ is continuous at $x = a$, and the function $g \colon B \to \mathbb{R}$ is continuous at $f(a)$, where $f(A) \subseteq B$, then $g \circ f$ is continuous at $x = a$.

That is, a continuous function of a continuous function is continuous.

Proof. Let $(a_n) \to a$ be a sequence in A. By the continuity of f at a, $\big(f(a_n)\big) \to f(a)$. Then, since g is continuous at $f(a)$, $\big(g(f(a_n))\big) \to g(f(a))$. Hence $g \circ f$ is continuous at $x = a$. □

Corollary 2.11. If $f : E \to \mathbb{R}$ is continuous at c and $f(c) \neq 0$ then $\frac{1}{f}$ is well-defined in a neighbourhood of c and is continuous at c.

Examples 2.12. Continuous functions are so familiar that to clarify the meaning of this definition we need some examples of discontinuity, illustrating the absence of continuity. The continuity of a function at a point a is recognised by the way the function affects ***every*** sequence (a_n) tending to a. Discontinuity is established by finding just one sequence $(a_n) \to a$ for which $\big(f(a_n)\big)$ does not tend to $f(a)$.

1. Consider again the integer part function $f(x) = [x]$. f is continuous for any non-integer x and discontinuous at any integer x. Thus f has an infinity of isolated discontinuities.
2. Consider the function $f \colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = 0$ when x is rational, and $f(x) = x$ when x is irrational. This is continuous at just a single point, $x = 0$.

3. Consider the function $f(x) = \sin\frac{1}{x}$. This is continuous if $x \neq 0$, but what about $x = 0$? Let $a_n = \frac{1}{2n\pi}$ and $b_n = \frac{1}{(2n+\frac{1}{2})\pi}$. Now $(a_n) \to 0$ and $(b_n) \to 0$, but $f(a_n) = 0$ for all n, and $f(b_n) = 1$ for all n. Hence, if we try to extend f to a continuous function by defining

$$\tilde{f}(x) = \begin{cases} f(x) & \text{if } x > 0 \\ k & \text{if } x = 0 \end{cases}$$

then we will fail, since whatever value of k we choose there will exist sequences (a_n) such that $\big(f(a_n)\big)$ does not tend to k.

Lemma 2.13. For $x \in (0, \frac{\pi}{2})$ we have that $0 < \sin x < x < \tan x$

Proof. Firstly, take a unit circle with centre O and two distinct points A,B on its circumference with the angle between OA and OB less than $\frac{\pi}{2}$. Then take a point E outside the circle such that OBE is a right angle and notice that the area of triangle OAB < sector OAB < area of triangle OBE. Finally, by denoting the angle of the sector by x we get the desired inequality. ☐

Theorem 2.14. The function $f : \mathbb{R} \to [-1, 1]$ given by $f(x) = \sin x$ is continuous.

Proof. Firstly, note that $\sin(x) = \sin(\frac{x-c}{2} + \frac{x+c}{2})$ and $\sin(c) = sin(\frac{x+c}{2} - \frac{x-c}{2})$ and then set $\delta = \epsilon$ (where $|x - c| < \delta$). Using the identities above and then the addition formula for sine we get $|\sin(x) - \sin(c)| = |2\sin(\frac{x-c}{2})\cos(\frac{x+c}{2})|$. Hence, $|2\sin(\frac{x-c}{2})\cos(\frac{x+c}{2})| < |2\sin(\frac{x-c}{2})| < |x - c| < \delta = \epsilon$ and therefore $\sin x$ is continuous. ☐

2.2 Continuity and Completeness

Up until now, our results on continuity have not depended on the property of completeness. However, using completeness gets us some of the most useful results to do with continuity.

Intervals, or connected sets, on the real line are subsets of the real line which contain all the real numbers lying between any two points of the subset. The set I is an *interval* if when $r, s \in I$, and $r < s$, then every x such that $r < x < s$ also belongs to I. The seven types of interval are distinguished by their boundedness and their boundaries:

bounded above and below:
- (i) singleton point, $\{a\}$
- (ii) closed [3] interval, $[a, b] = \{x \in \mathbb{R} \mid a \leq x \leq b\}$
- (iii) open[4] interval, $(a, b) = \{x \in \mathbb{R} \mid a < x < b\}$
- (iv) half-open interval, $[a, b) = \{x \in \mathbb{R} \mid a \leq x < b\}$ or $(a, b] = \{x \in \mathbb{R} \mid a < x \leq b\}$

bounded above or below, but not both:
- (v) open half-ray, $(a, \infty) = \{x \in \mathbb{R} \mid x > a\}$ or $(-\infty, a) = \{x \in \mathbb{R} \mid x < a\}$
- (vi) closed half-ray, $[a, \infty) = \{x \in \mathbb{R} \mid x \geq a\}$ or $(-\infty, a] = \{x \in \mathbb{R} \mid x \leq a\}$

unbounded:
- (vii) the whole real line, $\mathbb{R}$

Definition 2.15. Let $E \subseteq \mathbb{R}$ function $f\colon E \to \mathbb{R}$ is strictly increasing if $\forall x, y \in E, x > y \implies f(x) > f(y)$, increasing if $\forall x, y \in E, x > y \implies f(x) \geq f(y)$, decreasing if $\forall x, y \in E, x > y \implies f(x) \leq f(y)$ and strictly decreasing if $\forall x, y \in E, x > y \implies f(x) < f(y)$

The *Intermediate Value Theorem*, or IVT, tells us that a continuous function maps intervals onto intervals:

Theorem 2.16 (IVT version 1)**.** If $f\colon [a, b] \to \mathbb{R}$ is a continuous function with $f(a) < f(b)$, then for each k such that $f(a) < k < f(b)$, there exists $c \in (a, b)$ such that $f(c) = k$.

[3] The word *closed* as we have used it here indicates that every convergent sequence within each closed set converges to a point of the set (see the Closed Interval Rule, theorem 1.3).

[4] The word *open* as we have used it here indicates that there is space *within* each open set around each point of the set. For example, if $c \in (a, b)$, then $a < \frac{1}{2}(a + c) < c < \frac{1}{2}(c + b) < b$. This is also described by saying that an open set contains a neighbourhood of each of its points.

Proof. Let $S = \{x \in [a,b] \mid f(x) \le k\}$, so $a \in S$ and $x \in S \implies x \le b$. Therefore S is non-empty and bounded above, and so has a least upper bound, $\sup S$. Set $c = \sup S$. We claim $f(c) = k$. Now, since $c = \sup S$, there exists a sequence (x_n) in S with $x_n \to c$ as $n \to \infty$. As f is continuous at c, $f(x_n) \to f(c)$ as $n \to \infty$. But $\forall n$, $f(x_n) \le k$ by definition of S, hence $f(c) \le k$ (by the Closed Interval Rule).

Next, for contradiction, assume $f(c) < k$. Take $\varepsilon = \frac{1}{2}(k - f(c)) > 0$. Since f is continuous at c, $\exists \delta > 0$ such that $x \in [a,b]$ and $|x - c| < \delta \implies |f(x) - f(c)| < \frac{1}{2}(k - f(c))$. Hence $f(x) < f(c) + \frac{1}{2}(k - f(c)) = \frac{1}{2}(k + f(c)) < k$, so $x \in S$. But this means that $x \in (c, c+\delta) \implies x \in S$, which is impossible since $c = \sup S$. Hence $f(c) \ge k$. So $f(c) = k$ as required. □

Corollary 2.17 (IVT version 2)**.** If $f\colon [a,b] \to \mathbb{R}$ is a continuous function with $f(b) < f(a)$, then for each k such that $f(b) < k < f(a)$, there exists $c \in (a,b)$ such that $f(c) = k$.

