PRELIMS: ANALYSIS II

CONTINUITY AND DIFFERENTIABILITY

Hilary Term 2018

H.A. Priestley

With acknowledgements to the lecturers who have previously shaped this course and who produced online notes for it, on which the present notes have drawn.

Contents

0. Sequences: basic toolkit from Analysis I

0.1. Standard notation.

- $\mathbb{C}$: set of all complex numbers (the complex plane);
- $\mathbb{R}$: set of all real numbers (the real line);
- $\mathbb{Q}$: the rational numbers;
- $\mathbb{N}$: the natural numbers, $1, 2, \ldots$.

$$\mathbb{N} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}.$$

Infinity (∞) and negative infinity ($-\infty$) are a convenient device for expressing certain notions concerning real numbers. *They are not themselves real numbers.*

Intervals.

Given $a, b \in \mathbb{R}$, we define intervals as follows:

$$\begin{aligned} (a,b) &:= \{\, x \in \mathbb{R} \mid a < x < b \,\} \quad (\text{for } a < b), \\ [a,b] &:= \{\, x \in \mathbb{R} \mid a \leqslant x \leqslant b \,\} \quad (\text{for } a \leqslant b), \\ (-\infty, a) &:= \{\, x \in \mathbb{R} \mid x < a \,\}, \\ (a, \infty) &:= \{\, x \in \mathbb{R} \mid x > a \,\}, \end{aligned}$$

and so on.

Quantifiers.

- $\forall$: "for all" or"for every" or "whenever".
- $\exists$: "there exist(s)" or "there is (are)".

Quantifiers matter. Treat them with care and respect. As we shall see later, the order in which quantifiers are written down is important. Good discipline is to put quantifiers at the front of a statement, not at the back an afterthought, and to read carefully from left to right. This is particularly helpful when proving results by contradiction. See 0.4.

0.2. Basic assumptions about the real numbers: arithmetic and order.

The real numbers, with their usual arithmetic operations and usual order, form an **ordered field**. See Analysis I notes for the formal details.

Key facts about modulus and inequalities:

(i) **Triangle Inequality**: $|a \pm b| \leqslant |a| + |b|$ (for all $a, b \in \mathbb{R}$.
(ii) **Reverse Triangle Inequality**: $|a \pm b| \geqslant ||a| - |b||$ (for all $a, b \in \mathbb{R}$.
(iii) For $a, x \in \mathbb{R}$ and $b > 0$,

$$|x - a| < b \iff a - b < x < a + b.$$

0.3. Boundedness properties and the Completeness Axiom.

Recall that a subset S of $\mathbb{R}$ is **bounded above** there exists $b \in \mathbb{R}$ such that $x \leqslant b$ for all $x \in S$ and **bounded below** if there exists $c \in \mathbb{R}$ such that $c \leqslant x$ for all $x \in S$. A set S is **bounded** if it is bounded above and below; This happens if and only if there exists $M \in \mathbb{R}$ such that $|x| \leqslant M$ for all $x \in S$.

We assume that $\mathbb{R}$ satisfies the **Completeness Axiom**:

> *A non-empty subset S of $\mathbb{R}$ which is bounded above has a least upper bound, also called the* **supremum** *of S. This is unique and is denoted* $\sup S$.

The Completeness Axion can equivalently be formulated as the assertion that non-empty subset of $\mathbb{R}$ which is bounded below has a greatest lower bound, or **infimum**.

The Completeness Axiom underpins the deeper results in Analysis I and the same is true in Analysis II.

The brief summary of results from Analysis I that follows serves three purposes:

(1) we shall use many of these results directly;
(2) we shall want to prove analogues of these results for limits of functions, and in some cases can, if we so choose, exploit the corresponding sequences results to do this;
(3) important functions we shall wish to study in the context of Analysis II are defined, or can be obtained, a point at a time as the limit of a sequence of functions or as the sum of a series of functions (see 0.11).

0.4. **Limits of sequences.**

A sequence (x_n) of real numbers **converges to the limit** $L \in \mathbb{R}$ if

$$\forall \varepsilon > 0\ \exists N \in \mathbb{N}\ \forall n > N\ |x_n - L| < \varepsilon.$$

As usual, we then write $x_n \to L$ as $n \to \infty$ or $\lim_{n\to\infty} x_n = L$. We say (x_n) **converges** if there exists $L \in \mathbb{R}$ such that $x_n \to L$.

Important fact: the limit of a convergent sequence is unique.

Useful fact: a convergent sequence is bounded.

0.5. **Non-convergent sequences.**

Consider the contrapositive of the convergence definition. A sequence (x_n) fails to converge if :

$$\forall L \in R\ \exists \varepsilon > 0\ \forall N \in \mathbb{N}\ \exists n > N\ |x_n - L| \geqslant \varepsilon.$$

We say the sequence (x_n) **tends to infinity** if

$$\forall M \in \mathbb{R} > 0\ \exists N \in \mathbb{N}\ \forall n > N \quad x_n > M.$$

We then use the (notation: $x_n \to \infty$. [Aside: this is compatible with our usage $n \to \infty$ in the convergence definition. Remember the Archimedean Property of the reals.]

0.6. **The Algebra of Limits; limits and inequalities.**

This is bread-and-butter stuff for working with sequences. Just brief reminders here. See Analysis I notes for the statements of the results.

(AOL), for sequences which converge: applies to addition, scalar multiplication, product, reciprocal (of a sequence with a non-zero limit), and hence also to linear combinations and suitable quotients.

Remember (AOL) results do not extend in general to sequences which tend to $\pm\infty$.

Limits and inequalities: for *real* sequences, preservation of *weak* inequalities; sandwiching.

Now for the Big Theorems.

0.7. **BOLZANO–WEIERSTRASS THEOREM.** *A bounded sequence of real numbers has a convergent subsequence.*

0.8. **CAUCHY CONVERGENCE PRINCIPLE.**

A sequence (x_n) of real numbers is a **Cauchy sequence** if

$$\forall \varepsilon > 0\ \exists N \in \mathbb{N}\ \forall n, m > N \quad |x_n - x_m| < \varepsilon.$$

Theorem. *A sequence (x_n) of real numbers converges if and only if it is a Cauchy sequence.*

0.9. **Rationals and irrationals.**

We shall often want to make use of the following density results:

(a) *Given $x \in \mathbb{R}$, there exists a sequence (r_n) of rational numbers such that $r_n \to x$.*
(b) *Given $x \in \mathbb{R}$, there exists a sequence (q_n) of irrational numbers such that $q_n \to x$.*

0.10. **Complex sequences.**

Much of the theory of convergence of real sequences extends, with minor adaptation, to sequences of complex numbers. Analysis II is predominantly a course about real-valued functions defined on subsets of $\mathbb{R}$ so we have little need to make use of complex sequences. We just issue a warning here that inequalities between complex numbers do not make sense and that neither does the notion of a complex sequence tending to infinity.

0.11. **Sequences and series of functions.**

The most important class of functions which fall under this heading are the **power series**, to which you received an introduction at the end of Analysis I. Far down the track in Analysis II, we shall develop the tools one needs to work with such functions, as functions defined over an interval and not just point by point. To this end we are led to study the notion of **uniform convergence**.

Meantime, this is an alert that you will later need to brush up on such things as limits of well-known sequences, and simple tests for convergence of series and the calculation of the radius of convergence of a power series. In addition, having a catalogue of useful inequalities at your fingertips is worthwhile (facts on relative magnitudes, like $|\sin x| \leqslant |x|$, $\log(1 + x) \leqslant x$ for $x \geqslant 1$, and so on). If you don't already have an amalgamated list of inequalities involving everyday functions, now would be a good time to put one together for future reference.

1. Functions, and limits of functions

1.1. **Functions.**

Analysis II is a course about **functions**. Nearly always a function for us will be a function $f\colon X \to Y$, where X and Y are subsets of $\mathbb{R}$. This means that f assigns to each element x of X (the **domain** of f) a unique element $f(x)$ of Y. The **image** of f is $\{\, f(x) \mid x \in X\,\}$. This is a subset, possibly a proper subset, of Y.

There is no expectation here that the mapping $x \mapsto f(x)$ has to be specified by a single formula, or even a formula at all. Specification of a function 'by cases' will be common in this course. The modulus function is just one example of this.

Thus we shall allow our examples to include functions like the following:

(1) $f(x) = \begin{cases} \cos(\sin(1/x)) & x > 0, \\ 0 & x \leqslant 0. \end{cases}$

(2) $f(x) = \sqrt{1-x^2}$ if $x \in [-1,1]$.

(3) $f(x) = \arctan x$ on $(-\pi/2, \pi/2)$.

(4) f defined on $(0,1]$ by

$$f(x) = \begin{cases} \frac{1}{p+q} & \text{if } x \text{ is rational, } x = p/q \text{ in lowest terms,} \\ 0 & \text{otherwise.} \end{cases}$$

(5) $f(x) = x^2 \mathrm{e}^{-x^{2018}}$ on $\mathbb{R}$.

(6) $f(x) = \displaystyle\sum_{n=0}^{\infty} \frac{2^n}{(n!)^3} x^n$ on $\mathbb{R}$.

(7) $f(x) = \left(! + \dfrac{1}{x} \right)^x \quad$ for $x \neq 0$.

And so on and so on . . . .

We want to encompass the familiar functions of everyday mathematics: polynomials; exponential functions; trigonometric functions; hyperbolic functions—all of which can be defined on the whole of $\mathbb{R}$. We shall also encounter associated inverse functions, logarithm; arcsin, etc. You will know from Analysis I that many of these functions can be *defined* using power series. One of our objectives in Analysis II will be to develop properties of functions defined by power series (continuity, differentiability, useful inequalities and limits, . . .). But until our general theory of functions has been developed far enough to cover this material *we shall make use of the properties we need, as unpaid debts we shall later discharge.*

The material in this section is unashamedly technical, but necessary if we are to build firm foundations for the study of real-valued functions defined on subsets of $\mathbb{R}$, many of them having graphs neither you nor Matlab can hope to sketch effectively.

We want to define what is meant by the limit of a function. Intuitively f has a limit L at the point p if the values of $f(x)$ are close to L when x is close to (but not equal to) p. But for the definition of limit to be meaningful it is necessary that f is defined at 'enough' points close to p. So we are interested only in points p that x can get close to, where x is in the domain of f. This leads us to the definition of a limit point.

1.2. **Definition (limit point).**

Let $E \subseteq \mathbb{R}$. A point $p \in \mathbb{R}$ is called a **limit point** (or an **accumulation point**) of E if

$\forall \varepsilon > 0$ [illegible] $\in E \quad 0 < |x - p| < \varepsilon.$

Note that here p may be in E, but need not be in E.

The condition $0 < |x - p|$ is important in the case that $p \in E$. It says exactly that $x \neq p$. This restriction ensures that a point p in E only qualifies as a limit point if there are points of E arbitrarily close to p, *other than p itself.*

There is a useful equivalent definition of limit point in terms of limits of sequences.

1.3. **Proposition (limit points characterised in terms of sequences).** *A point $p \in \mathbb{R}$ is a limit point of $E \subseteq \mathbb{R}$ if and only if there exists a sequence (p_n) in E such that $\lim_{n\to\infty} p_n = p$ and $p_n \neq p$, $\forall n \in \mathbb{N}$.*

Proof. If p is a limit point of E, for any $n \in \mathbb{N}$, choose $\varepsilon = 1/n$. Thus there exists $p_n \neq p$ such that $|p_n - p| < 1/n$ and $p_n \in E$. Hence $p_n \to p$ as $n \to \infty$, as required.

Conversely, given $\varepsilon > 0$, $\exists N \in \mathbb{N}$ such that $\forall n > N$, $|p_n - p| < \varepsilon$. So in particular $|p_{N+1} - p| < \varepsilon$ and $p_{N+1} \neq p$. □

1.4. **Simple examples of limit points.**

(i) Let $a < b$ in $\mathbb{R}$ and let $E = (a, b]$. Then p is a limit point of E if and only if $p \in [a, b]$. To prove this, there are three cases to consider (by Trichotomy): $p < a$, $a \leqslant p \leqslant b$ and $p > b$. In the first case take $\varepsilon := (a - p)/2$ and get a contradiction, in the third take $\varepsilon := (p - b)/2$, similarly. If $p \in [a, b)$, given $\varepsilon > 0$ choose $x = p + \frac{1}{2}\min\{\varepsilon, (b - p)\}$. The case $p = b$ is similar.

The same conclusion holds when $E = (a, b)$, $E = [a, b)$ or $E = [a, b]$.

Intervals of the form $[a, b]$ $(a, b \in \mathbb{R})$ will feature in many of our important theorems. They have the important property of being **closed**: they contain all their limit points.

(ii) Let $E = [0, 1] \cup \{2\}$. Here p is a limit point of E if and only if $p \in [0, 1]$. The 'isolated point' 2 is not a limit point.

(iii) Let $E = \mathbb{Q}$. Recall (see 0.9) that every irrational number is the limit of a sequence of rational numbers. Proposition 1.3 implies that every $p \in \mathbb{R} \setminus \mathbb{Q}$ is a limit point of E. Also every $p \in \mathbb{Q}$ is the limit of a sequence of rationals $\neq p$: take $p_n = p + n^{-1}$, for example. Hence p is a limit point. We conclude that every point of $\mathbb{R}$ is a limit point of E.

(iv) Let $E = \mathbb{R} \setminus \mathbb{Q}$. Then every point of $\mathbb{R}$ is a limit point of E. See Problem sheet 1, Q. 1.

The notion of limit point is important well beyond the present course, in which we shall encounter only simple instances of it. Much more exotic examples exist. The structure of the real line is rich, with $\mathbb{R}$ having many subsets which are very complicated (and hard to visualise). Such complexities are important in topology and in the measure theory employed in advanced probability theory, for example.

1.5. **Limits of functions.**

Let $E \subseteq \mathbb{R}$ and $f\colon E \to \mathbb{R}$ be a real-valued function. Let p be a *limit point* of E (note that p is not necessarily in E). Let $L \in \mathbb{R}$. We say that f **tends to** L **as** x **tends to** p if

$$\forall \varepsilon > 0\ \exists \delta > 0\ \forall x \in E\ \big(0 < |x - p| < \delta \implies |f(x) - L| < \varepsilon\big)\ .$$

In symbols we write this as $\lim_{x\to p} f(x) = L$ or $f(x) \to L$ as $x \to p$.

Note that in the definition δ may, and usually will, depend on both p and ε.

1.6. **Important note.**

in the limit definition it may or may not happen that f is defined *at* p. And when $f(p)$ is defined, its value has no influence on whether or not $\lim_{x\to p} f(x)$ exists. Moreover, when the limit L does exist and $f(p)$ is defined, there is no reason to assume that $f(p)$ will equal L. See Example 1.8(1).

We should also be aware that our original motivation for considering function limits stems from differential calculus. The recipe from school calculus of the derivative of f can be cast in the form

$$\frac{\mathrm{d}}{\mathrm{d}x} f(x) := \lim_{\delta x\to 0} \frac{f(x+\delta x) - f(x)}{\delta x},$$

where δx is *non-zero*.

The following result validates our definition. Compare with the corresponding result for sequences and its proof.

1.7. **Technical theorem (uniqueness of function limits).** *Let* $f\colon E\to\mathbb{R}$ *and* p *be a limit point of* E. *If* f *has a limit as* $x\to p$, *then this limit is unique.*

Proof. Suppose $f(x)\to L_1$ and also $f(x)\to L_2$ as $x\to p$, where $L_1\neq L_2$. Then $\frac{1}{2}|L_1-L_2|>0$. We now apply the limit definition with $\varepsilon := \frac{1}{2}L_1 - L_2|$. By definition,

$$\exists\delta_1>0\ \forall x\in E\ \big(0<|x-p|<\delta_1 \implies |f(x)-L_1|<\varepsilon\big),$$
$$\exists\delta_2>0\ \forall x\in E\ \big(0<|x-p|<\delta_2 \implies |f(x)-L_2|<\varepsilon\big).$$

Let $\delta := \min\{\delta_1,\delta_2\}$. Since p is a limit point of E and $\delta>0$, $\exists x_0\in E$ such that $0<|x_0-p|<\delta$. However

$$\begin{aligned} |L_1-L_2| &= |(f(x_0)-L_1)-(f(x_0)-L_2)| && \text{[add and subtract technique]}\\ &\leqslant |f(x_0)-L_1| + |f(x_0)-L_2| && \text{[Triangle Law]}\\ &< \varepsilon+\varepsilon\\ &= |L_1-L_2|, \end{aligned}$$

and we have a contradiction. □

1.8. **Examples (limits of functions).**

We begin with some rather artificial examples, to highlight how the limit definition operates.

(1) Let f be defined on $E=\mathbb{R}\setminus\{0\}$ by $f(x)=42$. Then 0 is a limit point of E. Let $L=42$. Then for $x\neq 0$ we have $|f(x)-L|=0$. So, for any $\varepsilon>0$, we can take $\delta:=1$, say, to get that

$$0<|x-0|<\delta \implies |f(x)-L|<\varepsilon.$$

So $f(x)\to L$ as $x\to 0$.

(2) Let f be defined on $\mathbb{R}$ by

$$f(x) = \begin{cases} x & \text{if } x\in\mathbb{Q},\ x>0,\\ 10^{23} & \text{if } x=0,\\ -x & \text{otherwise.}\end{cases}$$

We claim that $f(x)\to 0$ as $x\to 0$. To prove this, simply note that $|f(x)-0|=|x|<\varepsilon$ if $0<|x-0|<\delta:=\varepsilon$. (Here, following the definition of limit, we omit consideration of $f(0)$, even though f is defined at 0. For $\varepsilon<10^{23}$ there is no $\delta>0$ such that $|f(x)-0|<\varepsilon$ for all x such that $|x-0|<\varepsilon$.)

(3) Let $\alpha > 0$. Consider the function $f(x) = |x|^\alpha \sin \frac{1}{x}$ on the domain $E = \mathbb{R} \setminus \{0\}$. We claim that $f(x) \to 0$ as $x \to 0$.

Since $|\sin \theta| \leqslant 1$ for any $\theta \in \mathbb{R}$, we have $\left|x^\alpha \sin \frac{1}{x}\right| \leqslant |x|^\alpha$ for any $x \neq 0$. Therefore, $\forall \varepsilon > 0$, choose $\delta = \varepsilon^{1/\alpha}$. Then

$$\left|x^\alpha \sin \frac{1}{x} - 0\right| \leqslant |x|^\alpha < \varepsilon \qquad \text{whenever } 0 < |x - 0| < \delta.$$

According to the definition, $|x|^\alpha \sin \frac{1}{x} \to 0$ as $x \to 0$.

(4) Consider the function $f(x) = x^2$ on the domain $E = \mathbb{R}$. Let $a \in \mathbb{R}$. We claim that $f(x) \to a^2$ as $x \to a$.

Note that $|x^2 - a^2| = |x - a||x + a| \leqslant |x - a|(|x| + |a|)$. So we want to get a bound on x. Suppose that $|x - a| < 1$, then

$$|x| = |x - a + a| \leqslant |x - a| + |a| < 1 + |a|.$$

So $\forall \varepsilon > 0$, choose $\delta = \min\{1, \frac{\varepsilon}{1+2|a|}\}$. Then

$$|x^2 - a^2| \leqslant |x - a|((1 + |a|) + |a|) < \varepsilon \text{ whenever } |x - a| < \delta.$$

This example serves to illustrate that going back to first principles to establish the limiting value of a function may be a tedious task. Help will soon be at hand; see 1.13 below.

Notice that all the examples presented so far have shown that function limits do exist. Now let's explore how to prove that a limit fails to exist.

1.9. **Limit definition: contrapositive.**

An exercise in quantification: f *doesn't converge* to L as $x \to p$ (that is, either f has no limit or $f(x) \to a \neq L$ as $x \to p$) means that

$$\exists \varepsilon > 0 \ \forall \delta > 0 \ \exists x \in E \ \big(0 < |x - p| < \delta \text{ and } |f(x) - L| \geqslant \varepsilon\big).$$

The proof of the following theorem illustrates how to work with the contrapositive of the limit definition. The theorem translates questions about function limits to questions about sequence limits, and vice versa, and so allows to draw on results from Analysis I. For example we will be able to deduce an Algebra of Limits for functions from (AOL) for sequences. Informally, $f(x) \to L$ as $x \to p$ if and only if f tends to the *same* limit L along *any* sequence in E tending to p.

In the proof some careful bookkeeping is needed to handle the $x \neq p$ condition.

1.10. **Theorem (function limits related to limits of sequences).** *Let $f\colon E \to \mathbb{R}$ where $E \subseteq \mathbb{R}$, p be a limit point of E and $L \in \mathbb{R}$. Then the following two statements are equivalent:*

(a) *$f(x) \to L$ as $x \to p$;*

(b) *For every sequence (p_n) in E such that $p_n \neq p$ and $\lim_{n\to\infty} p_n = p$ then $f(p_n) \to L$ as $n \to \infty$.*

Proof. $\Longrightarrow$ (from the general to the particular): Suppose $\lim_{x\to p} f(x) = L$. Then

$$\forall \varepsilon > 0 \ \exists \delta > 0 \ \forall x \in E \ \big(0 < |x - p| < \delta \implies |f(x) - L| < \varepsilon\big).$$

Now suppose (p_n) is a sequence in E, with $p_n \to p$ and $p_n \neq p$. Then, taking the ε in the convergence of a sequence definition to be δ.

$$\exists N \in \mathbb{N} \ \forall n > N \quad |p_n - p| < \delta.$$

Putting the conditions together and using the fact that $p_n \neq p$, we get

$$\forall \varepsilon > 0 \ \exists N \in \mathbb{N} \ \forall n > N \quad |f(p_n) - L| < \varepsilon.$$

Hence, $\lim_{n\to\infty} f(p_n) = L$.

$\Longleftarrow$ (from the particular to the general): Argue by contradiction. Suppose $\lim_{x\to p} f(x) = L$ is not true. Then, by 1.9, $\exists\varepsilon_0 > 0$ such that 'no δ works'. We can choose δ to be $1/n$ for arbitrary n to show $\exists x_n \in E$, with $0 < |x_n - p| < 1/n$ but

$$|f(x_n) - L| \geqslant \varepsilon_0.$$

Therefore we have found a sequence (x_n) which converges to p but for which $(f(x_n))$ does not tend to L. Contradiction. □

1.11. **Examples: exploiting Theorem 1.10 to prove function limits fail to exist.**

The contrapositive of (a) $\Longrightarrow$ (b) is principally used to prove that a limit $\lim_{x\to p} f(x)$ does not exist, by finding two rival values for the limit, assuming it did exist.

(1) Consider the function f defined in Example 1.8(2). We claim that, for any $p \neq 0$, the limit $\lim_{x\to p} f(x)$ fails to exist.

Assume $p \neq 0$ and $p \in \mathbb{Q}$. Then there exists a sequence (p_n) such that each $0 \neq p_n \neq p$, $p_n \in \mathbb{Q}$ and $p_n \to p$ and there exists a sequence (q_n) such that $q_n \in \mathbb{R}\setminus\mathbb{Q}$ (and so necessarily $q_n \neq p$) and $q_n \to p$. Then

$$f(p_n) = p_n \to p \quad \text{and} \quad f(q_n) = -q_n \to -p.$$

Since $p \neq 0$, the non-existence of $\lim_{x\to p} f(x)$ follows from the contrapositive of (a) $\Longrightarrow$ (b) in Theorem 1.10.

(2) To show that $\lim_{x\to 0} \sin\frac{1}{x}$ doesn't exist.

Let $f(x) = \sin 1/x$ for $x \neq 0$. Our strategy is to construct two sequences (x_n) and (y_n) converging to 0 but such that $(f(x_n))$ and $(f(y_n))$ converge to different limits.

Let $x_n = \dfrac{1}{\pi n}$ and $y_n = \dfrac{2}{\pi(2n+1)}$. Then both sequences (x_n) and (y_n) tend to 0, but

$$\lim_{n\to\infty} \sin\frac{1}{x_n} = 0 \quad \text{and} \quad \lim_{n\to\infty} \sin\frac{1}{y_n} = 1.$$

So $\lim_{x\to 0} \sin\frac{1}{x}$ cannot exist.

Our next task is to set up the basic machinery for working with function limits.

We start with a technical result which is frequently useful. The result parallels a corresponding result for sequences. It can be stated as follows.

1.12. **Non-zero limits.**

Let $g\colon E \to \mathbb{R}$ be a function. Assume that p is a limit point of E and that $\lim_{x\to p} g(x)$ exists and equals B, where $B \neq 0$. Then there exists $\eta > 0$ such that

$$|g(x)| \geqslant \tfrac{1}{2}|B| \qquad \forall x \in E \text{ such that } 0 < |x - p| < \eta.$$

In particular, $g(x) \neq 0$ for all $x \in E$ such that $0 < |x - p| < \eta$.

Proof. Apply the definition of function limit with $\varepsilon := |B|/2$. We can find $\eta > 0$ such that, for $x \in E$,

$$0 < |x - p| < \eta \Longrightarrow |g(x) - B| < \tfrac{1}{2}|B|.$$

Now the Reverse Triangle Inequality gives, for $0 < |x - p| < \eta$,

$$|g(x)| = |(g(x) - B) + B| \geqslant \left|\tfrac{1}{2}|B| - |B|\right| = \tfrac{1}{2}|B|,$$

as claimed. □

1.13. **Theorem (Algebra of Limits for functions).** *Let $E \subseteq \mathbb{R}$ and let p be a limit point of E. Let $f, g\colon E \to \mathbb{R}$, and let $\alpha, \beta \in \mathbb{R}$. Suppose that $f(x) \to A$, $g(x) \to B$ as $x \to p$. Then the following limits exist and have the values stated:*

- **Linear combination***:* $\lim_{x\to p}(\alpha f(x) + \beta g(x)) = \alpha A + \beta B$*;*
- **Product***;* $\lim_{x\to p}(f(x)g(x)) = AB$*;*
- **Reciprocal***: if $B \neq 0$ then $\exists \eta > 0$ s.t. $g(x) \neq 0$ $\forall x \in E$ such that $0 < |x-p| < \eta$, and* $\lim_{x\to p}(1/g(x)) = 1/B$*.*

For quotient, combine the results for product and reciprocal.

Proof. These can all be deduced from the Algebra of Limits of Sequences using Theorem 1.10. Alternatively they can be proved directly from the definitions: mimic the sequence proofs ("for N substitute δ", and remember the need for $0 < |x-p|$).

A sample of proofs to illustrate the techniques:

Proof of linear combinations result, via sequences: Assume (p_n) is any sequence with $p_n \in E$ and $p_n \neq p$. Then, by (a) implies (b) in Theorem 1.10,

$$f(p_n) \to A \text{ and } g(p_n) \to B.$$

By (AOL) for sequences (linear combinations),

$$(\alpha f + \beta g)(p_n) = \alpha f(p_n) + \beta g(p_n) \to \alpha A + \beta B.$$

Now $(\alpha f + \beta g)(x) \to \alpha A + \beta B$ as $x \to p$, by (b) implies (a) in Theorem 1.10.

Proof of product result, directly: Technical issues parallel to those for the corresponding sequences result need to be addressed. First note

$$|f(x)g(x) - AB| \leqslant |f(x)||(g(x) - B)| + |B||(f(x) - A)|.$$

So we need to bound $|f(x)|$ and can do this by taking $\varepsilon = 1$ in the limit definition:

$$\exists \delta_0 > 0 \ \forall x \in E \ \big(0 < |x-p| < \delta_0 \implies f(x) - A| < 1\big).$$

Hence

$$0 < |x-p| < \delta_1 \implies |f(x)| \leqslant |f(x) - A| + |A| < 1 + |A|.$$

Now

$$\forall \varepsilon > 0 \ \exists \delta_1 > 0 \ \forall x \in E \ \big(0 < |x-p| < \delta_1 \implies |f(x) - A| < \frac{\varepsilon}{1 + |A| + |B|}\big),$$
$$\forall \varepsilon > 0 \ \exists \delta_2 > 0 \ \forall x \in E \ \big(0 < |x-p| < \delta_2 \implies |g(x) - B| < \frac{\varepsilon}{1 + |A| + |B|}\big).$$

Thus, taking $\delta = \min\{\delta_0, \delta_1, \delta_2\}$, $\forall x \in E$ such that $0 < |x-p| < \delta$ implies

$$|f(x)g(x) - AB| \leqslant (1 + |A|)|g(x) - B| + |B||f(x) - A| < \varepsilon.$$

This completes the proof. □

Technical note. As is usual in proofs like the previous one, finishing up with $< \varepsilon$ at the end requires foresight earlier on, as when we chose δ_1 and δ_2. Observe that, to prove $h(x) \to L$ as $x \to p$ (where p is a limit point of $E = \operatorname{dom} h$), it suffices to show that for any $\varepsilon > 0$ there exists $\delta > 0$ and a *constant* K such that

$$x \in E \text{ and } 0 < |x-p| < \delta \implies |h(x) - L| < K\varepsilon.$$

Here it is crucial that K be a *finite constant*; it must not depend on x.

Proof of the reciprocal result: The first assertion was proved directly in 1.12. The limit claim can then be obtained either directly or via (AOL) for sequences. Note that we first need to restrict to points x close enough to p for $1/g(x)$ to be defined; then, and only then, can we embark on considering $1/g(x) - 1/L$. Here, and with product too, the final $\delta > 0$ may need to be the minimum of a finite number of positive numbers, each needed for a different purpose.

1.14. **Proposition (weak inequalities and limits; sandwiching).** *Let $f\colon E \to \mathbb{R}$ and let p be a limit point of E.*

(i) *if $f(x) \geqslant 0$ for all $x \in E$ and $f(x) \to A$ then $A \geqslant 0$.*
(ii) *Let $m, M\colon E \to \mathbb{R}$. Suppose that there exists $\eta > 0$ s.t. $m(x) \leqslant f(x) \leqslant M(x)$ for all $x \in E$ such that $0 < |x - p| < \eta$ and that $m(x) \to L$ and $M(x) \to L$ as $x \to p$. Then $\lim_{x \to p} f(x)$ exists and equals L.*

1.15. **Left-hand and right-hand limits.**

For functions defined on some suitable interval, we may usefully refine the function limit definition and talk about right-hand and left-hand limits. We include the definitions here for completeness but shall not make use of them until later.

Let p be a point in $\mathbb{R}$. Assume that f is a function defined on some interval (p, b); and let $L \in \mathbb{R}$. We say that L **is the right-hand limit of** f **at** p if

$$\forall \varepsilon > 0\ \exists \delta > 0\ \forall x \in [a, b)\ \big(0 < x - p < \delta \implies |f(x) - L| < \varepsilon\big).$$

We write this as

$$\lim_{x \to p+} f(x) = L; \text{ or as } \lim_{\substack{x \to p \\ x > p}} f(x); \text{ or sometimes as } f(p+) = L.$$

Similar considerations apply to the left-hand limit as $x \to p$ of a function defined on some interval (a, p).

Here the crucial thing is that, for the notion of right-hand (left-hand) limit of f at p to be defined, we require the function to be defined in a non-empty open interval with right (left) endpoint p.

The proof of the following claim is good practice in using the definitions.

Let $f\colon (a, b) \to \mathbb{R}$ and let $p \in (a, b)$. Then the following are equivalent:

(a) $\lim_{x \to p} f(x) = L$;
(b) *Both* $\lim_{x \to p+} f(x) = L$ *and* $\lim_{x \to p-} f(x) = L$.

1.16. **Extensions: functions tending to infinity and limits at infinity.**

Sometimes we want to extend the notion '$f(x) \to L$ as $x \to p$' to cover 'infinity'. Note that although ∞ appears in our vocabulary, we have *not* given it the status of a number: it can only appear in certain phrases in our mathematical language which are shorthand for quite complicated statements about real numbers.

There are various possible scenarios. The idea on the domain side is to think of infinity as being e limit point of any set which is not bounded above. To capture the idea of $f(x)$ tending to infinity, we use an arbitrary real number in place of ε to measure closeness to infinity.

So, let $E \subseteq \mathbb{R}$ and $f\colon E \to \mathbb{R}$ and let p be a limit point of E. We say that $f(x)$ **tends to** $+\infty$ **as** $x \to p$ if

$$\forall B > 0\ \exists \delta > 0\ \forall x \in E\ \big(0 < |x - p| < \delta \implies f(x) > B\big).$$

We may write this as $f(x) \to +\infty$ as $x \to p$. Note that now we do not exclude $x = p$: exclusion is neither needed nor wanted

Now suppose that $E \subseteq \mathbb{R}$ is a set which is unbounded above and $f\colon E \to \mathbb{R}$. Then we write $f(x) \to \infty$ **as** $x \to \infty$ to mean that:

$$\forall B > 0\ \exists D > 0\ \forall x \in E\ \big(x > D \implies f(x) > B\big).$$

The definition of $f(x) \to L \in \mathbb{R}$ as $x \to \infty$ is formulated analogously. This time we require

$$\forall \varepsilon > 0\ \exists D > 0\ \forall x \in E\ \big(x > D \implies |f(x) - L| < \varepsilon\big).$$

Variants involving negative infinity can be handled in the expected way.

---

2. Continuity of functions at individual points

We all have a good informal idea of what it means to say that a function has a continuous graph: we can draw it without lifting the pen from the paper. But we want now to use our precise definition of '$f(x) \to L$ as $x \to p$' to discuss the idea of continuity. That is, we want to discuss the precise question of whether f is continuous at a particular pint p. We continue the ε-δ theme of the previous section: the material in this section is a necessary technical precursor to what comes later. [For substantial results we need to consider functions which are continuous at *every* point of the domain, rather than those which are continuous at individual points. See Section 3.]

2.1. **Definition (continuity at a point).**

Again let us consider $E \subseteq \mathbb{R}$ and $f\colon E \to \mathbb{R}$ and assume p is a limit point of E.

In the definition of $\lim_{x\to p} f(x)$ in Section 1, the point p need not belong to the domain E of f. Indeed, even when $p \in E$ and $f(p)$ is defined, we steadfastly refused to acknowledge this when considering the limiting behaviour of $f(x)$ as x *approaches* p. Now we change our focus and consider the scenario in which $f(p)$ *is* defined and ask whether $\lim_{x\to p} f(x) = f(p)$.

Here is our key definition. Let $f\colon E \to \mathbb{R}$, where $E \subseteq \mathbb{R}$ and $p \in E$. We say that f **is continuous at** p if

$$\forall \varepsilon > 0\ \exists \delta > 0\ \forall x \in E\ \big(|x - p| < \delta \implies |f(x) - f(p)| < \varepsilon\big).$$

Note that we do not exclude $x = p$: to do so would be neither necessary nor appropriate.

2.2. **Proposition (limit definition and continuity definition reconciled).**

Let $f\colon E \to \mathbb{R}$, where $E \subseteq \mathbb{R}$, and let $p \in E$.

(i) *f is continuous at any isolated point p of E, meaning that p is not a limit point of E.*

(ii) *If $p \in E$ is a limit point of E, then f is continuous at p if and only if*

$$\lim_{x\to p} f(x) \text{ exists and } \lim_{x\to p} f(x) = f(p).$$

Proof. (i) is immediate, since we may choose $\delta > 0$ such that $\{\, x \in E \mid 0 < |x - p| < \delta \,\} = \emptyset$. For such δ, we have $x \in E$ and $|x - p| < \varepsilon$ only if $x = p$ and then $|f(x) - f(p)| < \varepsilon$, trivially.

(ii): It is clear that if the continuity condition holds then the limit one does too. In the other direction, the limit condition, provided the limit is $f(p)$, delivers all that we need for continuity; the inequality $|f(x) - l| < \varepsilon$ holds for $x = p$ as well as the other points x in $|x - p| < \delta$. $\square$

2.3. **Examples (continuity at a point).**

This group of examples illustrates the difference between the definition of a limit in general and the definition of continuity at a point.

(1) Let $f(x) = |x|$ for $x \in \mathbb{R}$. Then f is continuous at each point $p \in \mathbb{R}$. To prove this, note that the Reverse Triangle Inequality gives

$$||x| - |p|| \leqslant |x - p|.$$

Hence, given $\varepsilon > 0$ we can take $\delta := \varepsilon$ in the $\varepsilon - \delta$ definition of continuity at p.

(2) Let $c \in \mathbb{R}$. Consider f defined on $\mathbb{R}$ by

$$f(x) := \begin{cases} c & \text{if } x \neq 0, \\ 1 & \text{otherwise.} \end{cases}$$

Then $\lim_{x\to 0} f(x) = 1$. Hence f is continuous at 0 if and only if $c = 1$. (Compare with Example 1.8(1).)