Note that we most definitely require completeness for the Intermediate Value Theorem – if we work over $\mathbb{Q}$ the theorem vanishes in a puff of smoke. For example, consider $f\colon \mathbb{Q} \to \mathbb{Q}$ given by $f(x) = -1$ if $x^2 < 2$ and $f(x) = 1$ if $x^2 > 2$. Then f is continuous with $f(0) = -1$ and $f(2) = 1$, but there is no c with $f(c) = 0$. Notice also that the IVT does not work if a function, f, is not continuous at a point. E.g. $f(x) = -1$ for $x \neq 0$ and $f(x) = 1$ for $x = 0$ then $f(0) = 1$ and $f(1) = 0$ but there is no point $c \in [0,1]$ such that $f(c) = \frac{1}{2}$.

Lemma 2.18. Any odd degree polynomial has a real root.

Proof. Consider $P(x) = \sum_{j=1}^{2n+1} a_j x^j$ without loss of generality let $a_{2n+1} > 0$ then for $x > 1$, $P(x) > a_{2n+1}x^{2n+1} - \sum_{j=1}^{2n} |a_j|$ and so for $x > \frac{\sum_{j=1}^{2n} |a_j|}{a_{2n+1}}$ we have that $P(x) > 0$ and for $x < -\frac{\sum_{j=1}^{2n} |a_j|}{a_{2n+1}}$ we have $P(x) < 0$ and so by IVT there exists a root somewhere in the range $(-\frac{\sum_{j=1}^{2n} |a_j|}{a_{2n+1}}, \frac{\sum_{j=1}^{2n} |a_j|}{a_{2n+1}})$. □

If we consider $f : [a,b] \to [a,b]$ and apply the IVT to $f(x) - x$, we get a simple *fixed point theorem*[5]:

Theorem 2.19 (Fixed point theorem)**.** If $f : [a,b] \to [a,b]$ is continuous, then there exists at least one $c \in [a,b]$ such that $f(c) = c$.

The Intermediate Value Theorem tells us that the range of a continuous function on an interval is also an interval. A continuous function on a *closed interval* has special properties that lead to a stronger conclusion.

2.3 Limits

Definition 2.20 (Continuous limit)**.** For $f\colon (a,b)\setminus\{c\} \to \mathbb{R}$ then we say that $f(x)$ tends to α as x tend to c, denoted $f(x) \to \alpha$ as $x \to c$, if $\forall \epsilon > 0 \exists \delta > 0$ such that for $|x - c| < \delta \implies |f(x) - \alpha| < \epsilon$

Proposition 2.21 (Sandwich Rule)**.** Suppose that $f, g, h\colon (a,b)\setminus\{c\} \to \mathbb{R}$, and that $f(x) \le g(x) \le h(x)$ then if $\lim_{x \to c} f(x) = \lim_{x \to c} h(x) = \alpha$ then $lim_{x \to c} g(x) = \alpha$

Proposition 2.22 (Uniqueness of Limits)**.** If $\lim_{x \to a} f(x) = l_1$ and $\lim_{x \to a} f(x) = l_2$, then $l_1 = l_2$.

Some kinds of discontinuities may be identified and discussed by considering the two sides of the point in question separately. On this basis we define *one-sided limits*:

Definition 2.23. Let $f\colon A \to \mathbb{R}$. If, for any sequence (a_n) in A with $a_n < a$ and $(a_n) \to a$, we have $\big(f(a_n)\big) \to l$, we write $\lim\limits_{x \to a^-} f(x) = l$. Similarly, if, for any sequence (a_n) in A with $a_n > a$ and $(a_n) \to a$, we have $\big(f(a_n)\big) \to l$, we write $\lim\limits_{x \to a^+} f(x) = l$.

For example, for some integer $a \in \mathbb{Z}$, we have that $\lim_{x \to a^-} [x] = [a] - 1$ and $\lim_{x \to a^+} [x] = [a]$. Note that we do *not* require $a \in A$; we do *not* take account of the value of $f(a)$ – in fact we do not even require that f be defined at a, only that you can get arbitrarily close to a and still take values of $f(x)$.

[5] In higher dimensions this is completely non-trivial, since we do not have any inequalities when dealing with vectors. The full version is known as *Brouwer's fixed point theorem* and is covered in detail in MA3H5 Manifolds.

Theorem 2.24 (Algebra of Limits)**.** Suppose $\lim_{x\to a} f(x) = l$ and $\lim_{x\to a} g(x) = m$. Then $\lim_{x\to a}\big(f(x)+g(x)\big) = l+m$, $\lim_{x\to a}\big(f(x)\,g(x)\big) = l\,m$, and if $m \neq 0$ then $\lim_{x\to a}\frac{f(x)}{g(x)} = \frac{l}{m}$.

Just as we can define continuity in terms of sequences and neighbourhoods, we can do the same for limits. We prove this in a similar way to the proof of theorem 2.4.

Definition 2.25. Given a function $f\colon A \to \mathbb{R}$, $\lim_{x\to a^-} f(x) = l$ if and only if for each $\varepsilon > 0$, there is a $\delta > 0$ such that $x \in A$ and $a-\delta < x < a \implies |f(x)-l| < \varepsilon$.

Similarly, $\lim_{x\to a^+} f(x) = l$ if and only if for each $\varepsilon > 0$, there is a $\delta > 0$ such that $x \in A$ and $a < x < a+\delta \implies |f(x)-l| < \varepsilon$.

If, for a particular point a, we have $\lim_{x\to a^-} f(x) = l$ and $\lim_{x\to a^+} f(x) = l$ then it is customary to write $\lim_{x\to a} f(x) = l$. We then have:

The following properties of so-called *two-sided limits* follow immediately:

Moreover, by chasing definitions it is easy to show that for continuous functions the limit of the function is the function of the limit:

Proposition 2.26. $f\colon A \to \mathbb{R}$ is continuous at $a \in A$ if and only if $\lim_{x\to a^-} f(x) = \lim_{x\to a^+} f(x) = f(a)$, i.e. if and only if $\lim_{x\to a} f(x) = f(a)$.

Definition 2.27. We say that $f(x) \to \alpha$ as $x \to \infty$ if $\forall\epsilon > 0\ \exists$ R such that $\forall x > R$ we have $|f(x)-\alpha| < \epsilon$ and that $f(x) \to \alpha$ as $x \to -\infty$ if $\forall\epsilon > 0\ \exists$ R such that $\forall x < R$ we have $|f(x)-\alpha| < \epsilon$

Definition 2.28. We say that $f(x) \to \infty$ as $x \to c$ if $\forall R\ \exists\delta > 0$ such that $f(x) > R$ for all x with $0 < |x-c| < \delta$ and we say that $f(x) \to -\infty$ as $x \to c$ if $\forall R\ \exists\delta > 0$ such that $f(x) < R$ for all x with $0 < |x-c| < \delta$

Definition 2.29. We say that $f(x) \to \infty$ as $x \to \infty$ if $\forall R\ \exists\delta$ such that $f(x) > R$ with $x > \delta$ and we say that $f(x) \to -\infty$ as $x \to \infty$ if $\forall R\ \exists\delta$ such that $f(x) < R$ with $x > \delta$

Many of the results we established for continuous functions – such as sum and product rules – carry over easily into limits, and are proved easily using the analogous results for sequences.