On the other hand, f is continuous at every point $p \neq 0$, irrespective of the value of c.

(3) Let $\alpha > 0$. The function $f(x) = |x|^\alpha \sin \frac{1}{x}$ is not defined at $x = 0$ so it makes no sense to ask if it is continuous there. In such circumstances we modify f in some suitable way. So we look at

$$g(x) := \begin{cases} |x|^\alpha \sin \frac{1}{x} & \text{if } x \neq 0. \\ 0 & \text{if } x = 0. \end{cases}$$

Then 0 is a limit point of the domain, and we calculated before that $\lim_{x\to 0} g(x) = 0 = g(0)$, so g is continuous at 0.

We will be able to enlarge our catalogue of examples of discontinuous function by using sequences to witness discontinuity. at a particular point.

The following theorem follows immediately from Proposition 2.2(ii) and the proof of Theorem 1.10. Now we do not need to avoid sequences which hit the point.

2.4. **Theorem (continuity at a point via sequences).** *Let $f\colon E \to \mathbb{R}$ where $E \subseteq \mathbb{R}$ and $p \in E$. Then the following two statements are equivalent:*

(a) *f is continuous at p;*
(b) *For every sequence (p_n) in E such that $\lim_{n\to\infty} p_n = p$ it is the case that that $f(p_n) \to f(p)$ as $n \to \infty$.*

As for establishing a limit fails to exist, sequences, for example of rationals or irrationals, can be a good way to prove that a function is **discontinuous**; recall 0.9.

2.5. **Example of a function discontinuous at rationals and continuous at irrationals.**

Consider the function f on $E = (0, 1]$ given by

$$f(x) = \begin{cases} \dfrac{1}{q} & \text{if } x = \dfrac{r}{q} \text{ in lowest terms,} \\ 0 & \text{when } x \text{ is irrational.} \end{cases}$$

We first consider $p \in E$ with p rational. Then certainly $f(p) \neq 0$. We can construct a sequence (x_n) of irrationals with $x_n \in E$ and $x_n \to p$ as $n \to \infty$. But then $f(x_n) = 0 \to 0 \neq f(p)$, so f is not continuous at any rational.

Now consider $p \in E$ with p irrational and let $\varepsilon > 0$. Since the natural numbers are not bounded above (proved in Analysis I) we can choose $N \in \mathbb{N}$ with $1/N < \varepsilon$. Consider the set $\mathbb{Q}_N$ whose members are those rationals r/q in E for which $1 \leqslant q \leqslant N$. The set $\mathbb{Q}_N$ is *finite*. This means that we can find $\eta > 0$ such that $r/q \in E \cap (p - \eta, p + \eta)$ implies $q > N$. Therefore, for $x \in E \cap (p-\eta, p+\eta)$, either x is rational with $0 \leqslant f(x) < 1/N < \varepsilon$ or x is irrational and $f(x) = 0$. Either way $|x-p| < \eta$ implies $|f(x) - f(p)| < \varepsilon$. So f is continuous at each irrational in E.

We can use our characterisation of continuity at limit points in terms of $\lim_{x \to p} f(x)$ (2.2(ii)), together with the Algebra of Function Limits 1.13 to prove that the class of functions continuous at p is closed under all the usual algebraic operations.

As we did for function limits (recall 1.12), we first separate out a technical lemma. This one finds many uses, in this course and beyond.

2.6. **Lemma (continuous function non-zero at a point).** *Let $f\colon E \to \mathbb{R}$ be continuous at $p \in \mathbb{R}$ and assume that $f(p) > 0$. Then there exists $\eta > 0$ such that*

$$x \in E \text{ and } |x-p| < \eta \; f(x) > f(p)/2.$$

In particular $f(x) > 0$ if $x \in E$ and $|x-p| < \eta$.

Proof. Since f is continuous at p we may take $\varepsilon := f(p)/2$ in the continuity condition to obtain $\eta > 0$ such that $x \in E$ and $|x-p| < \eta$ implies $|f(x) - f(p)| < f(p)/2$. The result follows from the Reverse Triangle Inequality:

$$|f(x)| = |f(x) - f(p) + f(p)| \geqslant f(p) - \frac{1}{2}f(p) = \frac{1}{2}f(p). \qquad \square$$

2.7. **Theorem (functions continuous at a point: algebraic operations).** *Let $E \subseteq \mathbb{C}$ and let $p \in E$. Let $f, g\colon E \to \mathbb{R}$, and let $\alpha, \beta \in \mathbb{R}$. Suppose that f, g are continuous at p. Then the following functions are also continuous at p:*

- **Linear combination:** *$\alpha f(x) + \beta g(x)$;*
- **Product:** *$f(x)g(x)$;*
- **Quotient:** *$f(x)/g(x)$, provided $g(p) \neq 0$ (which guarantees that there exists $\eta > 0$ such that $f(x)/g(x)$ is defined $\forall x \in E$ such that $|x-p| < \eta$).*

Proof. This follows directly from the corresponding results on the Algebra of Function Limits. $\square$

2.8. **Example: polynomials and rational functions.**

Let $f\colon \mathbb{R} \to \mathbb{R})$ be a polynomial. Then f is continuous at every point of $\mathbb{R}$).

Further, consider $f(x) = \dfrac{r(x)}{q(x)}$, where $r, q\colon \mathbb{R} \to \mathbb{R}$ are polynomials. Then f is continuous at p provided $q(p) \neq 0$.

This follows immediately from the above theorem because the function $f(x) = x$ with domain $\mathbb{R}$ is continuous at every point.

However we can obtain more than trivial algebraic results.

2.9. **Theorem (composition of continuous functions).** *Let $f\colon E \to \mathbb{R}$ and $g\colon f(E) \to \mathbb{R}$, and define $h\colon E \to \mathbb{R}$ by*

$$h(x) = (g \circ f)(x) := g(f(x)) \qquad \text{for } x \in E.$$

If f is continuous at $p \in E$ and g is continuous at $f(p)$, then h is continuous at p.

Proof. Since g is continuous at $f(p)$,

$$\forall \varepsilon > 0 \; \exists \eta > 0 \; \forall y \in f(E) \; \big(|y - f(p)| < \eta \implies |g(y) - g(f(p))| < \varepsilon\big).$$

Hence

$$\forall x \in E \; \big(|f(x) - f(p)| < \eta \implies |g(f(x)) - g(f(p))| < \varepsilon\big).$$

In addition, f is continuous at p so

$$\forall \eta > 0\ \exists \delta > 0\ \forall x \in E\ \big(|x - p| < \delta \implies |f(x) - f(p)| < \eta\big).$$

Combining these assertions, with η as in the first assertion used to determine δ.

$$\forall \varepsilon > 0\ \exists \delta > 0\ \forall x \in E\ \big(|x - p| < \delta \implies |g(f(x)) - g(f(p))| < \varepsilon\big).$$

Hence h is continuous at p. □

Note in this proof where the assumption that g is defined on $f(E)$ was used. See Problem sheet 1 for a (careful!) formulation of an analogous composition result for function limits.

2.10. **Extending the realm: more examples of continuous functions.**

Recall from Analysis I that certain functions from $\mathbb{R} \to \mathbb{R}$—$\exp(x)$, $\sin(x)$, $\cos(x)$, $\sinh(x)$ and $\cosh(x)$ etc—are defined by their power series, each of which has infinite radius of convergence. You were told in Analysis I that a power series defines a function which is continuous at each point within its interval of convergence. Later on (Section 5) we shall justify this claim.
Unpaid debt
For now, we shall continue to take this fact on trust. This will allow us to use the algebra of continuous functions and the composition of continuous functions to prove the continuity of a wide variety of functions, including many arising in everyday mathematics and in Prelims courses from MT onwards.

2.11. **Example.**

We claim that the function $g\colon \mathbb{R} \to \mathbb{R}$ given by

$$g(x) := \begin{cases} x \sin \frac{1}{x} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

is continuous at every point of $\mathbb{R}$.

We have already proved that g is continuous at 0 (special case of Example 2.3(3)).

If $p \neq 0$: $1/x$ is continuous at p as $p \neq 0$ [Quotient of continuous functions] and $\sin(x)$ is continuous at $1/p$. Hence $\sin(1/x)$ is continuous at p [Theorem 2.9]. Hence $x \sin(1/x)$ is continuous at p [Product of continuous functions].

2.12. **Extensions to complex-valued functions defined on a subset of the complex plane.**

So far we have restricted attention to functions $f\colon E \to \mathbb{R}$, where $E \subseteq \mathbb{R}$. We have done this for three reasons. Firstly, Prelims Analysis II is first and foremost a course about real analysis. Secondly, and less important, we have not wanted to clutter our statements by presenting real and complex versions of results at the same time. Thirdly, what we can do in the complex case on limits and continuity is merely a mechanical adaptation of what we do in the real case, with very little new emerging. Differentiation is a very different matter: the subject known as Complex Analysis is developed in its full glory in MT of the second year.

Accordingly we collect together here a summary of which results from Sections 1 and 2 do carry over to the complex case, with a warning about those which do not.

2.13. **For $\mathbb{R}$ read $\mathbb{C}$?**

In what has gone before, consider replacing $\mathbb{R}$ by $\mathbb{C}$. This leads to

- the definition of a **limit point** pf a subset E of $\mathbb{C}$;
- the definition of $\lim_{x \to p} f(x)$ for a function $f\colon E \to \mathbb{C}$, where p is a limit point of E;
- basic properties of complex limits (capture via complex sequences; (AOL));
- definition of continuity at a point and basic properties (capture via complex sequences; (AOL); algebraic operations on continuous functions and composition).

Properties which do NOT carry over to complex-valued functions are those which involve inequalities or sandwiching; Proposition 1.14 does not have a complex-valued analogue.

As with limits of complex sequences, we can link the limiting behaviour of a function $f\colon E \to \mathbb{C}$ (where $E \subseteq \mathbb{R}$ or $\mathbb{C}$) to the limiting behaviour of its real and imaginary parts $\operatorname{re} f$ and $\operatorname{im} f$. For example, f is continuous at p if and only if $\operatorname{re} f$ and $\operatorname{im} f$ are continuous at p. Moreover, f continuous at p implies that the real-valued function $|f|$ is continuous at p.

Also, when f is complex-valued we can no longer define $\lim_{x\to\infty} f(x)$.

2.14. **Functions of several variables.**

If $E \subseteq \mathbb{R}^n$ and $f\colon E \to \mathbb{R}$ is a function, we may define the notion of continuity of f at a point $\mathbf{p} \in E$. To do this we will use the Euclidean norm of a vector $\mathbf{y} = (y_1, \dots, y_n)$, given by

$$\|\mathbf{y}\| = \|(y_1, \dots, y_n)\| := \sqrt{y_1^2 + \cdots + y_n^2}.$$

Let $f\colon E \to \mathbb{R}$ (or $\mathbb{C}$), where $E \subseteq \mathbb{R}^n$, and $\mathbf{p} \in E$. We say that f **is continuous at p** if $\forall \varepsilon > 0\ \exists \delta > 0$ such that $\forall \mathbf{x} \in E$ such that

$$\|\mathbf{x} - \mathbf{p}\| < \delta \implies |f(\mathbf{x}) - f(\mathbf{p})| < \varepsilon.$$

This notion will be developed in other courses. Suffice it to say for the present that almost all of the properties we will establish in the present course extend painlessly to continuous functions of several variables.

---

3. Functions continuous on an interval

Results in this section are stated and proved for real-valued functions defined on intervals of $\mathbb{R}$. We shall flag up clearly any results do have complex analogues.

We have so far concentrated on what it means to say that a function is continuous at a point. But usually what is more important is not continuity one point at a time, but continuity at all points of a set, and the interesting sets are usually intervals.

3.1. **Definition (continuous function on a set).**

Let $f\colon E \to \mathbb{R}$. We say that f is **continuous on** E if f is continuous at every point of E. For later use we cast this in $\varepsilon - \delta$ form. The function f is continuous on E if

$$\forall p \in E\ \forall \varepsilon > 0\ \exists \delta > 0\ \forall x \in E\ \big(|x - p| < \delta \implies |f(x) - f(p)| < \varepsilon\big).$$

Here δ may depend both on ε and on the point p.

Continuous functions on closed bounded intervals will be revealed to have especially good properties. We reiterate that a closed bounded intervals are the sets of the form $[a, b]$, where $a, b \in \mathbb{R}$ and $a \leqslant b$. Every point of $[a, b]$ is a limit point, but approach to either of the endpoints a and b will be from one side only.

We do not restrict exclusively to functions on closed intervals. In particular it will sometimes be useful to know whether a function f on a half-open interval, $[a, b)$ for example, can be defined at b so that f becomes continuous at b.

3.2. **Left-continuity and right-continuity.**

The definitions of one-sided limits from 1.15) lead on to notions of left- and right-continuity. We say that a function f is **right-continuous at** p **if** $f(p+)$ exists and $f(p+) = f(p)$. Likewise f is **left-continuous at** p **if** $f(p-)$ exists and equals $f(p)$.

The following simple result links continuity at a point to left- and right-continuity (proof: exercise).

Let $f\colon (a,b) \to \mathbb{R}$ and let $p \in (a,b)$. Then the following are equivalent:

(a) *f is continuous at p;*
(b) *f is both left-continuous at p and right-continuous at p.*

3.3. **Examples (one-sided continuity).**

Analysing left- and right-sided behaviour separately may be useful for deciding whether functions given by 'by-cases' specifications are continuous.

(1) Consider $f\colon \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = \begin{cases} x & \text{if } x \geqslant 0, \\ x+1 & \text{if } x < 0. \end{cases}$$

Then $f(0+) = 0$ and $f(0-) = 1$. But $\lim_{x\to 0} f(x)$ does not exist and f fails to be continuous at 0. It is right-continuous but not left-continuous at 0.

(2) Consider $f\colon \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = \begin{cases} 2x^5 & \text{if } x > 0, \\ 5x^2 & \text{if } x \leqslant 0. \end{cases}$$

Then $\lim_{x\to 0+} f(x)$ and $\lim_{x\to 0-} f(x)$ both exist and equal 0. Hence f is continuous at 0. In addition f is continuous at each point of $(0,\infty)$ and at each point of $(-\infty, 0)$, by 3.2. Therefore f is continuous on $\mathbb{R}$. Note how calling on the result in 3.2 avoids a messy ad hoc $\varepsilon - \delta$ argument.

(3) **Periodic extension** of a function on $[0, 2\pi)$. Let $f(x) = x^2$ if $0 \leqslant x < 2\pi$. Note that the family of intervals $[2k\pi, 2(k+1)\pi)$ $(k \in \mathbb{Z})$ slot together without overlaps. We can create a 2π-periodic function F on $\mathbb{R}$ by defining $F(x) = f(x - 2k\pi)$ when $x \in [2k\pi, 2(k+1)\pi)$. You are recommended to sketch the graph of F. We then have

$$\lim_{x\to 0+} F(x) = 0 \quad \text{and} \quad \lim_{x\to 0-} F(x) = \lim_{y\to 2\pi-} f(y) = 4\pi^2.$$

Hence F fails to be continuous at 0 and likewise it is discontinuous at each point $2k\pi$ for $k \in \mathbb{Z}$. On the other hand it is clear that the original function f is continuous on $[0, 2\pi)$.

This example is of relevance to the understanding of the behaviour of Fourier series. Fourier's idea was to represent functions by a superposition of sine and cosine functions. Fourier series are periodic, of period 2π, so the functions they can represent at any individual points would have to be periodic too. Hence our interest in periodic extension. Don't be fooled into thinking that the Fourier series for x^2 on $[0, 2\pi)$ will converge to 0 at 0. It doesn't! (In fact it can be proved that it converges to $\frac{1}{2}(F(0+) + F(0-))$.)

3.4. **Boundedness.**

We have already encountered boundedness in connection with subsets of $\mathbb{R}$. Now we want to consider boundedness of functions.

Let $f\colon E \to \mathbb{R}$. We say that f is **bounded on** E if $\exists M \geqslant 0$ such that

$$\forall x \in E \ |f(x)| \leqslant M.$$

We also say that f is **bounded by** M **on** E and M is a **bound** for f on E.

Saying that a real-valued function f on E is bounded is equivalent to saying that the image of f, that is, the set

$$f(E) := \{\, f(x) \mid x \in E \,\}$$

is bounded both above and below.

When the set $f(E)$ is bounded above, the Completeness Axiom tells us that

$$\sup f := \sup\{\, f(x) \mid x \in E \,\}$$

exists. When $\sup f \in f(E)$ we say that f **attains its sup(remum)**. Corresponding definitions apply to a real-valued function which is bounded below.

While the notion of boundedness is also available for a *complex-valued* function f, the notions of $\sup f$ and $\inf f$ make sense only when f is *real-valued.*

3.5. **Examples: boundedness of functions defined on intervals.**

Note that each of the functions in the examples which follow is continuous on the interval on which it is defined.

(1) Let $E = (0, 1]$ and $f\colon E \to \mathbb{R}$ be given by $f(x) = 1/x$. Then f is bounded below and attains its inf: $\inf f = f(1)$. On the other hand f is not bounded above: $f(x) \to \infty$ as $x \to 0$.

(2) Let $E = [1, \infty)$ and let $f(x) = 1/\mathrm{e}^x$. Then (by properties of exponential) $f(E) \subseteq (0, 1/\mathrm{e}]$. Here $\sup f = 1/\mathrm{e}$ and is attained at $x = 1$ whereas $\inf f = 0$ and is not attained.

(3) Let $E = [0, 1)$ and $f(x) = x\cos(1/(1-x))$. Then $f(E) \subseteq (-1, 1)$ (we assume familiar properties of the cosine function here). So f is bounded.

    Now consider points close to 1. Suppose $x_k := 1 - 1/(k\pi)$ $(k = 1, 2, \ldots)$. Then $f(x_{2n}) \to 1$ as $n \to \infty$ and $f(x_{2n-1}) \to -1$ as $n \to \infty$. Hence $\sup f = 1$ and $\inf f = -1$; neither the sup nor the inf is attained.

(4) Let f be a polynomial. Let $E = [a, b]$, where $-\infty < a < b < \infty$. Then we claim that f is bounded above and below and attains its sup and its inf. But how might you prove this?

    Now consider $E = (-\infty, \infty)$. Then f is bounded on E if and only if f is a constant. It is not bounded, either above or below, if f has odd degree. If f has even degree and in not constant, then it is is bounded either above or below, but not both. But how might you prove this?

Here is a central theorem of the course.