Theorem 2.30 (Algebra of Limits)**.** Suppose $\lim_{x\to a} f(x) = l$ and $\lim_{x\to a} g(x) = m$. Then $\lim_{x\to a}\big(f(x)+g(x)\big) = l+m$, $\lim_{x\to a}\big(f(x)\,g(x)\big) = l\,m$, and if $m \neq 0$ then $\lim_{x\to a}\frac{f(x)}{g(x)} = \frac{l}{m}$.

Note, however, that while the composition of continuous functions is continuous, limits do not behave as nicely under composition, and there is no analogous simple statement about limits of composed functions.

By replacing "for all $\varepsilon > 0$" with "for all $C > 0$" and changing $|f(x)-l| < \varepsilon$ to $f(x) > C$, we get *infinite limits*. We can also replace "there exists $\delta > 0$" with "there exists $C > 0$" and change $|x-a| < \delta$ with $x > C$ to get *limits at infinity*. Practise playing with the definitions, and also with evaluating all manner of limits; such questions are popular on exams.

2.4 Continuous functions on closed bounded intervals

Theorem 2.31 (Extreme Value Theorem)**.** If $f\colon [a,b] \to \mathbb{R}$ is continuous, then the range of f, i.e. $f([a,b])$ is bounded. Furthermore, f attains its bounds, i.e. there exist $x_1, x_2 \in [a,b]$ such that $f(x_1) = \inf_{x\in[a,b]} f(x)$ and $f(x_2) = \sup_{x\in[a,b]} f(x)$.

Notice that if the range isn't closed, e.g. $(0,1]$ or $(0,1)$ then the EVT may not apply, as an example consider $f\colon (0,1) \to \mathbb{R}$, $f(x) = x$ then both $f(x_1) = \inf_{x\in(0,1)} f(x)$ and $f(x_2) = \sup_{x\in(0,1)} f(x)$ don't exist.

We used the Intermediate Value Theorem to find individual solutions, i.e. when we have a function on a closed interval $[a,b]$ we can say that the function takes every value between $f(a)$ and $f(b)$. What would be much more useful would be to be able to say that f is a bijection – that f has a well-defined inverse function. However, continuity alone is not sufficient to ensure that f is a bijection – we need $f(x)$ to occur exactly once in the range in order that we can invert it, i.e. we require f to be injective. Once we assume this, we get:

Proposition 2.32. Suppose $f\colon [a,b] \to \mathbb{R}$ is continuous, injective and $f(a) < f(b)$. Then the range of f is the interval $[f(a), f(b)]$, and $f\colon [a,b] \to [f(a), f(b)]$ is a strictly increasing bijection, whose inverse $f^{-1}\colon [f(a), f(b)] \to [a,b]$ is also continuous and strictly increasing.

Proposition 2.33. If $f\colon [a,b] \to \mathbb{R}$ is continuous with $f(a) < f(b)$ but is not strictly increasing then it is not injective.

Proposition 2.34. If $f\colon [a,b] \to [f(a), f(b)]$ is non-decreasing and surjective then it is continuous.

Corollary 2.35. Let $f\colon [a,b] \to \mathbb{R}$ be continuous and strictly monotonic. Then $f^{-1}\colon f([a,b]) \to [a,b]$ is continuous and strictly monotonic.

Example 2.36. The function $\sin\colon [-\frac{\pi}{2}, \frac{\pi}{2}] \to \mathbb{R}$ is continuous and injective; hence it is strictly increasing and bijective onto $[-1,1]$, and its inverse $\arcsin\colon [-1,1] \to [-\frac{\pi}{2}, \frac{\pi}{2}]$ is also continuous.

3 Differentiation

Having dealt with continuous functions, the next class of "useful" functions which we come to are the *differentiable functions*. We first ask a simple question: how can we find the gradient, or rate of change, of f? If we consider a point $(a, f(a))$, then the line through this point with gradient m is $y - f(a) = m(x - a)$. Now, the equation of the chord joining $(a, f(a))$ and $(a+h, f(a+h))$ is $y - f(a) = \frac{f(a+h)-f(a)}{h}(x - a)$. If we let $m = \lim_{h \to 0} \frac{f(a+h)-f(a)}{h}$, then we would surely describe the line $y - f(a) = m(x - a)$ as being a tangent to the curve at $x = 0$, and m as being the gradient of this tangent. On the strength of these ideas, we define the derivative at a point a of its domain:

Definition 3.1. For a function $f\colon A \to \mathbb{R}$ and a point $a \in A$, if

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

for some real number m, then m is called the *derivative* of f at a, usually denoted by $f'(a)$, and f is said to be *differentiable* at a. (The two limits are equivalent; the first exists if and only if the second exists, and they are equal if they both exist.)

Although derivatives arise naturally from considering the geometric notion of tangent, there is still one circumstance when a tangent to the graph of a function may exist without a derivative of the function at the point in question; this occurs when the slope of the chord tends to ∞. The derivative does not exist (since ∞ is not a real number), but the tangent still exists and is vertical on the graph.

3.1 Basic Properties of Derivatives

We can easily establish that a constant function must have zero derivative, simply by working with the definition:

Proposition 3.2. If $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = c$ for some $c \in \mathbb{R}$, then $f'(a) = 0$ for all $a \in \mathbb{R}$.

The attempt to establish a converse to Proposition 3.2 exposes some unexpected subtleties; we will come back to this after the Mean Value Theorem. Similarly, we can easily establish the derivative of a straight line:

Proposition 3.3. If $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = mx + c$ for some $m, c \in \mathbb{R}$, then $f'(a) = m$ for all $a \in \mathbb{R}$.

The following rules will be familiar from A-level:

Theorem 3.4. Suppose $f, g\colon A \to \mathbb{R}$ are both differentiable at a. Then:
$f + g\colon A \to \mathbb{R}$, $(f+g)(x) = f(x) + g(x)$ is differentiable at a, and $(f+g)'(a) = f'(a) + g'(a)$;
$fg\colon A \to \mathbb{R}$, $(fg)(x) = f(x) \cdot g(x)$ is differentiable at a, and $(fg)'(a) = f'(a)g(a) + f(a)g'(a)$;
provided that $g(x) \neq 0$ for all $x \in A$, $\frac{f}{g}\colon A \to \mathbb{R}$, $\left(\frac{f}{g}\right)(x) = \frac{f(x)}{g(x)}$ is differentiable at a, and $\left(\frac{f}{g}\right)'(a) = \frac{f'(a)g(a) - f(a)g'(a)}{(g(a))^2}$

Theorem 3.5 (Chain Rule[6]). Let $A \subseteq \mathbb{R}$ and $B \subseteq \mathbb{R}$. If $f\colon A \to \mathbb{R}$ is differentiable at a, and $g\colon B \to \mathbb{R}$ is differentiable at $f(a)$, where $f(A) \subseteq B$, then the composite function $g \circ f\colon A \to \mathbb{R}$ defined by $(g \circ f)(x) = g(f(x))$ is differentiable at a and $(g \circ f)'(a) = g'(f(a)) \cdot f'(a)$.