3.6. **BOUNDEDNESS THEOREM.** *Let $[a, b]$ be a closed bounded subinterval of $\mathbb{R}$. Let $f\colon [a, b] \to \mathbb{R}$ be a continuous function. Then*

(i) *f is bounded.*

(ii) *f attains its sup and its inf. That is, there exist points x_1 and x_2 in $[a, b]$ such that*

$$f(x_1) = \sup_{x \in [a,b]} f(x) \quad \text{and} \quad f(x_2) = \inf_{x \in [a,b]} f(x).$$

[In general x_1 and x_2 will not be unique.]

Proof. Consider (i). Argue by contradiction. Suppose f were unbounded. Then for any $n \in \mathbb{N}$, there exists $x_n \in [a, b]$ such that $|f(x_n)| \geqslant n$. Since (x_n) is bounded, by the Bolzano–Weierstrass Theorem, there exists a subsequence (x_{n_k}) converging to p, say. Then p is a limit point of the interval $[a, b]$ so $p \in [a, b]$. Now f is continuous at p and hence

$$f(p) = \lim_{k \to \infty} f(x_{n_k})$$

so in particular the sequence $(f(x_{n_k}))$ is convergent. Therefore, by an Analysis I result, this sequence is bounded. But $|f(x_{n_k})| \geqslant n_k \geqslant k$, so this is a contradiction. Therefore f must be bounded.

We now prove (ii). Let us prove by contradiction that the supremum M of f is attained. Assume the contrary, that is,

$$f(x) < M \quad \text{for all } x \in [a,b].$$

Consider the function g defined on $[a,b]$ by

$$g(x) = \frac{1}{M - f(x)}$$

which is well defined, positive and continuous on $[a,b]$. Therefore by (i) g is bounded on $[a,b]$, by M_0 say, where $M_0 > 0$. Then

$$\frac{1}{M - f(x)} = g(x) \leqslant M_0.$$

It follows that

$$f(x) \leqslant M - \frac{1}{M_0}$$

for all $x \in [a,b]$ which is a contradiction to the fact that M is the *least* upper bound.

A similar argument deals with the infimum, or we can apply what we have done to $-f$ and get the result at once since for any bounded non-empty subset S of $\mathbb{R}$,

$$\inf\{\, s \mid s \in S \,\} = -\sup\{\, -s \mid s \in S \,\}. \qquad \square$$

3.7. **Remarks on the Boundedness Theorem.**

(1) Assume f is a continuous complex-valued function defined on $[a,b]$.
Then $|f|$ is continuous and real-valued and the Boundedness Theorem applies to $|f|$. Hence f is bounded. Part (ii) of the theorem involves order notions: we can no longer define $\sup f$ and $\inf f$ when f is complex-valued.

(2) We employed sequences to prove part (i) of the theorem. An argument via sequences and the Bolzano–Weierstrass Theorem can also be given for part (ii), starting from a sequence of points (x_n) in $[a,b]$ such that $f(x_n) > M - 1/n$, where $M = \sup f$ (this exists by the Approximation Property characterising supremum which was given in Analysis I). Exercise: complete this proof.

3.8. **Examples 3.5 revisited.**

We can now view these examples in context. We can see that the assumptions that the interval in the Boundedness Theorem on which f is considered needs to be both closed and bounded. For a continuous function on an interval E which does not contain one (or both) of its endpoints, and/or whose domain is unbounded, a continuous real-valued function on E may fail to be bounded, or may be bounded but fail to attain either its sup or its inf, or both.

So far we have concentrated on extreme values, the supremum and the infimum.of a continuous real-valued function on a closed bounded interval. What can we say about possible values between these? Here is the second of our Big Theorems.

3.9. **INTERMEDIATE VALUE THEOREM (IVT).** *Let $f\colon [a,b] \to \mathbb{R}$ be continuous, and let c be a real number between $f(a)$ and $f(b)$. Then there is at least one point $\xi \in [a,b]$ such that $f(\xi) = c$.*

[The restriction that f be real-valued is essential.]

Proof. By considering $-f$ instead of f if necessary, we may assume that $f(a) \leqslant c \leqslant f(b)$. The cases $c = f(a)$ and $c = f(b)$ are trivial, so assume $f(a) < c < f(b)$.

Define $g(x) = f(x) - c$. Then $g(a) < 0 < g(b)$. Hence it is sufficient to prove that if $g\colon [a,b] \to \mathbb{R}$ is continuous and $g(a) < 0 < g(b)$, then there exists $\xi \in (a,b)$ such that $g(\xi) = 0$.

Define

$$S = \{\, x \in [a,b] \mid g(x) < 0 \,\}.$$

Then $a \in S$ so $E \neq \emptyset$ and E is bounded above by b. So, by the Completeness Axiom, $\xi = \sup S$ exists. Since $a \in S$ we have $\xi = \sup S \geqslant a$ and since b is an upper bound for S we have $\xi = \sup S \leqslant b$. Therefore $\xi \in [a,b]$. We shall now prove, by contradiction, that $g(\xi) = 0$.

Suppose first that $g(\xi) > 0$ (so $\xi > a$). Lemma 2.6 implies there exists η_1 such that $g(x) > 0$ in some interval $(\xi - \eta_1, \xi]$, where $0 < \eta_1 < \xi - a$. This contradicts the definition of ξ as $\sup S$.

Suppose now that $g(\xi) < 0$ (so $\xi < b$). Lemma 2.6 applied to $-g$ implies that $g(x) < 0$ in some interval $[\xi, \xi + \eta)$ where $0 < \eta < b - \xi$. This contradicts the definition of ξ. □

Remark. There is an alternative proof of the IVT based on repeated bisection of intervals. This can be found for example in Bartle and Sherbert. Its strategy can be used as an aid to homing in on a root of an equation by successive approximation. Observe that in the two examples which follow we use the IVT to prove the existence of a root of an equation in a certain interval but do not get information on how to locate that root more precisely.

3.10. **Example.** There exists a unique positive number ξ such that $\xi^2 = 2$.

To prove this we consider $f(x) = x^2 - 2$. Note that $f(0) = -2$ and $f(2) = 2$. So $f\colon [0,2] \to \mathbb{R}$, $f(0) < 0 < f(2)$ and also, as f is a polynomial, it is continuous. Thus, by the IVT, there exists $\xi \in (0,2)$ such that $f(\xi) = 0$, as required. Uniqueness can be proved as in Analysis I.

[Note that the proof of existence of $\sqrt{2}$ given in Analysis I also relied crucially on the Completeness Axiom and on a trichotomy argument, as did our proof of the IVT.]

3.11. **Example (IVT).**

(1) Let $f\colon [0,1] \to [0,1]$ be continuous. Then we claim that there exists $\xi \in [0,1]$ such that $f(\xi) = \xi$.
   We shall apply the IVT to g where $g(x) = f(x) - x$. Then g is continuous on $[0,1]$ and $g(0) = f(0) \geqslant 0$ and $g(1) = f(1) - 1 \leqslant 0$. The result now follows from the IVT.

(2) Suppose $f\colon [0,\infty) \to [0,\infty)$ is continuous and bounded. We claim that there exists ξ such that $f(\xi) = \xi$. Consider the continuous function g on $[0,\infty)$ given by $g(x) = f(x) - x$. Then $g(0) \geqslant 0$. Also, because f is bounded, g is not bounded below. Hence there exists β such that $g(\beta) < 0$. We can now apply the IVT to g on $[0,\beta]$ to show that there exists ξ such that $g(\xi) = 0$.

We can combine the Boundedness Theorem and the IVT to arrive at the following very useful theorem. It tells us that a continuous real-valued function maps a closed bounded interval onto a closed bounded interval.

3.12. **Theorem.** *Let $f\colon [a,b] \to \mathbb{R}$ be a real valued continuous function. Then $f([a,b]) = [m,M]$ for some $m, M \in \mathbb{R}$.*

Proof. By the Boundedness Theorem, part (i), we can define

$$m := \inf_{x\in[a,b]} f(x) \quad \text{and} \quad M := \sup_{x\in[a,b]} f(x).$$

Clearly $f([a,b]) \subseteq [m,M]$.

By the Boundedness Theorem, part (ii), there exist $\alpha \in [a,b]$ and $\beta \in [a,b]$ such that $f(\alpha) = m$ and $f(\beta) = M$. Hence $m, M \in f([a,b])$.

Now let $y \in [m,M]$, so $f(\alpha) \leqslant y \leqslant f(\beta)$. By applying the IVT to f restricted to the interval $[\alpha,\beta]$ (or $[\beta,\alpha]$ as the case may be) we find $x \in [\alpha,\beta] \subseteq [a,b]$ such that $f(x) = y$. Therefore $y \in f([a,b])$. Hence $[m,M] \subseteq f([a,b])$. □

3.13. **Monotonic functions.**

The following definitions require the ordered structure of real numbers, and so apply only in $\mathbb{R}$. Let $E \subseteq \mathbb{R}$ and $f\colon E \to \mathbb{R}$. We say that

(i) f is **increasing** if $f(x) \leqslant f(y)$ whenever $x \leqslant y$;
(ii) f is **strictly increasing** if $f(x) < f(y)$ whenever $x < y$.

We define likewise a **decreasing function** and a **strictly decreasing function**. A function is called **monotonic** on E if it is increasing or decreasing on E.

Note that a function which is strictly monotonic is necessarily injective.

We are now ready to prove the third Big Theorem of this section. It will tell us that a continuous, strictly increasing function on a closed bounded interval has a continuous strictly increasing inverse function.

3.14. **The Continuous Inverse Function Theorem.** *Let $f\colon [a,b] \to \mathbb{R}$ be strictly increasing and continuous, where $a < b$. Then*

(i) $f([a,b]) = [f(a), f(b)]$.
(ii) *f is a bijection from $[a,b]$ to $[f(a), f(b)]$.*

Let $g\colon [f(a), f(b)] \to [a,b]$ be the nap inverse to f.

(iii) *g is strictly increasing.*
(iv) *g is continuous on $[f(a), f(b)]$.*

[The requirement that f be real-valued is essential.]

Proof. Parts (i) and (ii) come from Theorem 3.12 and the fact that f is strictly increasing.

It remains to establish the properties of g. Part (iii) is obtained by a simple trichotomy argument. We must prove that g is continuous.

To avoid awkwardness with endpoints (getting involved with half-open intervals, etc), we first extend f to a strictly increasing continuous function F on an interval $[A,B]$, where $A < a < b < B$. It is easy to construct a suitable F. Take any $A < a$ and $B > b$. Define F on $[A,a]$ to be linear, of gradient 1 (which will ensure it is strictly increasing) and such that $F(a) = f(a)$. Define F on $[b,B]$ similarly. Continuity of F follows from 1.15.

Then F has an inverse function $G\colon [F(A), F(B)] \to [A,B]$. We shall show that G is continuous at each point of $(F(A), F(B))$. The restriction of G to $[f(a), f(b)]$ is g, and so g will be continuous at each point of its domain.

To prove G is continuous at p, where $F(A) < p < F(B)$, we proceed as follows.

Claim 1: For any α, β with $A \leqslant \alpha < \beta \leqslant B$,

$$F((\alpha,\beta)) = (F(\alpha), F(\beta)),$$

so that F maps any open interval contained in $[A,B]$ onto an open interval.

Proof. As in (i), F maps $[\alpha,\beta]$ onto $[F(\alpha), F(\beta)]$. Then, because F is strictly increasing and so also injective,

$$\begin{aligned} F((\alpha,\beta)) = F([\alpha,\beta] \setminus \{\alpha,\beta\}) &= F([\alpha,\beta]) \setminus \{F(\alpha), F(\beta)\} \\ &= [F(\alpha), F(\beta)] \setminus \{F(\alpha), F(\beta)\} = (F(\alpha), F(\beta)). \end{aligned}$$

Claim 2: Let $p \in (F(A), F(B))$. Then G is continuous at p.

Proof. $G(p) \in (A, B)$. Let $\varepsilon > 0$ and assume without loss of generality that $A \leqslant G(p) - \varepsilon < G(p) + \varepsilon \leqslant B$. By Claim 1, F maps the open interval $(G(p) - \varepsilon, G(p) + \varepsilon)$ to an open interval, J say, contained in $(F(A), F(B))$ and J contains p. Therefore there exists $\delta > 0$ such that $(p - \delta, p + \delta) \subseteq J$. This means that

$$|y - p| < \delta \Longrightarrow |G(y) - G(p)| < \varepsilon.$$

□

3.15. **Corollary to Continuous IFT.** *Let $f\colon [a, b] \to \mathbb{R}$ be a strictly increasing continuous function. Let $a \leqslant c < d \leqslant b$. Then f maps (c, d) onto $(f(c), f(d))$.*

3.16. **Remarks about inverse functions.**

Problem sheet 3 asks you to prove that, if $f\colon [a, b] \to \mathbb{R}$ is a continuous, 1–1 function with $f(a) < f(b)$, then f is strictly increasing on $[a, b]$. So in the statement of the IFT it is sufficient to assume that $f\colon [a, b] \to \mathbb{R}$ is continuous and 1–1.

If you choose to use the notation f^{-1} for the inverse function of f, when this exists, then you must make very clear what you intend the domains of f and f^{-1} to be. For example sine and cosine are only invertible on a part of their domain where they are increasing or decreasing.

3.17. **Remarks: a wider context for this section's results.**

In Part A, you will find out more about continuity and how to capture this property more elegantly than with the $\varepsilon - \delta$ definition, and discover what is really going on with the IVT (and the IFT). The proof given here for the IFT is a first step along this road.

All the results in this section are specialisation of results true in more general settings than the real line, bringing in the ideas of topology. For example, the Boundedness Theorem is a result about continuous real-valued functions whose domains are compact sets and the IVT can properly be seen as a theorem about connectedness.

3.18. **Inverse functions, a prototype example: exponential and logarithm.**

Your likely first encounter with inverse functions would have occurred when you were introduced to the (natural) logarithm function as the inverse of the exponential function. Here, assuming familiar but as yet unproved properties of the exponential function, we show how to exploit the IFT to establish the existence and basic properties of $\log x$ (or $\ln x$ as you may have known it at school).

3.19. **Properties of the (real) exponential function.** We define $\exp(x)$, also written e^x, by

$$\exp(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!} \quad (x \in \mathbb{R}).$$

We record the following properties of the exponential function:

(a) exp is continuous on $\mathbb{R}$;

(b) exp is differentiable at each $x \in \mathbb{R}$ with $\exp'(x) = \exp(x)$;

(c) $\exp(x)\exp(y) = \exp(x + y)$ (for $x, y \in \mathbb{R}$);

(d) $\exp(0) = 1$ and $\exp(-x) = 1/\exp(x)$ (for $x > 0$);

(e) $\exp(x) > 0$ and $\exp(x) > 1 + x$ $(x \geqslant 0)$;

(f) $\exp(x) \to \infty$ as $x \to \infty$;

(g) $\exp(x) \to 0$ as $x \to -\infty$;

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

Treat (a) & (b) as unpaid debts

Here we elect to retain (a) within our set of unpaid debts. A direct proof is possible, but uninstructive, A more informative, and simpler, proof will be available in Section 5 (uniform convergence). Part (b) needs to draw on our later study of differentiability in general and of the Differentiation Theorem for Power Series in particular. It was stated in Analysis I and its use illustrated there in the proof of (c). Part (d) makes use of (c). Part (e) comes from the definition of $\exp(x)$ and (f) is then immediate, ,and (g) follows from this and (d) (recall the definitions in 1.16). We can obtain (h) from (c) and (d) (an alternative proof via (b) will be available later).

3.20. **Proposition (invertibility of the exponential function).** *The function* exp *is a bijection from* $\mathbb{R}$ *onto* $(0, \infty)$*. It has a well-defined inverse function which is defined and continuous on* $(0, \infty)$*.*

Proof. From the fact that exp is strictly increasing it is 1–1. Now we prove that $\exp(x)$ maps $\mathbb{R}$ *onto* $(0, \infty)$. Observe that

$$(0, \infty) = \bigcup_{N=0}^{\infty} [N^{-1}, N]$$

(from results in Analysis I). Fix N. By **??**(f) and (g), we can find a and b, with $a < 0 < b$, such that $\exp(a) < N^{-1}$ and $\exp(b) > N$. By the IVT applied to the continuous function exp, this function attains every value in $[\exp(a), \exp(b)]$. In particular, every point in $[N^{-1}, N]$ is a value of exp. We deduce that exp maps $\mathbb{R}$ onto $(0, \infty)$.

We want the inverse of exp to be continuous at each point q of $(0, \infty)$. Choose N such that $N^{-1} < q < N$. The IFT applied to exp on $[N^{-1}, N]$ implies that the inverse function is continuous at q. □

3.21. **The logarithm.**

By Proposition 3.20 the logarithm function can be defined on $(0, \infty)$ as the function inverse to the exponential, and is continuous on $(0, \infty)$.

We have $\log \mathrm{e} = 1$, where $\mathrm{e} = \exp(1) = \sum \frac{1}{n!}$.

Moreover, for any $u, v > 0$,

$$\begin{aligned} \exp(\log u + \log v) &= \exp(\log u) \cdot \exp(\log v) \qquad \text{(by 3.19(b))} \\ &= u \cdot v. \end{aligned}$$

Since $\log(\exp x) = x$ for all $x \in \mathbb{R}$ we deduce that

$$\log(u \cdot v) = \log u + \log v \quad \text{for all } u, v > 0.$$

3.22. **Example on IVT (assuming properties of exponential and log functions).**

If you sketch the graphs of $y = \mathrm{e}^x$ and $y = \alpha x$, you will see that if $\alpha = \mathrm{e}$ the curves touch, if $\alpha < \mathrm{e}$ they do not meet, but if $\alpha > \mathrm{e}$ then they meet twice.

We now show how to make this graphical argument rigorous using the IVT. It shows that if $\alpha > \mathrm{e}$ there exist two solutions. Once we have covered differentiability you will be able to prove that there are exactly two solutions, by considering the sign of the derivative f'.

We are assuming continuity of e^x. So $f(x)$ is continuous on $[0, \infty)$. From the power series definition of e^x we see that $\mathrm{e}^x > x^2/2$. Thus $\mathrm{e}^X > \alpha X$ for any $X > 2\alpha$. Fix such an X $(> \log \alpha)$.