Theorem 3.6 (Carathéodory formulation of differentiability). $f\colon (a,b) \to \mathbb{R}$ and $x_0 \in (a,b)$, then f is differentiable at x_0 with derivative $f'(x_0)$ if and only if there exists a function ϕ that is continuous at x_0 and $\phi(x_0) = f'(x_0)$ and $f(x) = f(x_0) + \phi(x)(x - x_0)$

Proof. Supposed f is differentiable at x_0 then define $\phi(x) = \frac{f(x) - f(x_0)}{x - x_0}, x \neq x_0$ and $\phi(x_0) = f'(x_0)$ and so ϕ is continuous at x_0 by definition of differentiability and $\phi(x_0) = f'(x_0)$. Now suppose that there exists ϕ which is continuous at x_0 and $f(x) = f(x_0) + \phi(x)(x - x_0)$ so $\phi(x) = \frac{f(x) - f(x_0)}{x - x_0}$ and the limit of ϕ as $x \to x_0$ exists so f is differentiable. □

From these it is easy to show that:

Proposition 3.7. If $f\colon \mathbb{R} \to \mathbb{R}$, where $f(x) = x^n$ for some natural number n, then $f'(a) = n \cdot a^{n-1}$ for any $a \in \mathbb{R}$.

Theorem 3.4 can be proved using the algebra of limits with some fancy trickery; however, it is much more difficult (though possible) to prove the chain rule in such a manner, and the proof does not generalise to higher dimensions. The problem stems from the lack of a rule for limits of composite functions. However, we know how composite functions behave when they are continuous; so we seek to reduce the problem of differentiability to one of continuity.

One other problem we encounter when trying to generalise differentiation to higher dimensions is that division cannot always be defined. To solve both these problems, we reject the definition of the derivative as the slope of the tangent, and simply use the tangent itself. The most general straight line through the point $(a, f(a))$ is of the form $y = f(a) + k(x - a)$, and this is tangent to the graph of f precisely when $k = f'(a)$. Now, if we let k vary in such a way that $y = f(x)$, then we require that $k(x) \to f'(a)$ as $x \to a$; in other words we need the function k to be continuous. Stating this formally, we have:

Using we can prove the product, quotient and chain rules quite easily[7]. Furthermore, since each of the functions on the right-hand side of the expression $f(x) = f(a) + (x - a)\Delta_a f(x)$ are continuous, it is immediately clear that:

Lemma 3.8. If f is differentiable at a, then f is continuous at a.

The converse is *not* true. Some examples follow:

- Consider the absolute value function $f\colon \mathbb{R} \to \mathbb{R}$ given by $f(x) = |x|$. Since for $x \geq 0$ we have $f(x) = x$ and for $x < 0$ we have $f(x) = -x$, f is clearly continuous and differentiable at any point except 0; however, f is continuous *but not differentiable* at $x = 0$.
- We saw earlier that the function $f(x) = \sin \frac{1}{x}$ for $x > 0$ could not be extended to a continuous function on $x \geq 0$ no matter what we defined $f(0)$ as. It follows that f cannot be differentiable at $x = 0$, by the contrapositive of Lemma 3.8. However, $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x \sin \frac{1}{x}$ for $x \neq 0$ with $f(0) = 0$ *is* continuous at $x = 0$, but it fails to be differentiable. On the other hand, $f\colon \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2 \sin \frac{1}{x}$ is continuous *and* differentiable at $x = 0$.

Definitions 3.9. A function $f\colon (a,b) \to \mathbb{R}$ is *differentiable* if it is differentiable at every $c \in (a,b)$. A function $f\colon [a,b] \to \mathbb{R}$ is differentiable if, in addition, the limits $\lim_{x \to a^+} \frac{f(x) - f(a)}{x - a}$ and $\lim_{x \to b^-} \frac{f(x) - f(b)}{x - b}$ exist as real numbers; we call these the right derivative of f at a and the left derivative of f at b, respectively.

Definitions 3.10. If a function $f\colon A \to \mathbb{R}$ is differentiable, we define the derived function $f'\colon A \to \mathbb{R}$. If the derived function is differentiable at a point $a \in A$, we say that f is *twice differentiable*, and denote $(f')'(a)$ as $f''(a)$. Similarly, we say f is *n times differentiable* at a if $f^{(n)}(a)$ exists, where we define inductively by $f^{(n)}(a) = (f^{(n-1)})'(a)$ and $f^{(0)}(a) = f(a)$.

If $f\colon A \to \mathbb{R}$ is n times differentiable, and $f^{(n)}\colon A \to \mathbb{R}$ is continuous, then we say that f is $C^n(A)$. If f is n times differentiable for all n, then we say f is $C^\infty(A)$.

[6] In Leibniz notation, the chain rule says that $\frac{dg}{dx}(x) = \frac{dg}{df}(f(x)) \cdot \frac{df}{dx}(x)$.

[7] These proofs are fairly simple, involving mainly algebraic manipulation, and are examinable.

3.2 Derivatives and Completeness

We now consider maxima and minima of functions.

Definition 3.11. For $f\colon A \to \mathbb{R}$, $a \in A$ is a *local maximum* of f if $\exists\, \delta > 0$ s.t. $|x - a| < \delta \implies f(x) \leq f(a)$. Similarly, $a \in A$ is a *local minimum* of f if $\exists\, \delta > 0$ s.t. $|x - a| < \delta \implies f(x) \geq f(a)$.

You may be used to finding local maxima or minima by setting $f'(x) = 0$ and solving for x; however, $f'(x) = 0$ is a necessary condition for x to be a maximum or minimum, but not sufficient:

Lemma 3.12. If $f\colon A \to \mathbb{R}$ has a local maximum or local minimum at some $a \in A$, and f is differentiable at a, then $f'(a) = 0$.

To better understand the situation, we turn to more general questions. The following powerful theorem *seems* obvious at first glance; however, it relies on the completeness of $\mathbb{R}$ and is hence non-trivial

Theorem 3.13 (Rolle's Theorem)**.** If $f\colon [a, b] \to \mathbb{R}$ is continuous on $[a, b]$ and differentiable on (a, b), and additionally $f(a) = f(b)$, then there is a point $c \in (a, b)$ such that $f'(c) = 0$.

Proof. As f is continuous, it is bounded and attains its bounds, i.e. there exist $x_1, x_2 \in [a, b]$ such that $f(x_1) \leq f(x) \leq f(x_2)$ for all $x \in [a, b]$. Set $A = f(a) = f(b)$. If $f(x_1) < A$, then $x_1 \in (a, b)$ is a local minimum, and hence $f'(x_1) = 0$. If $f(x_2) > A$, then $x_2 \in (a, b)$ is a local maximum, and hence $f'(x_2) = 0$. Otherwise $f(x_1) = f(x_2) = A$ and hence f is constant, so $f'(x) = 0$ for all $x \in (a, b)$. □

Notice that the interval must be closed in order to use Extreme Value Theorem. Geometrically, Rolle's Theorem states that there is a tangent to the curve $y = f(x)$, at some point between $x = a$ and $x = b$, which is parallel to the chord joining the points $(a, f(a))$ and $(b, f(b))$. In this theorem, the chord is horizontal; we now generalise this to allow the chord to have any gradient.