Then $f(0) = 1 > 0$, $f(\log \alpha) = \alpha(1 - \log \alpha) < 0$, $f(X) > 0$. So we can apply the IVT to the two intervals $[0, \log \alpha]$, and $[\log \alpha, X]$ to find that there exist $x_1 \in [0, \log \alpha]$ such that $f(x_1) = 0$, and $x_2 \in [\log \alpha, X]$ such that $f(x_2) = 0$ as required.

3.23. **Tactical tips on inverse functions.**

What we have done with the exponential and logarithm functions provides a prototype for other examples, such as trigonometric functions and their inverse functions.

The first point to be aware of is that in order for a function f to have a well-defined inverse function we need it to be a bijection and strictly monotonic. This may restrict the interval on which we can construct the inverse. For example the tan function satisfies these requirements provided we restrict it to the open interval $(-\pi/2, \pi/2)$.

The next thing to remember is that the Continuous IFT applies to a continuous function on a *closed bounded* interval. But it is common for functions to have a well-defined inverse, g say, on an interval I which is *not* closed and bounded: exp on $(0, \infty)$, tan on $(-\pi/2, \pi/2)$, for example. Then we need to identify $f(I)$. Finally, we want to show g is continuous on $f(I)$, that is, g is continuous at each point q of $f(I)$ (remember that continuity is a *local* property). To do this we aim to find a closed bounded subinterval $[\alpha, \beta]$ of I such that $f(\alpha) < q < f(\beta)$. Then IFT applied to the restriction of f to $[\alpha, \beta]$ to conclude that g is continuous at q.

4. Uniform continuity

This section and the next one are unashamedly technical. In them we look closely at conditions for continuity of functions and at convergence of sequences of functions. The pay-off will be theorems which are important throughout Analysis.

4.1. **Definition (uniform continuity).**

Let $f\colon E \to \mathbb{R}$. Then f is **uniformly continuous on** E if

$$\forall \varepsilon > 0 \;\; \exists \delta > 0 \; \forall p \in E \; \forall x \in E \;\; \big(|x-p| < \delta \implies |f(x)-f(p)| < \varepsilon\big). \quad \text{(uniform continuity on } E\text{)}$$

Compare this with the definition of f being continuous at every $p \in E$:

$$\forall p \in E \; \forall \varepsilon > 0 \;\; \exists \delta > 0 \; \forall x \in E \;\; \big(|x-p| < \delta \implies |f(x)-f(p)| < \varepsilon\big). \quad \text{(continuity at each } p \in E\text{)}$$

The difference between the two statements is in *the order of the quantifiers*. Swapping $\forall$'s doesn't affect the meaning, but swapping the order in which $\forall p$ and $\exists \delta$ occur does change the meaning. Read the expressions from left to right. For uniform continuity on E we need a δ which works universally—that is, for all p in E at the same time. For continuity on E we first choose any p and then find δ that works for that choice of p: in this case δ is allowed to depend on p.

Of course if $f\colon E \to \mathbb{R}$ is uniformly continuous on E then f is continuous on E. The converse is false, as we now demonstrate.

4.2. **Example: a continuous function which is not uniformly continuous.**

Consider $f(x) = \dfrac{1}{x}$ on $E = (0,1]$. Certainly f is continuous on E. We shall show that f fails to be uniformly continuous on E.

Take $\varepsilon = 1$. We show that there is no $\delta > 0$ that works in the condition for uniform continuity.

Take sequences $x_n = \frac{1}{n}$ and $y_n = \frac{1}{n+1}$. Then $|f(x_n) - f(y_n)| = 1$, but $|x_n - y_n| \to 0$. So for any $\delta > 0$, there exists n such that $|x_n - y_n| < \delta$ but $|f(x_n) - f(y_n)| \not< 1$. So f is not uniformly continuous.

4.3. **Example (Lipschitz continuous functions).**

Here is one class of functions that satisfy the uniform continuity condition. We say that f is **Lipschitz continuous on** E if there exists a constant $K > 0$ such that

$$\forall x, y \in E \;\; |f(x) - f(y)| \leqslant K|x-y|.$$

Assume f satisfies this condition. Given $\varepsilon > 0$ choose $\delta := \dfrac{\varepsilon}{2K}$. Then $\delta > 0$ and for $x, y \in E$ for which $|x-y| < \delta$,

$$|f(x) - f(y)| \leqslant K|x-y| \leqslant \tfrac{1}{2}\varepsilon < \varepsilon.$$

Thus f is uniformly continuous on E.

We claim that $f(x) = \sqrt{x}$ is Lipschitz continuous on $[1,\infty)$, so is uniformly continuous on that set. To obtain the Lipschitz condition note that, for all $x, y \geqslant 1$.

$$|\sqrt{x} - \sqrt{y}| = \frac{|x-y|}{\sqrt{x}+\sqrt{y}} \leqslant \tfrac{1}{2}|x-y|.$$

Uniform continuity is a condition that is found to be necessary in certain technical proofs in analysis which involve continuous functions on intervals. So the following theorem is important beyond the present course.

4.4. **Theorem (uniform continuity of a real-valued continuous function on a closed bounded interval).** *If $f\colon [a,b] \to \mathbb{R}$ is continuous, then f is uniformly continuous on $[a,b]$.*

Proof. Suppose for a contradiction that f were not uniformly continuous. By the contrapositive of the uniform continuity condition there would exist $\varepsilon > 0$ such that for any $\delta > 0$—which we choose as $\delta = \frac{1}{n}$ for arbitrary n—there exists a pair of points x_n, $y_n \in [a,b]$, such that

$$|x_n - y_n| < \tfrac{1}{n} \quad \text{but} \quad |f(x_n) - f(y_n)| \geqslant \varepsilon.$$

Since each $x_n \in [a,b]$, the sequence (x_n) is bounded, and by the Bolzano–Weierstrass Theorem there exists a subsequence (x_{n_k}) which converges to some p. Hence p must be a limit point of $[a,b]$, so $p \in [a,b]$. But

$$\begin{aligned}|y_{n_k} - p| &\leqslant |x_{n_k} - y_{n_k}| + |x_{n_k} - p| \\ &< \frac{1}{n_k} + |x_{n_k} - p| \to 0.\end{aligned}$$

Thus $x_{n_k} \to p$ and $y_{n_k} \to p$, so that by continuity at p we have

$$0 < \varepsilon \leqslant |f(x_{n_k}) - f(y_{n_k})| \leqslant |f(x_{n_k}) - f(p)| + |f(y_{n_k}) - f(p)| \to 0 \text{ as } k \to \infty.$$

This gives the required contradiction. □

4.5. **Another example of uniform continuity on an unbounded interval.**

We claim that $f(x) = \sqrt{x}$ is uniformly continuous in the unbounded interval $[0,\infty)$.

This example will illustrate how we can establish uniform continuity of a function known to be uniformly continuous on suitable subintervals.

We know from Theorem 4.4 that f is uniformly continuous on the closed bounded interval $[0,1]$ provided it is continuous there. Continuity at 0 is easy. Fix $\varepsilon > 0$. Then, for $x \geqslant 0$,

$$|\sqrt{x} - \sqrt{0}| < \varepsilon \iff 0 \leqslant x < \delta := \varepsilon^2.$$

Hence f is continuous at 0. The argument for continuity at $p \neq 0$ is similar to that used in 4.3. Let $\varepsilon > 0$. Take x such that $|x - p| < \eta := p/2$. Then $\sqrt{x} > \sqrt{p}/\sqrt{2}$. For such x,

$$|\sqrt{x} - \sqrt{p}| = \frac{|x-p|}{\sqrt{x} + \sqrt{p}} \leqslant \frac{\sqrt{2}|x-p|}{\sqrt{3p}} < \varepsilon \text{ if } |x - p| < \delta := \min\{\eta, \sqrt{3p/2}\varepsilon\}.$$

(Note that care would be needed to establish uniform continuity on $[0,1]$ directly from the $\varepsilon - \delta$ definition; Theorem 4.4 allows us to avoid this.)

We also proved in 4.3 that f is Lipschitz, and so uniformly continuous, on $[1,\infty)$.

We now need a patching-together argument to establish uniform continuity on $[0,1] \cup [1,\infty)$. This requires more $\varepsilon - \delta$ dexterity than a corresponding patching argument for continuity. We give the details to serve as an illustration of the method.

We know

$$\forall \varepsilon > 0 \; \exists \delta_1 > 0 \; \forall x, y \in [0,1] \; \big(|x - y| < \delta_1 \implies |\sqrt{x} - \sqrt{y}| < \tfrac{1}{2}\varepsilon\big)$$

and

$$\forall \varepsilon > 0 \; \exists \delta_2 > 0 \; \forall x, y \in [1,\infty) \; \big(|x - y| < \delta_2 \implies |\sqrt{x} - \sqrt{y}| < \tfrac{1}{2}\varepsilon\big).$$

Choose $\delta = \min\{\delta_1, \delta_2\} > 0$. Suppose that $|x - y| < \delta$. If $x, y \geqslant 1$ or $x, y \leqslant 1$ we are done.

So suppose that $x \in [0,1]$ and $y \geqslant 1$ and $|x - y| < \delta$. Then $|x - 1| < \delta$ and $|y - 1| < \delta$ so that

$$\begin{aligned}|\sqrt{x} - \sqrt{y}| &\leqslant |\sqrt{x} - \sqrt{1}| + |\sqrt{y} - \sqrt{1}| \\ &< \tfrac{1}{2}\varepsilon + \tfrac{1}{2}\varepsilon = \varepsilon\end{aligned}$$

Hence $|\sqrt{x} - \sqrt{y}| < \varepsilon$ whenever $x, y \in [0, \infty)$ such that $|x - y| < \delta$. By definition, $f(x) = \sqrt{x}$ is uniformly continuous on $[0, \infty)$.

4.6. **Uniform continuity of a function on an interval** $(a, b]$.

The proofs of the following assertions are sought on Problem sheet 3. The results give insight into uniform continuity of a function on an interval which is not closed.

(a) *Let* $f\colon (a, b] \to \mathbb{R}$ *be continuous and assume that* $\lim_{x\to a+} f(x)$ *exists. Then* f *is uniformly continuous on* $[a.b]$.

(b) *Assume that* $g\colon (a, b] \to \mathbb{R}$ *is uniformly continuous. Then:*
  - (i) *If* (x_n) *is a Cauchy sequence in* $(a, b]$*, then* $(g(x_n))$ *is also a Cauchy sequence.*
  - (ii) *Let* (x_n) *and* (y_n) *be sequences in* $(a, b]$ *which converge to* a*. If* $(g(x_n))$ *tends to* L *and* $(g(y_n))$ *tends to* M *then* $L = M$.
  - (iii) $\lim_{x\to a} g(x)$ *exists.*

---

5. Uniform convergence of sequences and series

This important section continues the study of continuous functions on closed bounded subintervals of $\mathbb{R}$, investigating limits and infinite sums of such functions. In particular we look at power series.

In Analysis one often wants to know how different limiting processes interact with one other. In particular, does a limiting process, such as that involved in continuity, commute with another type of limit? Sadly, the answer in general is 'no'. This leads us to try to find sufficient conditions under which the answer will be 'yes'. In this section we take a first excursion into problems of this kind.

5.1. **Pointwise convergence.**

Initially, we want to consider a sequence (f_n) of functions, where $E \subseteq \mathbb{R}$ and $f_n\colon E \to \mathbb{R}$ for $n = 1, 2, \ldots$. Observe that, for each *fixed* $x \in E$, the sequence $(f_n(x))$ is a sequence of real numbers, whose behaviour we can analyse by the techniques of Analysis I.

We say (f_n) **converges pointwise** to the function $f\colon E \to \mathbb{R}$ (and write $f = \lim f_n$ or $f_n \to f$ on E) if for each $x \in E$ the sequence $(f_n(x))$ converges to $f(x)$. That is,

$$\forall x \in E \ \forall \varepsilon > 0 \ \exists N \in \mathbb{N} \ \forall n > N \quad |f_n(x) - f(x)| < \varepsilon. \qquad \text{(pointwise convergence on } E\text{)}$$

[Here N is allowed to depend on both x and ε. Sometimes in proofs using the pointwise convergence condition (or other similar conditions) it is neater to work with '$\forall n \geqslant N$' rather than '$\forall n > N$'. This is no more than a notational adjustment; the condition itself does not change.]

Pointwise convergence is nothing unfamiliar. In saying, for example,

$$e^x = 1 + x + \frac{x^2}{2!} + \cdots \qquad \text{on } \mathbb{R}$$

we mean precisely that the partial sums of the series on the right-hand side converge pointwise to the exponential function for each $x \in \mathbb{R}$.

5.2. **A first cautionary example.**

Consider the sequence of functions (f_n), where $f_n\colon [0,1] \to \mathbb{R}$ given by

$$f_n(x) = \begin{cases} -nx+1 & \text{if } 0 \leqslant x < \frac{1}{n}, \\ 0 & \text{if } x \geqslant \frac{1}{n}. \end{cases}$$

Consider also the function $f\colon [0,1] \to \mathbb{R}$ given by

$$f(x) = \begin{cases} 1 & \text{if } x = 0, \\ 0 & \text{if } x > 0. \end{cases}$$

Sketch the graphs! What happens as n increases? Note that for each fixed $x \in [0,1]$ we have $f(x) = \lim_{n\to\infty} f_n(x)$ (separate cases $x \neq 0$ and $x = 0$—see 5.7(1) below for details). Hence (f_n) converges pointwise to f.

Note that although all the f_n are continuous the pointwise-limit function f is *not* continuous at 0:

$$\lim_{x\to 0} f(x) = 0 \qquad \text{but} \qquad f(0) = \lim_{n\to\infty} f_n(0) = 1.$$

Spelling this out,

$$\lim_{x\to 0}\Big(\lim_{n\to\infty} f_n(x)\Big) = \lim_{x\to 0} f(x) = 0 \qquad \text{but} \qquad \lim_{n\to\infty}\Big(\lim_{x\to 0} f_n(x)\Big) = \lim_{n\to\infty} 1 = 1.$$

The order in which the limits are taken affects the value of the iterated limit.

Moral: in general, iterated limits may squabble. They must be handled with care.

See 5.8–5.10 for more in the same vein.

Uniform continuity leads to stronger results than continuity one point at a time. The idea in the definition of uniform continuity was to require a 'universal δ'. There is a parallel with the key definition of this section, which we now give.

5.3. **Definition (uniform convergence of a sequence of functions on a set).**

Let E be a subset of $\mathbb{R}$. Let (f_n) be a sequence of functions $f_n\colon E \to \mathbb{R}$. Then (f_n) **converges uniformly to f on E** if

$$\forall \varepsilon > 0\ \exists N \in \mathbb{N}\ \forall n > N\ \forall x \in E\ \ |f_n(x) - f(x)| < \varepsilon. \qquad \text{(uniform convergence on } E\text{)}$$

If this holds we write $f_n \xrightarrow{\text{u}} f$ on E. Note that specifying the set E is an integral part of the definition. The order of the quantifiers matters: the uniform convergence condition demands a universal N which is independent of x.

It is immediate from the definitions that if $f_n \xrightarrow{\text{u}} f$ on E then (f_n) converges pointwise to f on E.

The next theorem gives a reason why uniform convergence is a Good Thing.

5.4. **Theorem (uniform convergence preserves continuity).** *Let (f_n) be a sequence of continuous real-valued functions on E which converges uniformly to f on E. Then f is continuous on E.*

Proof. We employ what you'll come to know as an "$\varepsilon/3$ argument". To prove this we fix the obligatory $\varepsilon > 0$ and let $p \in E$. We'll show f is continuous at p.

By uniform convergence we can find $N \in \mathbb{N}$ such that

$$n \geqslant N \implies \forall x \in E\ |f_n(x) - f(x)| < \varepsilon/3.$$

(Small technicality: it's going to be tidier to work with $n \geqslant N$ rather than $n > N$. Difference is immaterial.)

By continuity of f_N at p there exists $\delta > 0$ such that

$$|x - p| < \delta \implies |f_N(x) - f_N(p)| < \varepsilon/3$$

(δ depending on N—but N is fixed). Hence for $|x - p| < \delta$,

$$\begin{aligned}|f(x) - f(p)| &= |f(x) - f_N(x) + f_N(x) - f_N(p) + f_N(p) - f(p)| \\ &\leqslant |f(x) - f_N(x)| + |f_N(x) - f_N(p)| + |f_N(p) - f(p)| < \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon.\end{aligned}$$

This suffices to prove our claim. Note that uniformity of convergence is needed to handle the first term. □

We now convert the uniform convergence condition into a more amenable form.

5.5. **Proposition (testing for uniform convergence in practice).** *Let $E \subseteq \mathbb{R}$ and (f_n) be a sequence of functions with $f_n\colon E \to \mathbb{R}$ for each n. Assume that (f_n) converges pointwise to f on E. Then the following statements are equivalent:*

(a) $f_n \overset{\text{u}}{\to} f$ *on E;*

(b) *for each n (sufficiently large) the set $\{\, |f_n(x) - f(x)| \mid x \in E \,\}$ is bounded and*

$$\alpha_n := \sup_{x\in E} |f_n(x) - f(x)| \to 0 \text{ as } n \to \infty.$$

Proof. Assume (a). Then, given $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that for $n > N$ and for all $x \in E$ we have $|f_n(x) - f(x)| < \varepsilon/2$. So the first condition in (b) holds for such n and hence α_n is well defined. Fix n and take the supremum over $x \in E$ to get, for $n > N$,

$$0 \leqslant \alpha_n = \sup_{x\in E} |f_n(x) - f(x)| \leqslant \varepsilon/2 < \varepsilon.$$

Hence $\alpha_n \to 0$.

Conversely, assume (b). Given $\varepsilon > 0$, choose N so that $n > N$ implies $0 \leqslant \alpha_n < \varepsilon$. Then, for all $x \in E$,

$$n > N \implies |f_n(x) - f(x)| \leqslant \alpha_n < \varepsilon.$$

Hence $f_n \overset{\text{u}}{\to} f$. □

5.6. **Testing for uniform convergence: tactics.**

A few comments on working with Proposition 5.5 are in order. First of all, it allows us to reduce testing for uniform convergence of (f_n) on E to three steps:

Step 1: pointwise limit.

With $x \in E$ fixed, find $f(x) := \lim_{n\to\infty} f_n(x)$, or show it fails to exist (of course, if the pointwise limit fails to exist for any $x \in E$, then certainly (f_n) does not converge uniformly and we proceed no further). Look out for values of x which need special attention.

Step 2: calculate α_n (or show it does not exist).