Theorem 3.14 (The Mean Value Theorem)**.** If $f\colon [a, b] \to \mathbb{R}$ is continuous on $[a, b]$ and 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. Consider the line joining the points $(a, f(a))$ and $(b, f(b))$, whose gradient is $m = \frac{f(b) - f(a)}{b - a}$. This is given by $k\colon [a, b] \to \mathbb{R}$ where $k(x) = f(a) + \frac{f(b) - f(a)}{b - a}(x - a)$. Applying Rolle's theorem to $g(x) = f(x) - k(x)$, which has $g(a) = g(b) = 0$, yields some $c \in (a, b)$ with $g'(c) = 0$. But $k'(x) = \frac{f(b) - f(a)}{b - a}$ for all $x \in (a, b)$, so $g'(c) = f'(c) - \frac{f(b) - f(a)}{b - a} = 0$, i.e. $f'(c) = \frac{f(b) - f(a)}{b - a}$ as required. □

Corollary 3.15. If $f\colon [a, b] \to \mathbb{R}$ is continuous on $[a, b]$ and differentiable on (a, b), then
- if $f'(x) = 0$ for all $x \in (a, b)$, then f is constant on $[a, b]$;
- if $f'(x) > 0$ for all $x \in (a, b)$, then f is strictly increasing on $[a, b]$;
- if $f'(x) < 0$ for all $x \in (a, b)$, then f is strictly decreasing on $[a, b]$.

In higher dimensions, the Mean Value Theorem does not hold, but often the Mean Value *Inequality* does hold:

Proposition 3.16 (Mean Value Inequality)**.** Let $f\colon [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b). If there exists a real number K such that $|f'(x)| \leq K$, then for all $x, y \in [a, b]$ we have $|f(x) - f(y)| \leq K|x - y|$.

This says that if the derivative is bounded, f is Lipschitz. (A function $f\colon A \to \mathbb{R}$ is Lipschitz if there exists $K > 0$ such that $|f(x) - f(y)| \leq K|x - y|$; a Lipschitz function is necessarily continuous.) In particular, this occurs when $f'\colon [a, b] \to \mathbb{R}$ is continuous, since it is then automatically bounded. The Mean Value Inequality converts local information into global information: $|f'(x)| \leq K$ tells us that the *local* rate of change is no greater than K, and then $|f(x) - f(y)| \leq K|x - y|$ says that the rate of change between *any* two points in the domain is no greater than K.

When f is invertible, we can relate the derivatives of f and f^{-1} as follows[8]:

[8] In Leibniz notation, the inverse rule says that $\frac{dx}{dy}(y) = 1/\frac{dy}{dx}(x)$, where $y(x) = f(x)$.

We turn now to Taylor's theorem. When we write the Mean Value Theorem in the form $f(b) = f(a) + (b - a)f'(c)$, for some $c \in (a, b)$, we can use the term $(b - a)f'(c)$ to give us an estimate of how near the value of f at b is to its value at a. This is a linear approximation to f around the point a. What if we try to approximate it by a polynomial? Well, let $p_n(x) = f(a) + k_1(x - a) + \cdots + k_n(x - a)^n$, and suppose $f(x) = p_n(x) + r_n(x)$, where $r_n(x)$ is a remainder term for the n^{th}-order approximation. If we choose the k_i such that f and p have the same derivatives of orders 1 to n at $x = a$, then we get $p_n^{(j)}(a) = j!k_j = f^{(j)}(a)$ for $1 \leq j \leq n$, and hence

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

Does this approximate f well? In order to know that, we need to know what form the remainder term, $r_n(x)$ takes. Taylor's theorem states that the remainder term for the approximation by an n^{th}-order polynomial depends on $f^{(n+1)}(c)$ for some $c \in (a, b)$.

Theorem 3.17 (Cauchy's Mean Value Theorem)**.** Suppose $f, g \colon [a, b] \to \mathbb{R}$ are continuous on $[a, b]$ and differentiable on (a, b), and $g'(x) \neq 0$ for all $x \in (a, b)$. Then there exists $c \in (a, b)$ for which $\frac{f(b)-f(a)}{g(b)-g(a)} = \frac{f'(c)}{g'(c)}$.

Theorem 3.18 (Taylor's Theorem)**.** If $f \colon [a, b] \to \mathbb{R}$ has a continuous n^{th} derivative on $[a, b]$ and is $n+1$ times differentiable on (a, b), then there is a point $c \in (a, b)$ such that

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

By putting $a = 0$, $b = x$ and $c = \theta x$ into Taylor's theorem, for some $\theta \in (0, 1)$, we get the special case known as ***Maclaurin's Theorem***. The final term, known as the Lagrange form of the remainder, is most like that in the Mean Value Theorem; a term dependent on a derivative of f at some point $c \in (a, b)$. As we do not, in general, know what the point c is, this does not always give useful information, so we consider an alternative formulation of Taylor's Theorem, with Cauchy's form of the remainder:

Theorem 3.19 (Taylor's Theorem with Cauchy's form of the remainder)**.** If $f \colon [a, b] \to \mathbb{R}$ has a continuous n^{th} derivative on $[a, b]$ and is $n + 1$ times differentiable on (a, b), then there is a point $c \in (a, b)$ such that

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

Proposition 3.20. Let $A \subseteq \mathbb{R}$ and $B \subseteq \mathbb{R}$. For a bijection $f \colon A \to B$, with inverse $f^{-1} \colon B \to A$, if f is differentiable at a and f^{-1} is differentiable at $f(a)$ then $(f^{-1})'(f(a)) = \frac{1}{f'(a)}$.

Note that this means that if f and f^{-1} are inverse functions, then if $f'(a) = 0$ then f^{-1} cannot be differentiable at $f(a)$. However, if the domain of f is an interval, then the differentiability of f at a, along with the requirement that $f'(a) \neq 0$, is sufficient to show that f^{-1} is differentiable at the point $f(a)$.

Proposition 3.21. Suppose $f \colon (a, b) \to (h, k)$ is a continuous bijection with inverse $f^{-1} \colon (h, k) \to (a, b)$. Then if f is differentiable at $x \in (a, b)$ with $f'(x) \neq 0$, then f^{-1} is differentiable at $f(x)$.

Using this, we get the Inverse Function Theorem[9]:

Theorem 3.22 (The Inverse Function Theorem)**.** If $f \colon (a, b) \to \mathbb{R}$ is continuous and differentiable on (a, b) with $f'(x) > 0$ for all $x \in (a, b)$, then $f \colon (a, b) \to (h, k)$ is a bijection with inverse $f^{-1} \colon (h, k) \to (a, b)$ which is continuous and differentiable on (h, k) with $(f^{-1})'(y) = \frac{1}{f'(f^{-1}(y))}$ for all $y \in (h, k)$.

[9] This is not the most general form of the Inverse Function Theorem; the lectures included a slightly stronger statement. However, this is good enough for all intents and purposes.

4 Power Series

So far, the only interesting functions we have considered have been polynomials, $f(x) = x^n$, and what can be made by combining finitely many through addition, multiplication, or even division. To obtain functions other than rational functions, we must remove the restriction to a finite number of operations; in other words, we must admit limiting processes.

In the previous section, we considered polynomial approximations to differentiable functions using Taylor's Theorem. In each case we had a finite polynomial of the form $\sum_{k=0}^{n} a_k x^k$, and the difference between the function and the polynomial was given by a remainder term depending on the values of $f^{(n+1)}$. We now ask the question: what if these remainder terms tend to 0 as $n \to \infty$? Do we get an "infinite polynomial" representation of the function?