Assuming all f_n and f are continuous and E is an interval $[a, b]$ (the most common scenario), the Boundedness Theorem applied to the continuous function $|f_n - f|$ tells us the sup is attained, so we want to know the maximum value of $|f_n - f|$. Frequently $f_n - f$ will be of constant sign so we can get rid of the modulus signs. Then, if the functions f_n and f are differentiable the supremum (or infimum) of $f_n - f$ will be achieved either at a or at b or at some interior point

unpaid debt

where $\dfrac{\mathrm{d}}{\mathrm{d}x}(f_n(x) - f(x)) = 0$. It is fine to use school calculus to find maxima and minima by differentiation, when the derivative exists—we'll validate this technique later. See examples below for illustrations.

Step 3: does α_n tend to 0?

Note that (α_n) is a sequence of real numbers. We are back in Analysis I territory, and can use standard techniques and standard limits from the MT course.

Note that in Step 1 we work with fixed x and in Steps 2 & 3 we work with fixed n: we do not need to consider both x and n varying at the same time.

5.7. **Testing for uniform convergence: examples.**

(1) Consider again (f_n) on $E = [0, 1]$, from 5.2:

$$f_n(x) = \begin{cases} -nx + 1 & \text{if } 0 \leqslant x < \frac{1}{n}, \\ 0 & \text{if } x \geqslant \frac{1}{n}. \end{cases}$$

Step 1: Fix x. Suppose first that $x \neq 0$. Then $\exists N \in \mathbb{N}$ such that $0 < \frac{1}{N} < x$ (Archimedean Property). This implies $f_n(x) = 0$ for all $n \geqslant N$. Therefore $f_n(x) \to 0$ as $n \to \infty$ whenever $x \neq 0$. If $x = 0$, then $f_n(0) = 1$ and $f_n(0) \to 1$.

We deduce that, as we claimed in 5.2, the pointwise limit indeed exists and equals f, where

$$f(x) = \begin{cases} 1 & \text{if } x = 0, \\ 0 & \text{if } x > 0. \end{cases}$$

Steps 2 & 3: Now fix n and calculate α_n.

$$\alpha_n := \sup_{x \in [0,1]} |f_n(x) - f(x)| = \sup_{x \in (0,1/n]} |-nx + 1| = 1.$$

Trivially, $\alpha_n \to 1 \neq 0$. Hence (f_n) is not uniformly convergent.

Of course the contrapositive of Theorem 5.4 gives an alternative proof that convergence cannot be uniform.

(2) Let $E = [0, 1)$ and let $f_n(x) = x^n$.

Step 1: For fixed $x \in [0, 1)$, we have $x^n \to 0$ as $n \to \infty$. Hence the pointwise limit is $f = 0$.

Step 2 & 3: Trivially, $\alpha_n = \sup_{x \in [0,1)} |x^n| = 1$. To see this, observe that x^n is a monotonic increasing function on $[0, 1)$ and $x^n \to 1$ as $x \to 1$. Hence $\alpha_n \not\to 0$ so convergence is not uniform.

Now consider what happens if, with f_n as before, we work on $[0, b]$, where b is a constant with $0 \leqslant b < 1$. The pointwise limit is unchanged but now

$$\alpha_n = \sup_{x \in [0,b]} x^n = b^n \to 0 \text{ as } n \to \infty.$$

Hence convergence is uniform on $[0, b]$ for each fixed $b < 1$.

This example highlights that uniform convergence, or not, depends on the set E. It makes no sense to say '(f_n) converges uniformly' without specifying the set E on which the functions are considered.

(3) Let $E = [0, 1]$ and let

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

Step 1: Clearly $\lim_{n\to\infty} f_n(x) = 0$ for every $x \in [0, 1]$.

But $f_n(1/n) = 1/2$, so that

$$\sup_{x \in [0,1]} |f_n(x) - f(x)| \geqslant \frac{1}{2} \not\to 0 \text{ as } n \to \infty$$

and so (f_n) converges to 0 but *not uniformly* on $[0, 1]$.

Sketch diagrams of the graphs of the functions f_n are instructive.

(4) Consider $f_n(x) := nx^3\mathrm{e}^{-nx^2}$ on $[0,1]$.

Step 1: Fix x. For $x=0$, $f_n(x)=0$ for all n. For $x \neq 0$ we have, from the exponential series,

$$0 \leqslant f_n(x) = nx^3/(1 + nx^2 + \frac{(nx^2)^2}{2!} + \cdots) \leqslant \frac{2}{nx} \to 0 \text{ as } n \to \infty.$$

So, by sandwiching, $f_n(x) \to 0$ and this is true for $x=0$ too, trivially.

Step 2: Fix n and compute $\alpha_n := \sup\{\, nx^3\mathrm{e}^{-nx^2} \mid x \in [0,1]\,\}$. We have

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(nx^3\mathrm{e}^{-nx^2}\right) = 3nx^2\mathrm{e}^{-nx^2} - 2n^2x^4\mathrm{e}^{-nx^2}$$

and this is zero when $x=0$ (giving a minimum) and when $2nx^2=3$ (giving a maximum). Hence

$$\alpha_n = nx^3\mathrm{e}^{-nx^2}\Big|_{x=\sqrt{3/2n}} = C\sqrt{3/2n},$$

where C is a constant independent of n.

Step 3: From Step 2, $\alpha_n \to 0$ as $n \to \infty$. Therefore $f_n \xrightarrow{\mathrm{u}} 0$ on $[0,1]$.

It is instructive to draw the graphs of the first few functions in the sequence (f_n).

(5) [The partial sums of the geometric series] On $E = (-1,1)$ consider (f_n) given by

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

Step 1: Fix x with $|x|<1$ and let $n \to \infty$. Then $f_n(x) \to f(x) := 1/(1-x)$.

Step 2: Fix n. Here

$$\left\{\left|\frac{1-x^{n+1}}{1-x} - \frac{1}{1-x}\right| \mid |x|<1\right\} = \left\{\frac{|x|^{n+1}}{1-x} \mid |x|<1\right\}$$

is not bounded above. To see this, consider what happens at $x = (n+1)/(n+2)$.

[Alternative proof: Let $u = 1-x$ so we need to consider $(1-u)^{n+1}/u$ as $u \to 1-$. Expanding the numerator by the binomial theorem (remember n is fixed) and using an (AOL) argument, we see that $(1-u)^{n+1}/u \to \infty$ as $u \to 1-$. Hence Proposition 5.5 implies convergence is not uniform on $(-1,1)$.]

For more on this important example see 5.14 and 5.20.

5.8. Another occurrence of iterated limits: integrals and limits.

We now pick up again the theme introduced in 5.2.

You were almost certainly introduced at school to a definite integral $\int_a^b f(x)\,\mathrm{d}x$ representing the area under the graph of f and would have seen this calculated approximately using areas of rectangles. In Analysis III in Trinity Term you'll revisit Integration: you'll be given a formal definition of $\int_a^b f(x)\,\mathrm{d}x$ for suitable functions $f\colon [a,b] \to \mathbb{R}$, including continuous functions, and be shown that integration and differentiation are connected in the way you expect (for well-behaved functions anyway).

The point here is that you should think of integration as a limiting process. Hence it is natural to be sceptical as to whether, for a pointwise convergent sequence (f_n) of continuous functions $f_n\colon [a,b] \to \mathbb{R}$,

$$\lim_{n\to\infty} \int_a^b f_n(x)\,\mathrm{d}x \quad \text{and} \quad \int_a^b \lim_{n\to\infty} f_n(x)\,\mathrm{d}x$$

are equal.

As we shall now show, this is not always the case.

5.9. **Examples: integration and limits.**

Consider the following examples of sequences of functions. Here we shall stray a little beyond school calculus in places, making the commonsense assumption that we should define

$$\int_0^1 g(x)\,\mathrm{d}x = c(b-a)$$

when g takes the constant value c on a bounded subinterval of $\mathbb{R}$ with endpoints a, b where $a \leqslant b$ (the interval is allowed to include or exclude either endpoint), and $g(x) = 0$ otherwise.

We leave as exercises the detailed verifications of the claims we make.

(1) Let $f_n(x) = (n+x)^{-2}$ on $[0,1]$. For each fixed x, $\lim_{n\to\infty} f_n(x) = 0$. By school calculus,

$$\lim_{n\to\infty}\int_0^1 f_n(x)\,\mathrm{d}x = \lim_{n\to\infty} -(n+x)^{-1}\Big|_0^1 = 0 \quad\text{and}\quad \int_0^1 \lim_{n\to\infty} f_n(x)\,\mathrm{d}x = 0.$$

In this case $f_n \overset{\mathrm{u}}{\to} 0$ on $[0,1]$: $\alpha_n = 1/n^2$.

(2) Let $f_n(x) = x^n$ on $[0,1]$. Then $f_n(x) \to 0$ as $n \to \infty$ for each $x \neq 1$ and $f_n(1) = 1 \to 1$. Hence

$$\lim_{n\to\infty}\int_0^1 f_n(x)\,\mathrm{d}x = \lim_{n\to\infty} \frac{x^{n+1}}{n+1}\Big|_0^1 = 0 \quad\text{and}\quad \int_0^1 \lim_{n\to\infty} f_n(x)\,\mathrm{d}x = 0.$$

We already know that convergence of (f_n) is not uniform on $[0,1)$ and hence it cannot be uniform on $[0,1]$ either.

(3) Let $f_n \colon [0,1] \to \mathbb{R}$ where $f_n(x) = n^{-1}$ on $((1/n - 1/n!), 1/n)$ and $f_n(x) = 0$ otherwise.
If $x = 0$ then $f_n(x) = 0$ for all n. If $0 < x \leqslant 1$ then $f_n(x) = 0$ for $n > 1/x$. Thus $\lim f_n(x) = 0$. Then

$$\lim_{n\to\infty}\int_0^1 f_n = \lim_{n\to\infty} \frac{1}{n\cdot n!} = 0 \quad\text{and}\quad \int_0^1 \lim_{n\to\infty} f_n(x)\,\mathrm{d}x = 0.$$

Here convergence is uniform on $[0,1]$: $\alpha_n = 1/n$.

(4) Let $f_n \colon [0,1] \to \mathbb{R}$ where $f_n(x) = n^2$ on $(0, 1/n)$ and 0 otherwise. As in (3) we have $f_n \to 0$ pointwise on $[0,1]$. This time

$$\lim_{n\to\infty}\int_0^1 f_n(x)\,\mathrm{d}x = \lim_{n\to\infty} n = \infty \quad\text{and}\quad \int_0^1 \lim_{n\to\infty} f_n(x)\,\mathrm{d}x = 0.$$

Here (f_n) does not converge uniformly on $[0,1]$: α_n is not defined.

(5) Let $f_n \colon \mathbb{R} \to \mathbb{R}$ be given by $f_n(x) = n^{-2}(n-x)$ on $[0,n]$ and $f_n(x) = 0$ otherwise. Clearly $f_n(x) \to 0$ for each $x \in \mathbb{R}$ and $\alpha_n = n^{-1}$. Hence $f_n \overset{\mathrm{u}}{\to} 0$ on $\mathbb{R}$. But

$$\int_0^n n^{-2}(n-x)\,\mathrm{d}x = n^{-2}(nx - \frac{1}{2}x^2)\Big|_0^n \to \frac{1}{2}.$$

We observe from these examples that a limit and an integral may or may not commute:

- Example (4) provides a striking example of limit and integral not commuting.
- We also note that in examples (1), (2) and (3), limit and integral do commute. We deduce from (2) that uniform convergence is not a necessary condition for this to happen.
- We have not provided an example of a sequence of functions on a closed bounded interval for which convergence is uniform and limit and integral do not commute. There is a general result which underlies this, and which is easy to prove from well-known properties of integrals.
- From (5) we see that uniform convergence is not a sufficient condition for limit and integral to commute when we deal with a sequence of functions on an unbounded set.

5.10. **A wider perspective: iterated limiting processes more generally.**

In Analysis III in Trinity Term, you will learn how to define integrals using a limiting process, and the proofs of the following theorems will be given.

Assume $f_n \xrightarrow{u} f$ *on* $[a, b]$ *and that each* f_n *is continuous, then*

$$\int_a^b f(x)\,\mathrm{d}x = \int_a^b \lim_{n\to\infty} f_n(x)\,\mathrm{d}x = \lim_{n\to\infty} \int_a^b f_n(x)\,\mathrm{d}x.$$

Likewise, we may consider series, via their sequences of partial sums (see 5.12 for details). Then we have the following result.

If the series $\sum_{k=1}^\infty u_k$ *converges uniformly on* $[a, b]$ *and if all* u_k *are continuous, then we may integrate the series term-by-term:*

$$\int_a^b \sum_{k=1}^\infty u_k(x)\,\mathrm{d}x = \sum_{k=1}^\infty \int_a^b u_k(x)\,\mathrm{d}x.$$

Thus, for continuous functions on a closed bounded interval, uniform convergence provides a *sufficient* condition for interchanging an integral and either a limit or an infinite sum: instances of two limiting processes commuting.

The proofs of these theorems will turn out to be easy, but the conditions are stringent. Often we encounter non-uniform convergence, and working only with continuous functions on closed bounded intervals is very restrictive. The Part A option on (Lebesgue) Integration supplies tools which are much more powerful and are applicable for a much wider class of 'integrable' functions.

Before leaving integration we should also mention that applied courses make considerable use of 'multiple integrals'. This is an another situation in which two limiting processes are in play at the same time—and they might not commute. The examples in Prelims Applied will involve well-behaved functions and you can expect that you'll get the same answer whichever variable you integrate with respect to first. But this need not be true in general. Again, you need to wait for Part A Integration to hear the full story, with some interesting surprises.

Differentiation is another limiting process, and we shall want to know whether it commutes with other limiting processes, like taking the limit of a sequence of functions, or forming an infinite sum of functions. You were already introduced to the Differentiation Theorem for power series in Analysis I, without proof: for a real power series with radius of convergence R we can legitimately 'differentiate term-by-term' at each point in its interval of convergence $(-R, R)$. We shall prove this theorem later on. For now we note that, as is verified in Theorem 5.18, it will be important that a power series converges uniformly on any proper closed subinterval of $(-R, R)$.

Just as we found for sequences of real numbers, there is a characterisation of uniform convergence which does not depend on knowing the limit function. This will be essential to have when we consider series, since we often use infinite series to *define* functions.

5.11. **Theorem (Cauchy criterion for uniform convergence of a sequence).** *Let* $E \subseteq \mathbb{R}$ *and let* $f_n \colon E \to \mathbb{R}$ *for* $n = 1, 2, \ldots$. *Then* (f_n) *converges uniformly on* E *if and only if*

$$\forall \varepsilon > 0\ \exists N \in \mathbb{N}\ \forall n, m \geqslant N\ \forall x \in E \quad |f_n(x) - f_m(x)| < \varepsilon.$$

Proof. ($\Longrightarrow$): Suppose (f_n) converges uniformly on E with limit function f. Then

$$\forall \varepsilon > 0\ \exists N \in \mathbb{N}\ \forall n > N\ \forall x \in E \quad |f_n(x) - f(x)| < \tfrac{1}{2}\varepsilon.$$

So,
$\forall \varepsilon > 0 \; \forall n, m > N \; \forall x \in E \quad |f_n(x) - f_m(x)| \leqslant |f_n(x) - f(x)| + |f_m(x) - f(x)| < \frac{1}{2}\varepsilon + \frac{1}{2}\varepsilon = \varepsilon.$
Hence the uniform Cauchy condition holds.

($\Longleftarrow$) Suppose the uniform Cauchy condition holds. Then for each $x \in E$, $(f_n(x))$ is a Cauchy sequence in $\mathbb{R}$, so that it is convergent. Let us denote its limit by $f(x)$. Now

$$\forall \varepsilon > 0 \; \exists N \in \mathbb{N} \; \forall n, m > N \; \forall x \in E \quad |f_n(x) - f_m(x)| < \varepsilon/2.$$

Fix $n > N$ and $x \in E$, and let $m \to \infty$ in the above inequality. By the preservation of weak inequalities,

$$|f_n(x) - f(x)| = \lim_{m \to \infty} |f_n(x) - f_m(x)| \leqslant \varepsilon/2 < \varepsilon.$$

Hence $f_n \to f$ uniformly on E. □

[**A (very) technical note**: Observe the two different reasons why we introduced $\varepsilon/2$'s in the above proof in order to arrive at $< \varepsilon$ at the end.]

An important role of the Cauchy condition for real sequences comes from its application to series: it allows us to prove that absolute convergence implies convergence. Likewise, the uniform Cauchy condition is important in respect of series of functions.

5.12. **Uniform convergence of series of functions.**

As usual, we handle a series by considering its sequence of partial sums. Accordingly, given a sequence of real-valued functions (u_k) on a set $E \subseteq \mathbb{R}$ we say that the series $\sum u_k$ **converges pointwise** (**uniformly**) **on** E if (f_n) converges pointwise (uniformly) on E, where

$$f_n := u_1 + \cdots + u_n = \sum_{k=1}^{n} u_k.$$

[Note: when working both with individual terms of the series and with the partial sums of the series it is sensible to use different dummy variables: we use k for the first and n for the second.]

Assume each u_k is continuous on E. Then each f_n is also continuous on E. As a corollary of Theorem 5.4 we deduce that if $\sum u_k$ converges uniformly on E then $\sum_{k=1}^{\infty} u_k$ is continuous on E.

5.13. **Corollary to** 5.11 **(uniform Cauchy condition for series).** *Let (u_k) be a sequence of real-valued functions on E. Then the series $\sum u_k$ converges uniformly on E if and only if*

$$\forall \varepsilon > 0 \; \exists n \in \mathbb{N} \; \forall n > m \geqslant N \; \forall x \in E \quad |u_{m+1}(x) + \cdots + u_n(x)| < \varepsilon.$$

Proof. Apply 5.11 to the sequence (f_n) of partial sums given by $f_n = \sum_{k=1}^{n} u_k$. □

5.14. **Example 5.7(4) revisited.**

We apply 5.13 with $u_k = x^{k-1}$ on $E = (-1, 1)$. Take $n = 2m$. then

$$\begin{aligned} |u_{m+1}(x) + \cdots + u_{2m}(x)| &= |x^m + \cdots + x^{2m-1}| \\ &\geqslant (2m - m)x^{2m-1} \qquad \text{(no. of terms} \times \text{smallest term).} \end{aligned}$$

Given $\varepsilon > 0$ it is impossible to find N such that the LHS is $< \varepsilon$ for all $x \in (-1, 1)$ and all $m > N$. [You'll have seen similar arguments in Analysis I.] We have the same conclusion as before: the series does not converge uniformly on $(-1, 1)$.