We must first consider what we mean by an "infinite polynomial". One likely candidate is a series of the form $\sum_{n=0}^{\infty} a_n x^n$, where the a_n are fixed coefficients; this is known as a *power series*. Power series have many beautiful and interesting properties which make using them in analysis very useful indeed.

4.1 Differentiability and Taylor Series

We begin with a few examples:

Examples 4.1.

- Lemma 1.9 from Analysis I tells us that the power series $\sum x^n$ converges for $|x| < 1$ and diverges for $|x| \geq 1$.
- Consider the power series $\sum n x^n$. By the ratio test, this converges if $|x| < 1$ and diverges if $|x| > 1$.
- Consider the power series $\sum \frac{x^n}{n!}$. By the ratio test, this converges for any value of x.
- Consider the power series $\sum n! x^n$. By the ratio test, this diverges for every $x \neq 0$, but converges for $x = 0$.

This motivates our definition of radius of convergence:

Theorem 4.2. For any power series $\sum a_n x^n$, either:

1. the series converges absolutely for all $x \in \mathbb{R}$; or
2. there is a real number R, such that the series is absolutely convergent for $|x| < R$ and divergent for $|x| > R$; or
3. the series converges only if $x = 0$.

We call R the *radius of convergence* of the power series; in the first case we write $R = \infty$, and in the third case we set $R = 0$.

We can use the ratio test (although most of the time Cauchy's n^{th} Root is better) to find the radius of convergence as in the above examples; again this is popular in exam questions. However, note that in the second case, we know nothing about the convergence of $\sum a_n x^n$ on the boundary where $|x| = R$.

What is even more beautiful about power series is that they are guaranteed to be differentiable within their radius of convergence, and that the derivative of a function defined by a power series is simply the power series obtained by differentiating each term:

Lemma 4.3. The power series $\sum a_n x^n$, $\sum n a_n x^{n-1}$ and $\sum \frac{a_n x^{n+1}}{n+1}$ have the same radius of convergence.

Theorem 4.4 (Term by Term Differentiation)**.** Suppose $\sum a_n x^n$ has radius of convergence R with $R > 0$ (or $R = \infty$), and define $f\colon (-R, R) \to \mathbb{R}$ by $f(x) = \sum_{n=0}^{\infty} a_n x^n$. Then f is differentiable and $f'(x) = \sum_{n=1}^{\infty} n a_n x^{n-1}$ for $x \in (-R, R)$.

This is a powerful theorem: given a power series that is convergent on $(-R, R)$, we know automatically that it is differentiable, and its derivative is given by differentiating the power series termwise, as we would a polynomial. But if we differentiate a power series, we get another power series, which we can differentiate again and again; so power series can be differentiated *arbitrarily many times*! Furthermore, as differentiability implies continuity, not only is the power series "infinitely" differentiable, each derivative (not to mention the power series itself) is continuous. Power series really are *very* nice functions.

By applying theorem 4.4 to $f(x) = \sum a_n x^n$, and setting $x = 0$, we see that $a_k = \frac{1}{k!} f^{(k)}(0)$. Hence if f is equal to some power series, then we have

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

This looks very much like Taylor's theorem, but with infinitely many terms; we call it the *Taylor series* of f about 0. If we can differentiate f arbitrarily many times, i.e. if f is C^{∞}, then can we say that f is equal to its Taylor series? Many great mathematicians, such as Lagrange, thought so. But ***the answer is NO!***

Firstly, the Taylor series may not converge; and even if it does, it may not converge to the function. The starkest example is due to Cauchy: consider $f \colon \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) := \begin{cases} e^{-1/x^2} & \text{for } x \neq 0 \\ 0 & \text{for } x = 0 \end{cases}$$

Then f possesses derivatives of all orders at every point, but $f^{(n)}(0) = 0$ for all n, so its Taylor series about 0 is zero everywhere! Where does this example break down? Simple: we have neglected the statement of Taylor's theorem. This states that f is equal to a polynomial *plus a remainder term!* In this example, the remainder terms taken about $x = 0$ do not converge to 0 for any $x \neq 0$.

Definition 4.5. The function $f \colon A \to \mathbb{R}$ is called *analytic* if, for all $a \in A$, there is a radius $\delta > 0$ such that $f(x) = \sum_{n=0}^{\infty} a_n (x - a)^n$ for all x with $|x - a| < \delta$.

This says that if f is equals its Taylor series about *any* point in its domain, then f is analytic. Any analytic function is automatically C^{∞}, but we have seen that not all C^{∞} functions are analytic.

4.2 Upper and Lower Limits

Theorem 4.2 is beautiful: a power series must converge for every $x \in (-R, R)$ for some $R \in [0, \infty]$. Unfortunately, however, it does not tell us *how* to find the radius of convergence R. As we did in the examples, we can use the ratio test:

Proposition 4.6. For a power series $\sum c_n x^n$, suppose that $\left| \frac{c_{n+1}}{c_n} \right| \to k$ for some $0 < k < \infty$. Then $\sum c_n x^n$ is convergent if $|x| < \frac{1}{k}$ and divergent if $|x| > \frac{1}{k}$.

However, this limit may not exist. Hence, we search for an explicit expression for the radius of convergence which always exists. Consider the sequence $(0.9, 3.1, 0.99, 3.01, 0.999, 3.001, \dots)$. This does not have a limit, but we can find a subsequence, $(0.9, 0.99, 0.999, \dots)$ that approaches 1, and another subsequence $(3.1, 3.01, 3.001, \dots)$ which approaches 3. No subsequence approaches anything larger than 3. We say that 3 is the upper limit, or lim sup, of this sequence, and that the lower limit, or lim inf, is 1. It should be noted that 3 is not the least upper bound of the elements in this sequence. There are infinitely many terms that are strictly greater than 3. But if we take any positive ε, then eventually all of the terms will fall below $3 + \varepsilon$ and stay below. We thus define $\limsup a_n$:

Definition 4.7. The *upper limit* (or "limit superior") of a sequence (a_n), denoted $\limsup a_n$, is defined as $\limsup_{n \to \infty} a_n := \lim_{k \to \infty} \sup_{n \geq k} a_n$. Similarly, the *lower limit* (or "limit inferior") of a sequence (a_n), denoted $\liminf a_n$, is defined as $\liminf_{n \to \infty} a_n := \lim_{k \to \infty} \inf_{n \geq k} a_n$.

By definition, $(a_n) \to a$ if and only if, for each $\varepsilon > 0$ there is a number N such that $n \geq N$ implies $|a_n - a| < \varepsilon$. This can be characterised by saying that, for any $\varepsilon > 0$, $a - \varepsilon < a_n < a + \varepsilon$ for *all but finitely many* n. This is our working definition of the limit; but our definitions of the upper and lower limits bear no relation to this. However, for the lim sup and lim inf, one of the inequalities in $a - \varepsilon < a_n < a + \varepsilon$ holds for all but finitely many n, but the other holds only for *infinitely many* n.

Proposition 4.8. Given a sequence (a_n) in $\mathbb{R}$, we have $\limsup a_n = \alpha \in \mathbb{R}$ if and only if, for each $\varepsilon > 0$, the following two conditions hold:

(i) for all but finitely many n (that is, for all $n \geq N$ for some $N \in \mathbb{N}$), we have $a_n < \alpha + \varepsilon$; and

(ii) for infinitely many n, we have $a_n > \alpha - \varepsilon$.