There is a user-friendly sufficient condition for uniform convergence of a series. It is not a *necessary* condition.

5.15. **Weierstrass' M-test.** *The series $\sum u_k$ converges uniformly on E if there exist real constants M_k such that*

$$\forall k \ \forall x \in E \quad |u_k(x)| \leqslant M_k \quad \textit{and} \quad \sum M_k \textit{ converges.}$$

It is critically important in the M-test that $\sum M_k$ is a convergent series of *constants*: M_k must be independent of x as x varies over E.

Proof. We invoke the Cauchy criterion for $\mathbb{R}$ (0.8). A sequence (x_n) of real numbers converges implies it satisfies the Cauchy condition:

$$\forall \varepsilon > 0 \ \exists N \in \mathbb{N} \ \forall m, n > N \quad |x_m - x_n| < \varepsilon.$$

Let $f_n := u_1 + \cdots + u_n$. For each fixed $x \in E$ and $n > m$,

$$|f_m(x) - f_n(x)| = |u_{m+1}(x) + \cdots + u_n(x)| \leqslant M_{m+1} + \cdots + M_n \to 0 \text{ as } m, n \to \infty$$

by the Cauchy condition applied to the partial sums of the real series $\sum M_k$. Hence $(f_n(x))$ satisfies the Cauchy condition and so converges, to $f(x)$ say. Thus the series $\sum u_k$ converges pointwise.

To check that convergence is uniform, take the limit as $n \to \infty$ in the displayed line (with x and m fixed) to get

$$\forall m > N \ \forall x \in E \ |f_m(x) - f(x)| \leqslant \sum_{k=m+1}^{\infty} M_k \to 0 \text{ as } n \to \infty. \qquad \square$$

5.16. **Examples: M-test.**

On $E = [0,1]$ and for $k \geqslant 1$, let $u_k(x) = \dfrac{x^p}{1 + k^2x^2}$ where p is a constant.

(1) [A simple case] Assume $p \geqslant 2$). Then, for or $x \in [0,1]$,

$$|u_k(x)| \leqslant \frac{x^{p-2}}{k^2} \leqslant M_k := \frac{1}{k^2}.$$

Since $\sum k^{-2}$ converges, $\sum u_k(x)$ converges uniformly on $[0,1]$ by the M-test.

(2) [Optional: a harder example, aimed at more intrepid readers] Now assume $1 < p < 2$. The choice of M_k we used in (1) no longer works. Note that $u_k(x) \geqslant 0$. For fixed k, let's find the maximum value of $u_k(x)$ on $[0,1]$ by differentiation. We have

$$u_k'(x) = \frac{px^{p-1}(1 + k^2x^2) - 2k^2}{(1 + k^2x^2)^2}$$

and we see that the maximum of u_k on $[0,1]$ is achieved at $x_k \in [0,1]$ where $k^2x_k^2 := p/(2-p)$. We deduce that, for all $x \in [0,1]$,

$$0 \leqslant u_k(x) \leqslant u_k(x_k) \leqslant M_k := C\frac{1}{k^p},$$

where C is a positive constant depending on p but *independent of* x. The series $\sum 1/k^p$ converges for $p > 1$ by the Integral Test. Hence $\sum u_k$ converges uniformly on $[0,1]$ by the M-test.

Note: This method is useful when it works, but is not infallible. It investigates the maximun of each term separately rather than of the expression arising in the uniform Cauchy condition.

For Analysis II the most important examples of series of functions are (real) power series, so we now prioritise these.

5.17. **Power series: recap.**

We recall from Analysis I that the **radius of convergence** of a real power series $\displaystyle\sum_{k=0}^{\infty} c_k x^k$ is given by

$$R = \begin{cases} \infty & \text{if } \sum |c_x x^k| \text{ converges for all } x \in \mathbb{R}, \\ \sup\{\, |x| \mid \sum |c_x x^k| \text{ converges}\,\} & \text{otherwise.} \end{cases}$$

(Here, as usual, the symbol ∞ is a convenient shorthand.)

The following properties follow (with a little technical work required ((1) is not immediately obvious) and use of 'absolute convergence implies convergence' and its contrapositive):

(1) $\sum |c_k x^k|$ converges pointwise for every x in the **interval of convergence** $(-R, R)$;

(2) $\sum c_k x^k$ converges for $|x| < R$;

(3) $\sum c_k x^k$ diverges for $|x| > R$.

Accordingly, if $\sum c_k x^k$ has radius of convergence R then we have a well-defined function

$$f(x) = \sum_{k=0}^{\infty} c_k x^k \qquad \text{for } x \in (-R, R).$$

[Aside: We can likewise consider complex power series $\sum_{k=0}^{\infty} c_k z^k$, where c_k and z are now complex. We now have a **disc of convergence**, $\{\, z \in \mathbb{C} \mid |z| < R\,\}$ in which, pointwise, the series converges absolutely, and hence converges.]

Refer back to Analysis I notes for a full discussion of radius of convergence and how to calculate it for a given power series.

We now reach an important theorem.

5.18. **Theorem (power series: uniform convergence and continuity).** *Let* $\displaystyle\sum_{k=0}^{\infty} c_k x^k$ *be a real power series with radius of convergence* R*. Assume* $R > 0$*.*

(i) *Assume* R *is finite. Then for any* δ *with* $0 < \delta < R$, $\sum c_k x^k$ *converges uniformly on* $[-R + \delta, R - \delta]$.

Assume $R = \infty$*. Then* $\sum c_k x^k$ *converges uniformly on any bounded subinterval of* $\mathbb{R}$.

(ii) $f(x) := \displaystyle\sum_{k=0}^{\infty} c_k x^k$ *defines a continuous function* f *on* $(-R, R)$. debt paid!

Corresponding results hold for complex power series.

Proof. (i) We assume R is finite, leaving the case $R = \infty$ as an easy exercise. We apply the M-Test with $M_k := |c_k(R - \delta)^k|$:

$$x \in [-R + \delta, R - \delta] \implies |c_k x^k| \leqslant M_k$$

and $\sum M_k$ converges by 5.17(1).

Now consider (ii). Fix $p \in (-R, R)$ and choose $\delta > 0$ such that $\delta < R - |p|$. Then

$$-R + \delta < p < R - \delta.$$

By (i), $\sum c_k x^k$ converges uniformly on $E := [-R + \delta, R - \delta]$. Certainly each $u_k := c_k x^k$ is continuous on E and so too is each $f_n := \sum_{k=0}^{n} u_k$. We deduce from Theorem 5.4 that f is continuous on E and so is continuous in particular at p. Since p was *any* point in $(-R, R)$ we have proved (ii). □

5.19. **Corollary.** *The following functions, given by power series with infinite radius of convergence, are continuous on* $\mathbb{R}$*:*

debts paid!

$$\exp(x),\ \sin x,\ \cos x,\ \sinh x,\ \cosh x.$$

Functions derived from these via reciprocal and quotient, such as

$$\operatorname{cosec} x,\ \sec x,\ \tan x,\ \cot x$$

are continuous on any set on which the denominator is never zero.

Functions which can be derived from the above functions by application of the Continuous Inverse Function Theorem are themselves continuous. This includes $\log x$ *on* $(0, \infty)$ *and* $\arctan x$ *on* $(-\infty, \infty)$.

5.20. **Warnings!**

We cannot stress too strongly that Theorem 5.18 is subtle and needs applying with care. Let $\sum c_k x^k$ be a power series with radius of convergence $R > 0$.

- In general $\sum c_k x^k$ will *not* converge uniformly on $(-R, R)$. We have given two proofs earlier to show that $\sum x^k$, for which $R = 1$, fails to converge uniformly on $(-1, 1)$. Nevertheless, $\sum x^k$ does converge uniformly on any interval $[-\eta, \eta]$, where $0 < \eta < 1$: take $M_k = \eta^k$ in the M-Test.

  Uniform convergence of a sequence or series of functions on a specified domain E is a *global* condition: we consider behaviour on all points of E at the same time.
- Continuity of a function f on a set E is a *local* condition: we need f to be continuous at each *fixed* point p in E. It is only the behaviour of f close to p that determines whether or not f is continuous at p.

  Note the similarity here with the contortions we had to go through when we wanted to prove continuity of the inverse of a strictly monotonic continuous function when we wanted to apply the IFT but could not work on a single closed bounded interval. (Recall 3.20.)

Consider $(-R, R)$ with $R > 0$ and suppose that

$$(-R, R) = \bigcup_{m \geqslant 1} I_m, \quad \text{where } \{I_m\} \text{ is a sequence of proper subintervals.}$$

It is not true in general that uniform convergence on each I_m implies that convergence is uniform on $(-R, R)$.

It is true that a function which is continuous on each I_m is continuous on their union $(-R, R)$.

Example, revisited: $\sum x^k$ has $R = 1$. It converges uniformly on each

$$I_m := [-1 + m^{-1}, 1 - m^{-1}]$$

but does not converge uniformly on $\bigcup I_m = (-1, 1)$. We know that $\sum_{k=0}^{\infty} x^k = (1 - x)^{-1}$ on $(-1, 1)$ and this function is continuous on $(-1, 1)$. (Unusually, in this example, we can evaluate the infinite sum explicitly, and so confirm directly the theoretical result from Theorem 5.18(ii).)

5.21. **Further examples on uniform convergence of series, and continuity.**

(1) Consider the series $\sum_{k=0}^{\infty} x^k \cos(kx^2)$ on $E = [0,1)$. By the Comparison Test the series converges for each fixed $x \in [0,1)$.

For any η with $0 < \eta < 1$,

$$\forall x \in [0,\eta] \quad |x^k \cos(kx^2)| \leqslant M_k := |\eta|^k \text{ and } \sum M_k \text{ converges.}$$

By the M-test, the series converges uniformly on $[0,\eta]$.

We don't have a candidate for M_k which would show that the series is *uniformly* convergent on $[0,1)$. Nonetheless we claim that $f(x) = \sum_{k=0}^{\infty} x^k \cos(kx^2)$ defines a function f which is continuous on $[0,1)$. To do this, fix p with $0 \leqslant p < 1$ and choose $\eta > 0$ with $p < \eta < 1$. Then the series converges uniformly on $[0,\eta]$. Since each function $x^k \cos(kx^2)$ is continuous on $[0,\eta]$, Theorem 5.4 implies that f is continuous on $[0,\eta]$ and hence is continuous at p.

(2) Consider the series

$$\sum_{k=0}^{\infty} \frac{k^2 x}{1 + k^4 x^2}.$$

We claim that this converges uniformly on $[\delta, 1]$ for each δ with $0 < \delta < 1$. Let $M_k = k^{-2}\delta^{-1}$. Then, on $[\delta, 1]$.

$$|k^2 x/(1 + k^4 x^2)| \leqslant |k^2 x / k^4 x^2| \leqslant k^{-2}\delta^{-1} = M_k.$$

Since $\sum M_k$ converges, we do indeed have uniform convergence on each interval $[\delta, 1]$.

We shall now show that the series is not uniformly convergent on the interval $(0,1]$.

If the series were uniformly convergent, the uniform cauchy criterion would show that, for any $\varepsilon > 0$ there exists N such that for all $x \in (0,1]$, and all $n \geqslant N$,

$$\left| \sum_{n < k \leqslant n+1} \frac{k^2 x}{1 + k^4 x^2} \right| < \varepsilon$$

But if $\varepsilon = 1/2$ and $x = (n+1)^{-2}$ this would give $1/2 < 1/2$, a contradiction.

But, localising to a point $p \in (0,1]$ and choosing δ such that $0 < \delta < p$, we can prove that the series defines a function which is continuous on $(0,1]$.

---

6. Further results on continuity and uniform convergence

This section contains non-examinable material which goes beyond the syllabus for Prelims. It concerns continuous functions and monotonic functions and builds on and illuminates results we have proved already.

This supplementary material will not be covered in lectures. It has been included in the webnotes for two reasons. It introduces several important theorems which crop up in more advanced applications of Analysis. These don't find a natural place in later Analysis courses but should be known to well-educated students of mathematical analysis. Secondly, the results are interesting in their own right and add some spice to the basic treatment of continuity and uniform convergence presented so far.

We have shown that a real-valued function which is continuous and strictly monotonic increasing on a closed bounded interval behaves very well; recall the Continuous Inverse Function Theorem. But how do the notions of monotonicity and continuity interact? Recall that a real-valued function $f\colon E \to \mathbb{R}$ is **monotonic increasing** (resp. **decreasing**) if $x \leqslant y$ implies

$f(x) \leqslant f(y)$ (resp. $f(x) \geqslant f(y)$). Remember that any result about increasing functions can be translated into a result about decreasing functions by considering $-f$ instead of f.

6.1. **Theorem (left-hand and right-hand limits of a monotonic increasing function).** *Let $f\colon (a, b) \to \mathbb{R}$ be increasing. Then for every $p \in (a, b)$ the right-hand limit $f(p+)$ and the left-hand limit $f(p-)$ of f at p exist.*

Moreover, $f(p-) = \sup_{a<x<p} f(x)$ and $f(p+) = \inf_{p<x<b} f(x)$, and

$$f(p-) \leqslant f(p) \leqslant f(p+).$$

Proof. [*Notice the parallel with the proof of the Monotonic Sequences Theorem from Analysis I: the proof we require here is a function limit version of that argument.*]

The set $\{\, f(x) \mid a < x < p \,\}$ is non-empty and is bounded above by $f(p)$ since f is increasing. Therefore by the Completeness Axiom $A := \sup_{a<x<p} f(x)$ exists and $A \leqslant f(p)$. We have to show that $f(p) = A$. Let $\varepsilon > 0$. By the Approximation Property for sup, there exists $x_\varepsilon \in (a, p)$ such that

$$A - \varepsilon < f(x_\varepsilon) \leqslant A.$$

As $p - x_\varepsilon > 0$ choose $\delta := p - x_\varepsilon$. Then, $x \in (x_\varepsilon, p)$ if and only if $0 < p - x < \delta$, and thus, as f is increasing,

$$A - \varepsilon < f(x_\varepsilon) \leqslant f(x) \leqslant A \quad \text{when } 0 < p - x < \delta.$$

By definition $f(p) = A$ and we are done.

The other inequality can be obtained by a similar argument (a good exercise); or by applying what we have done to the function $-f(b-x)$ on $(0, b-a)$ and juggling with the inequalities. □

We say that f, as in Theorem 6.1, has a **jump discontinuity at** p if $f(p+) - f(p-)$ is non-zero and then, informally, we refer to $f(p+) - f(p-)$ as the **jump** of f at p.

We remark that it can also be proved that an increasing function on a bounded interval (a, b) can be shown to have a set of jump discontinuities which is finite or countably infinite.

Finally we stress that the behaviour of increasing functions is very special. Consider, for example, $f\colon (0, 1) \to \mathbb{R}$ given by

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

Then the left-hand and right-hand limits $f(p-)$ and $f(p+)$ fail to exist for every $p \in (0, 1)$. Moreover f is discontinuous at every point of the uncountable set $(0, 1)$.

Our next extension topic concerns monotonic sequences of continuous functions. We obtain a partial converse to Theorem 5.4 which asserted that 'uniform convergence preserves continuity'. We shall show that, on a closed bounded interval, if the sequence is monotonic then the continuity of the limit will give uniformity of convergence. The proof is unappealingly technical. A slicker proof, and in a more general setting than for functions on an interval $[a, b]$, can be given later, once ideas about compactness from Part A Metric Spaces are available.

6.2. **Dini's Theorem.** *Let (f_n) be a sequence of real-valued continuous functions on $[a, b]$ and let f be a real-valued continuous function on $[a, b]$. Assume that*

$$\lim_{n\to\infty} f_n(x) = f(x) \qquad \text{for every } x \in [a, b]$$

and that

$$f_n(x) \geqslant f_{n+1}(x) \qquad \text{for all } n \quad \text{and for all } x \in [a, b].$$

Then $f_n \to f$ uniformly on $[a, b]$.

Proof. Let $g_n(x) = f_n(x) - f(x)$. Then each g_n is continuous, $g_n \geqslant 0$ and $\lim_{n\to\infty} g_n(x) = 0$ for any $x \in [a, b]$.

Suppose (g_n) were not uniformly convergent on $[a, b]$. Write down the contrapositive to see that for some $\varepsilon > 0$, and every $k \in \mathbb{N}$ there exists $n_k \in \mathbb{N}$ with $n_k > k$ and some $x_k \in [a, b]$ such that

$$|g_{n_k}(x_k)| = g_{n_k}(x_k) \geqslant \varepsilon.$$

We may choose n_k so that the map $k \mapsto n_k$ is increasing. We may assume that (x_k) tends to some point p—otherwise use the Bolzano–Weierstrass theorem to extract a convergent subsequence of (x_k) and use it instead. Then $p \in [a, b]$. For any (fixed) k, since (g_n) is decreasing,

$$\varepsilon \leqslant g_{n_\ell}(x_\ell) \leqslant g_{n_k}(x_\ell)$$

for all $\ell > k$. Letting $\ell \to \infty$ in the above inequality, we obtain

$$\varepsilon \leqslant \lim_{\ell\to\infty} g_{n_k}(x_\ell) = g_{n_k}(p)$$

since g_{n_k} is continuous at p. This contradicts the assumption that $\lim_{k\to\infty} g_{n_k}(p) = 0$. □

6.3. **Example.**

Let $f_n(x) = \frac{1}{1+nx}$ for $x \in (0, 1)$. Then $f_n(x) \to 0$ for every $x \in (0, 1)$, Also (f_n) is decreasing, but (f_n) does not converge uniformly. Dini's theorem doesn't apply here, as $(0, 1)$ is not compact (compact $\equiv$ closed + bounded, for subsets of $\mathbb{R}$).

There are a number of different proofs of the following famous approximation theorem and many generalisations of it exist. The conclusion of the theorem is striking, given that uniform convergence is a strong condition, and polynomials are very amenable functions, whereas continuous functions can be wild enough not to be differentiable at any point. It would take us too far afield to include a proof of the theorem here but we remark that in one approach Dini's Theorem is used along the way (applied to an ancillary sequence of polynomials approximating $\sqrt{x}$ on $[0, 1]$).

6.4. **Weierstrass's Polynomial Approximation Theorem.** *Assume that $[a, b]$ is a closed bounded subinterval of R and that $f\colon [a, b] \to \mathbb{R}$ is continuous. Then there exists a sequence (P_n) of real polynomials such that $P_n \xrightarrow{\text{u}} f$.*

Our final extension topic involves power series and continuity.

6.5. **Endpoints of the interval of convergence of a real power series.**

You have been aware ever since you first studied the convergence of power series in Analysis I that a real power series $\sum_{k=0}^{\infty} c_k x^k$ with finite non-zero radius of convergence R converges absolutely for any x for which $|x| < R$. You also saw examples which show that the series may converge absolutely, may convergence non-absolutely, or may diverge, at each of the points $x = R$ and $x = -R$.

We showed in Section 5 that $f(x) := \sum_{k=0}^{\infty} c_k x^k$ defines a continuous function f on $(-R, R)$, irrespective of how the series behaves at $\pm R$.

We now investigate the behaviour of a real power series at the endpoints of its interval of convergence, assuming without loss of generality that $R = 1$. The theorem we prove has a variety of applications, notably in probability theory.

6.6. **Abel's Continuity Theorem.** *Assume that the real power series* $\sum_{k=0}^{\infty} c_k x^k$ *has radius of convergence* $R = 1$. *Assume further that* $\sum_{k=0}^{\infty} c_k$ *converges. Then*

$$\lim_{x \to 1-} \sum_{k=0}^{\infty} c_k x^k = \sum_{k=0}^{\infty} c_k.$$

Therefore f *given by* $f(x) = \sum_{k=0}^{\infty} c_k x^k$ *is continuous on* $(-1, 1]$, *with left-continuity at* 1.

Proof. The proof we give makes use of a standard theorem about multiplication of absolutely convergent series[1]. Applied to power series this theorem gives

$$\left(\sum a_r x^r\right)\left(\sum b_s x^s\right) = \sum \left(a_0 b_k + a_1 b_{k-1} + \cdots + a_k b_0\right) x^k.$$

(Intuitively the idea is to collect together all terms in $x^r x^s$ for which $r + s = k$. What is not obvious is that this reorganisation does not affect the convergence of the series or its sum.)

For $|x| < 1$ we know that

$$f(x) := \sum_{k=0}^{\infty} c_k x^k,$$

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k,$$

with both series converging absolutely. Hence for $|x| < 1$ the multiplication formula gives

$$\frac{f(x)}{1-x} = \sum \big(c_0 + c_1 + \cdots + c_k\big) x^k = \sum s_k x^k,$$

where $s_k := \sum_{m=0}^{k} c_m$—the partial sum of $\sum c_m$ obtained by adding from the first $k+1$ terms. Let $s = \sum_{m=0}^{\infty} c_m$. Now

$$f(x) - s = (1-x) \sum_{k=0}^{\infty} \big(s_k - s\big) x^k.$$

Then

$$\forall \varepsilon > 0 \ \exists N \in \mathbb{N} \ \forall k \geqslant N \quad |s_k - s| < \varepsilon/2$$

and

$$\exists C > 0 \ \forall k \in \mathbb{N} \quad |s_k - s| \leqslant C.$$

Now let $0 < x < 1$. We get

$$\begin{aligned} |f(x) - s| &< (1-x) \left(\sum_{k=0}^{N-1} C x^k + \frac{\varepsilon}{2} \sum_{k=N}^{\infty} x^k \right) \\ &< (1-x) \left(CN + \frac{\varepsilon}{2} \cdot \frac{x^N}{1-x} \right) \\ &< (1-x) CN + \frac{\varepsilon}{2} \\ &< \varepsilon \end{aligned}$$

if $0 < 1 - x < \varepsilon/(2CN)$. We conclude that $f(x) \to s$ as $x \to 1-$. □

We remark that, with the aid of the Cauchy condition for uniform convergence, it is possible also to prove that the series defining f converges uniformly on any interval $[b, 1]$, where $-1 < b < 1$. This is true by the M-test if $\sum |c_k|$ converges so the interesting case is that where $\sum c_k$ converges non-absolutely.

[1] See for example T.M. Apostol, *Mathematical Analysis*, 2nd edn, Section 8.24 or W. Rudin, *Principles of Mathematical Analysis*, Chapter 4; other Analysis texts also discuss multiplication of series.

7. Differentiability: the basics

In this section we look at differentiation, making use of the machinery of function limits with which Analysis II began. We rediscover all the familiar differentiation rules from school calculus and start to explore examples of functions which are and are not differentiable. Major theorems on differentiable functions come in the next section.

We shall restrict attention to real-valued functions defined on intervals in $\mathbb{R}$. Many of the results we obtained in earlier sections for real-valued functions of a real variable have obvious analogues when $\mathbb{R}$ is replaced by $\mathbb{C}$. But the theory of differentiability of complex valued functions on the complex plane turns out to be very different from that in the real case and is much more powerful: Complex Analysis is covered within the Part A Core. (We make just a brief excursion into the setting of $\mathbb{C}$ when we come to prove the Differentiation Theorem for power series.)

7.1. Definitions (differentiability notions).

Let $f\colon (a,b) \to \mathbb{R}$, and let $x_0 \in (a,b)$. By saying f **is differentiable at** x_0 we mean that the following limit exists:

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}.$$

When it exists we denote the limit by $f'(x_0)$ and we call it the **derivative of** f **at** x_0**.** In symbols,

$$\forall \varepsilon > 0 \; \exists \delta > 0 \; \forall x \in (a,b) \; \left(0 < |x - x_0| < \delta \implies \left| \frac{f(x) - f(x_0)}{x - x_0} - f'(x_0) \right| < \varepsilon \right).$$

[Note that the restriction to $0 < |x - x_0|$ ensures that the quotient makes sense.] We say that f **is differentiable on** (a,b) if f is differentiable at every point of (a,b).

Alternative notations: We shall, as convenient, adopt the various different ways of writing derivatives with which you'll be already familiar: for a function $y = y(x)$ differentiable on (a,b):

$$y' \qquad \text{or} \qquad \frac{\mathrm{d}y}{\mathrm{d}x} \qquad \text{or} \qquad \frac{\mathrm{d}}{\mathrm{d}x} y(x).$$

Sometimes it is helpful tor necessary o consider one-sided versions of derivatives. Let $f\colon [a,b) \to \mathbb{R}$ and let $x_0 \in [a,b)$. We say that f **has a right-derivative** at x_0 if

$$\lim_{x \to x_0+} \frac{f(x) - f(x_0)}{x - x_0}$$

exists. When it does, we denote it by $f'((x_0+)$. (Alternative notation: $f'_+(x_0)$.) Now let $f\colon (a,b] \to \mathbb{R}$ and let $x_0 \in (a,b]$. We say that f **has a left-derivative** at x_0 if

$$\lim_{x \to x_0-} \frac{f(x) - f(x_0)}{x - x_0}$$

exists. In this case we denote the limit by $f'(x_0-)$. (Alternative notation: $f'_-(x_0)$.)

For a function $f\colon [a,b] \to \mathbb{R}$ which is differentiable on (a,b) we may ask whether $f'(b-)$,and $f'(a+)$ also exist. When this is so we say that f is **differentiable on** $[a,b]$ **(with one-sided derivatives at the endpoints)**.

The following result is easily proved (compare what we did for left- and right-continuity). It is useful for showing differentiability/non-differentiability of certain functions given by 'by cases' definitions at those points at which the specification changes. See Examples 7.3.

7.2. **Proposition.** *Let $f\colon (a,b) \to \mathbb{R}$. Then the following are equivalent:*

(a) *f is differentiable at x_0;*

(b) *f has both left- and right-derivatives at x_0, and $f'(x_0-) = f'(x_0+)$.*

7.3. **First examples.**

(1) It is immediate that f given by $f(x) = x$ is differentiable on $\mathbb{R}$.

(2) Consider $f(x) = |x|$ on $\mathbb{R}$. Here f is differentiable at any $x_0 \neq 0$. At 0 we have one-sided derivatives:

$$f'(0+) = 1 \quad \text{and} \quad f'(0-) = -1;$$

By the contrapositive of Proposition 7.2, $f'(0)$ fails to exist. This example shows that a function which is continuous at a point x_0 need not be differentiable at x_0.

(3) Define $f\colon \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} x^2 & \text{for } x > 0, \\ 0 & \text{for } x \leqslant 0. \end{cases}$$

Then $f'(0-)$ exists and equals 0, obviously. Also

$$f'(0+) = \lim_{x \to 0+} \frac{x^2 - 0}{x - 0} = 0.$$

Hence, by Proposition 7.2, $f'(0)$ exists and equals 0. Alternatively, we can give a direct sandwiching argument:

$$\left| \frac{f(x) - f(0)}{x - 0} - 0 \right| \leqslant \left| \frac{x^2}{x} \right| \to 0 \text{ as } x \to 0.$$

We next present a reformulation of the definition of differentiability as a point. The central idea is to avoid the need for division and so the need to worry about the possibility that the denominator of a fraction might be zero.

7.4. **Lemma (alternative formulations of the differentiability definition).** *Let $f\colon (a,b) \to \mathbb{R}$ and let $x_0 \in (a,b)$. Then the following are equivalent:*

(a) $\displaystyle\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}$ *exists (and equals ℓ say);*

(b) *there exists a constant A such that*

$$f(x) = f(x_0) + A(x - x_0) + \varepsilon(x)\,(x - x_0) \quad \textit{where } \varepsilon(x) \to 0 \textit{ as } x \to x_0.$$

Hence, if $f'(x_0)$ exists, then

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \varepsilon(x)\,(x - x_0) \quad \textit{where } \varepsilon(x) \to 0 \textit{ as } x \to x_0.$$

Proof. In what follows x is assumed to lie in (a,b).

((a) $\implies$ (b)): Write

$$\varepsilon(x) := \frac{f(x) - f(x_0)}{x - x_0} - \ell.$$

Assumption (a) implies $\varepsilon(x) \to 0$ as $x \to x_0$. Rearranging shows that (b) holds with $A = \ell$.

((b) $\implies$ (a)): For $x \neq x_0$, divide by $x - x_0$ in (b). A simple (AOL) argument gives (a), with $\ell = A$.

The final statement is now immediate. $\square$

7.5. **Proposition (differentiability implies continuity).** *Let* $f\colon (a,b) \to \mathbb{R}$ *and* $x_0 \in (a,b)$. *Then* f *is differentiable at* x_0 *implies* f *is continuous at* x_0.

Proof. Immediate from Lemma 7.4 and (AOL) for function limits. $\square$

We have already shown (Example 7.3(2)) that the converse of Proposition 7.5 is not true.

Note: one-sided versions of the results above exist for a function on $[a,b]$, involving right-hand derivative at a and right-continuity, and likewise with left limits at b. This technical observation will be useful later when we want to apply Rolle's Theorem to successive derivatives, as in the proof of Taylor's Theorem, for example.

Now we start assembling the rules of differential calculus as you learned them at school, but now obtained as consequences of (AOL) for function limits.

7.6. **Theorem (differentiation: algebraic properties).** *Assume that* $f, g\colon (a,b) \to \mathbb{R}$ *are both differentiable at* $x_0 \in (a,b)$, *and that* $\lambda, \mu \in \mathbb{R}$. *Then the following hold.*

(i)] **Linearity of differentiation**: $\lambda \cdot f + \mu \cdot g$ *is differentiable at* x_0 *and*

$$(\lambda \cdot f + \mu \cdot g)'(x_0) = \lambda \cdot f'(x_0) + \mu \cdot g'(x_0).$$

(ii) **Product Rule:** $fg\colon x \mapsto f(x)g(x)$ *is differentiable at* x_0 *and*

$$(fg)'(x_0) = f(x_0)g'(x_0) + f'(x_0)g(x_0).$$

(iii) **Quotient Rule:** *Assume that* $g(x_0) \neq 0$. *Then* $x \mapsto \dfrac{f(x)}{g(x)}$ *is differentiable at* x_0 *and*

$$\left(\frac{f}{g}\right)'(x_0) = \frac{f'(x_0)g(x_0) - f(x_0)g'(x_0)}{g^2(x_0)}.$$

Proof. (i) The proof is a simple application of (AOL).

(ii) We give a proof based on Lemma 7.4 to illustrate its use. We have

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \varepsilon_1(x)(x - x_0) \quad \text{where } \varepsilon_1(x) \to 0 \text{ as } x \to x_0,$$
$$g(x) = g(x_0) + g'(x_0)(x - x_0) + \varepsilon_2(x)(x - x_0) \quad \text{where } \varepsilon_2(x) \to 0 \text{ as } x \to x_0.$$

We multiply and get

$$f(x)g(x) = f(x_0)g(x_0) + (f'(x_0)g(x_0) + f(x_0)g'(x_0))(x - x_0) + [g(x_0)\varepsilon_1(x) + f(x_0)\varepsilon_2(x) + \varepsilon_1(x)\varepsilon_2(x)(x - x_0)](x - x_0).$$

By standard (AOL) for function limits the expression in square brackets, $\eta(x)$ say, tends to 0 as $x \to x_0$. Now by Lemma 7.4 ((b) implies (a)), we deduce that fg is differentiable at x_0, with derivative $f'(x_0)g(x_0) + f(x)g'(x_0)$.

(iii) See Problem sheet 4, which seeks a proof establishing the rule for Reciprocal first, from which the general Quotient Rule can be derived by means of the Product Rule. $\square$

7.7. **Corollary: simple examples.** *The power* x^n *is differentiable at all points, for* $n \geqslant 0$, *and so are polynomials, and also rational functions at points where the denominator is non zero.*

7.8. **Higher Derivatives.** Suppose that $f\colon (a,b) \to \mathbb{R}$ is differentiable at every point of some $(x_0 - \delta, x_0 + \delta)$. Then it makes sense to ask if f' is differentiable at x_0. If it is differentiable then we denote its derivative by $f''(x_0)$.

We can seek to iterate this process. Suppose f, f',..., $f^{(n)}$ have been defined recursively at every point of (a,b) (we make this assumption to simplify matters). Then we say f **is** $(n+1)$**-times differentiable at** x_0 if $f^{(n)}$ is differentiable at x_0; we then write $f^{(n+1)}(x_0) := f^{(n)\prime}(x_0)$.

If f has derivatives of all orders on (a,b) (that is, $f^{(k)}(x_0)$ exists at each $x_0 \in (a,b)$, for all $k = 1, 2, \ldots$) we sometimes say it is **infinitely differentiable on** (a,b)**.**

The following is proved by an easy induction using Linearity and the Product Rule. (Compare with the proof of the binomial expansion of $(1+x)^n$ for n a positive integer.)

7.9. **Proposition (Leibniz' Formula).** *Let $f, g\colon (a,b) \to \mathbb{R}$ be n-times differentiable on (a,b). Then $x \mapsto f(x)g(x)$ is n-times differentiable and*

$$(fg)^{(n)}(x) = \sum_{j=0}^{n} \binom{n}{j} f^{(j)}(x) g^{(n-j)}(x).$$

7.10. **Theorem (Chain Rule).** *Assume that $f\colon (a,b) \to \mathbb{R}$, and that $g\colon (c,d) \to \mathbb{R}$. Suppose that $f((a,b)) \subseteq (c,d)$, so that $g \circ f\colon (a,b) \to \mathbb{R}$ is defined.*

Suppose further that f is differentiable at $x_0 \in (a,b)$, and that g is differentiable at $f(x_0)$.

Then $g \circ f$ is differentiable at x_0 and

$$(g \circ f)'(x_0) = g'(f(x_0))\, f'(x_0).$$

[Proof on separate page in landscape format.]

7.11. **Example.**

Let $f(x) = x^2 \cos(1/x)$ for $x \neq 0$ and $f(0) = 0$. We shall assume that cos and sin are differentiable with the expected derivatives. We know from Section 5 that they are continuous. (Their differentiability relies on the Differentiation Theorem for power series.). unpaid debt

On $\mathbb{R} \setminus \{0\}$ we can apply the standard differentiation rules, including the Chain Rule, and we get, for $\neq 0$,

$$f'(x) = 2x \cos(1/x) + \sin(1/x).$$

Now consider 0: for $x \neq 0$,

$$\left| \frac{f(x) - f(0)}{x - 0} \right| = |x \cos(1/x)| \leqslant |x| \to 0 \text{ as } x \to 0.$$

Therefore $f'(0)$ exists and equals 0.

Note that the formula for $f'(x)$ for $x \neq 0$ shows that $\lim_{x \to 0} f'(x)$ fails to exists (the first term tends to 0, the second one does not have a limit as $x \to 0$, so the sum cannot tend to a limit). We deduce that f' is not continuous at 0. By the contrapositive of Proposition 7.5, $f''(0)$ cannot exist. (Note that f does have derivatives of all orders on $\mathbb{R} \setminus \{0\}$.)

Like the other main results in this section, our final theorem tells us how to build new differentiable functions.

7.12. **Inverse Function Theorem: Differentiability of inverse function.** *Assume that $f\colon [a, b] \to [f(a), f(b)]$ is a strictly increasing continuous function from $[a, b]$ onto $[f(a), f(b)]$, with inverse function g mapping $[f(a), f(b)]$ onto $[a, b]$. Assume that f is differentiable at $x_0 \in (a, b)$ and that $f'(x_0) \neq 0$. Then g is differentiable at $f(x_0)$ and*

$$g'(f(x_0)) = \frac{1}{f'(x_0)}.$$

Proof. The statement includes all the assumptions we imposed for the IFT for continuous functions, 3.14. Hence there exists a continuous inverse $g\colon [(fa), f(b)] \to [a, b]$ for f. Moroever the proof showed that f maps any open subinterval of $[a, b]$ to an open interval (see Corollary 3.15). Hence when considering the existence of g' at $y_0 := f(x_0)$ we may assume that g is defined in an open interval containing y_0. We have

$$f(x) - f(x_0) = (f'(x_0) + \varepsilon(x))(x - x_0) \quad \text{where } \varepsilon(x) \to 0 \text{ as } x \to x_0.$$

Then, for suitably small $\delta > 0$,

$$|x - x_0| < \delta \implies \big[x \in (a, b) \text{ and } f'(x_0) + \varepsilon(x) \neq 0\big].$$

Now restrict to y such that $x = g(y)$ is such that $0 < |g(y) - g(y_0)| < \delta$. Then

$$g(y) - g(y_0) = x - x_0 = (y - y_0)/(f'(x_0) + \varepsilon(g(y))$$

Noting that continuity of g implies $x = g(y) \to g(y_0)$ as $y \to y_0$, we deduce by (AOL) that

$$\frac{g(y) - g(y_