Proposition 4.9. Given a sequence (a_n) in $\mathbb{R}$, we have $\liminf a_n = \beta \in \mathbb{R}$ if and only if, for each $\varepsilon > 0$, the following two conditions hold:

(i) for all but finitely many n (that is, for all $n \geq N$ for some $N \in \mathbb{N}$), we have $a_n > \beta - \varepsilon$; and

(ii) for infinitely many n, we have $a_n < \beta + \varepsilon$.

Proposition 4.10. Given a sequence (a_n) in $\mathbb{R}$, $(a_n) \to l$ if and only if $\liminf a_n = \limsup a_n = l$.

There are many uses of lim sup and lim inf; one of these is in analysing convergent sub-sequences Given a bounded sequence, the Bolzano-Weierstrass Theorem guarantees that it contains a convergent subsequence. It says nothing, however, about how many of them there are; if the sequence is convergent, then all sub-sequences converge to the same limit, that of the original sequence. On the other hand, a sequence which oscillates, e.g. $(\sin n)$, will have many convergent sub-sequences, all converging to different points. We call these points ***limit points***:

Definition 4.11. If (a_n) is a sequence in $\mathbb{R}$, we say that $l \in \mathbb{R}$ is a ***limit point*** of (a_n) if there exists a subsequence (a_{n_k}) with $(a_{n_k}) \to l$ as $k \to \infty$.

If the lim sup is the "limit towards which the greatest values converge" (as Cauchy, somewhat heuristically, defined it), then we would expect that the largest value to which a subsequence can converge is the lim sup of the sequence; similarly, we expect the smallest value to be the lim inf. This is indeed the case, and what is more, $\limsup a_n$ and $\liminf a_n$ are themselves limit points:

Proposition 4.12. Suppose (a_n) is a bounded sequence in $\mathbb{R}$. Then $\limsup a_n$ and $\liminf a_n$ are limit points of (a_n), and for any limit point l of (a_n) we have $\liminf a_n \leq l \leq \limsup a_n$.

A general sequence may not always have a limit, but it will always have a lim sup and a lim inf. We put this to good use in Cauchy's n^{th} root test:

Theorem 4.13 (Cauchy's n^{th} Root Test)**.** The series $\sum a_n$ is absolutely convergent if $\limsup |a_n|^{1/n} < 1$ and divergent if $\limsup |a_n|^{1/n} > 1$.

This leads straight away to our desired expression for the radius of convergence of a power series:

Corollary 4.14 (The Cauchy–Hadamard formula)**.** For a power series $\sum a_n x^n$, set $R = \frac{1}{\limsup \sqrt[n]{|a_n|}}$, with the convention that $R = 0$ if $\limsup \sqrt[n]{|a_n|} = +\infty$ and $R = \infty$ if $\limsup \sqrt[n]{|a_n|} = 0$. Then $\sum a_n x^n$ is absolutely convergent if $|x| < R$, and diverges if $|x| > R$.

5 Special functions of analysis

5.1 The Exponential,Logarithm Functions

One of the ultimate applications of mathematical analysis is to solve the problems which present themselves in the natural sciences, engineering, economics and other disciplines. Commonly the step from the experimental data or hypotheses to the conclusions that they can be made to yield lies in solving differential equations. In solving these problems, particularly in physics, a number of functions occur repeatedly which play a vital role in all these disciplines. The most important of these are the exponential and logarithm functions, of which we now strive to develop their principal properties.

Definition 5.1. Define $\exp\colon \mathbb{R} \to \mathbb{R}$ by $\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$.

We must now, of course, show that all its familiar properties can be retrieved from this definition. First of all, we note that $\exp 0 = 1$, since $0^0 = 1$ and $0^n = 0$ for all non zero n; comparing $\exp x$ with $e := \sum_{n=0}^{\infty} \frac{1}{n!}$ we see that $\exp 1 = e$. Using theorem 4.4, we can show that exp is its own derivative:

Proposition 5.2. For all $x \in \mathbb{R}$, $\exp'(x) = \exp(x)$.

Another important fact is the multiplicative formula:

Proposition 5.3. For all $x, y \in \mathbb{R}$, we have $\exp(x)\exp(y) = \exp(x + y)$.

Propositions 5.2 and 5.3 yield between them, in one way or another, just about every property we seek to show of exp. By letting $y = -x$ in proposition 5.3, we get that $\exp(-x)\exp x = \exp 0 = 1$ for all real numbers x; hence $\exp x \neq 0$ for all $x \in \mathbb{R}$, and $\exp(-x) = \frac{1}{\exp x}$.

Lemma 5.4. $\exp x > 0$ for all real numbers x.

Now, since $\exp'(x) = \exp x > 0$ for all $x \in \mathbb{R}$, using the Mean Value Theorem we get that

Proposition 5.5. $\exp\colon \mathbb{R} \to \mathbb{R}$ is strictly increasing (and hence injective).

We now consider some common limits involving exp. First of all, it is clear from the definition of the power series that $\exp x \to \infty$ as $x \to \infty$. So, as $\exp(-x) = \frac{1}{\exp x}$, we have that $\exp x \to 0$ as $x \to -\infty$. Secondly, we consider the ratio of any power of x and the exponential function:

Lemma 5.6. For all $n \in \mathbb{N}$, $\displaystyle\lim_{x\to\infty} \frac{x^n}{\exp x} = 0$.

In other words $\exp x \to \infty$ "faster" than any power of x as $x \to \infty$.

We know that exp is continuous and differentiable on all of $\mathbb{R}$; furthermore, $\exp' x > 0$ for all x. As $\lim_{x\to-\infty} \exp x = 0$ and $\lim_{x\to\infty} \exp x = \infty$, we can apply the Inverse Function Theorem to get that exp is bijective, with inverse $(0, \infty) \to \mathbb{R}$:

Theorem 5.7. $\exp\colon \mathbb{R} \to (0, \infty)$ is bijective with inverse $\log\colon (0, \infty) \to \mathbb{R}$ that is differentiable and $\log'(y) = \frac{1}{y}$ for all $y > 0$.

The usual properties of the logarithm now fall out almost immediately, again using proposition 5.3:

Proposition 5.8. For all $a, b > 0$ and any $r \in \mathbb{Q}$, we have
(i) $\log 1 = 0$;
(ii) $\log a + \log b = \log ab$; and
(iii) $\log(a^r) = r \log a$.

So far we have only defined a^x for rational x; here $a > 0$. By part (iii) of the above proposition, for any rational x we have $a^x = \exp(x \log a)$. In order to extend a^x to any real x, we *define* $a^x := \exp(x \log a)$ for any real x. In the spirit of the section, we prove the following obvious properties of a^x:

Proposition 5.9. For $a > 0$ we have
(i) $a^x > 0$ for all $x \in \mathbb{R}$;
(ii) $a^x a^y = a^{x+y}$ for all $x, y \in \mathbb{R}$;
(iii) $(a^x)^y = a^{xy}$ for all $x, y \in \mathbb{R}$.

In particular, since $\exp 1 = e$, we have that $\log e = 1$, and hence that $e^x = \exp x$. We can use this definition of a^x to extend our ability to differentiate x^k to all real numbers k, as well as differentiating a^x with respect to x.

Proposition 5.10. The functions $f, g\colon (0, \infty) \to \mathbb{R}$ defined by $f(x) = a^x$, $a > 0$, and $g(x) = x^k$, $k \in \mathbb{R}$, are differentiable with $f'(x) = a^x \log a$ and $g'(x) = kx^{k-1}$.

Finally, applying Taylor's theorem yields two key results:

Proposition 5.11. For $-1 < x \leq 1$, $\displaystyle\log(1 + x) = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{x^n}{n}$.

Proposition 5.12 (The Binomial Theorem for any real index)**.** For any $a \in \mathbb{R}$ and $-1 < x < 1$,

$$(1 + x)^a = \sum_{n=0}^{\infty} \frac{a(a-1)\dots(a-n+1)}{n!} x^n = 1 + ax + \frac{a(a-1)}{2!} x^2 + \frac{a(a-1)(a-2)}{3!} x^3 + \dots$$

5.2 sine and cosine functions

$s(x) = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{(2k+1)!}$, $c(x) = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k}}{(2k)!}$

Lemma 5.13. $c, s\colon \mathbb{R} \to \mathbb{R}, x \in \mathbb{R}$ then $c^2(x) + s^2(x) = 1$

Lemma 5.14. $c, s\colon \mathbb{R} \to \mathbb{R}$ $a, b \in \mathbb{R}$ then $s(a+b) = s(a)c(b) + s(b)c(a)$ and $c(a+b) = c(a)c(b) - s(a)s(b)$

Lemma 5.15. The function $c\colon \mathbb{R} \to \mathbb{R}$ has its smallest positive root γ between $\sqrt{2}$ and $\sqrt{3}$. Then 2γ is the smallest positive root of s and s(γ) = 1.

In some cases when we are required to prove there exists a root between 2 values a useful method is to group terms in a specific way where each group is greater than or less than zero. This gives the required inequality and we can then use the IVT.

6 L'Hôpital's Rule

The Mean Value Theorem is one of the most powerful theorems in all of analysis, and yet it is relatively simple to understand. It thus comes as something of a surprise that it can lead to such powerful theorems as L'Hôpital's Rule, a means of evaluating limits which would otherwise result in $\frac{0}{0}$ by differentiating both top and bottom without affecting the limit.

The first form of L'Hôpital's Rule is concerned with the evaluation of limits of the form $\lim_{x \to a} \frac{f(x)}{g(x)}$ when $f(a) = g(a) = 0$. The original Mean Value Theorem considered the ratio $\frac{f(b)-f(a)}{b-a}$, where f is a differentiable function. But we now have two functions, both (presumably) differentiable; how do we relate their rates of change? The answer comes in the form of Cauchy's Mean Value Theorem for two functions f and g:

Using this we get the first form of L'Hôpital's Rule, for right limits:

Proposition 6.1 (L'Hôpital's Rule for right limits, case I)**.** Suppose $f, g\colon [a, b] \to \mathbb{R}$ are continuous on $[a, b]$ and differentiable on (a, b), and $g'(x) \neq 0$ for all $x \in (a, b)$. If $f(a) = g(a) = 0$ and $\lim_{x \to a^+} \frac{f'(x)}{g'(x)} = l$ for some $l \in \mathbb{R} \cup \{\pm\infty\}$, then $\lim_{x \to a^+} \frac{f(x)}{g(x)} = \lim_{x \to a^+} \frac{f'(x)}{g'(x)} = l$.

Analogously, we can use left-sided limits; we can also use two-sided limits, giving us the standard form of L'Hôpital's Rule:

Theorem 6.2 (L'Hôpital's Rule, case I)**.** Suppose $f, g\colon [a, b] \to \mathbb{R}$ are differentiable on (a, b), and $g'(x) \neq 0$ for all $x \in (a, b)$. If for some $c \in (a, b)$ we have $f(c) = g(c) = 0$ and $\lim_{x \to c} \frac{f'(x)}{g'(x)} = l$ for some $l \in \mathbb{R} \cup \{\pm\infty\}$, then $\lim_{x \to c} \frac{f(x)}{g(x)} = \lim_{x \to c} \frac{f'(x)}{g'(x)} = l$.

Example 6.3. Suppose we wish to calculate the limit $\lim_{x \to 0} \frac{\sqrt{x+2}-\sqrt{2}}{\sqrt{x+1}-1}$. If we let $f(x) = \sqrt{x+2} - \sqrt{2}$ and $g(x) = \sqrt{x+1} - 1$ then we see that $f(0) = g(0) = 0$. We compute $f'(x) = \frac{1}{2\sqrt{x+2}}$ and $g'(x) = \frac{1}{2\sqrt{x+1}}$; then $g'(x) \neq 0$, so

$$\lim_{x \to 0} \frac{f'(x)}{g'(x)} = \lim_{x \to 0} \frac{\frac{1}{2\sqrt{x+2}}}{\frac{1}{2\sqrt{x+1}}} = \lim_{x \to 0} \frac{\sqrt{x+1}}{\sqrt{x+2}} = \frac{1}{\sqrt{2}}.$$

So $\lim_{x \to 0} \frac{f'(x)}{g'(x)} = \frac{1}{\sqrt{2}}$, and hence by L'Hôpital's Rule we have $\lim_{x \to 0} \frac{\sqrt{x+2}-\sqrt{2}}{\sqrt{x+1}-1} = \frac{1}{\sqrt{2}}$.

So far, L'Hôpital's Rule has dealt with limits of the form $\lim_{x \to a} \frac{f(x)}{g(x)}$ in which $f(x), g(x) \to 0$ as $x \to a$. However, if $g(x) \to \infty$ as $x \to a$, then $\lim_{x \to a} \frac{f(x)}{g(x)}$ will equal 0 if f is finite, but if also $f(x) \to \infty$ as $x \to a$ we have an indeterminate form. For this reason, we study Case II of L'Hôpital's Rule:

Proposition 6.4 (L'Hôpital's Rule for right limits, case II)**.** Suppose $f, g\colon (a, b) \to \mathbb{R}$ are differentiable, and that $g(x) \neq 0$ and $g'(x) \neq 0$ for all $x \in (a, b)$. If $\lim_{x \to a^+} g(x) = +\infty$ and $\lim_{x \to a^+} \frac{f'(x)}{g'(x)} = l$ for some $l \in \mathbb{R} \cup \{\pm\infty\}$, then $\lim_{x \to a^+} \frac{f(x)}{g(x)} = \lim_{x \to a^+} \frac{f'(x)}{g'(x)} = l$.

Analogously, we can use left-sided limits; we can also use two-sided limits:

Theorem 6.5 (L'Hôpital's Rule, case II)**.** Suppose $f, g\colon [a,b] \setminus \{c\} \to \mathbb{R}$ are differentiable on (a,c) and (c,b), and $g'(x) \neq 0$ for all $x \in (a,b)$. If for $c \in (a,b)$ we have $\lim_{x \to c} g(x) = +\infty$ and $\lim_{x \to c} \frac{f'(x)}{g'(x)} = l$ for some $l \in \mathbb{R} \cup \{\pm\infty\}$, then $\lim_{x \to c} \frac{f(x)}{g(x)} = \lim_{x \to c} \frac{f'(x)}{g'(x)} = l$.

The evaluation of limits using L'Hôpital's Rule is a very useful application of the Mean Value Theorem, and is popular in exam questions.

This guide would not be possible without our wonderful sponsors: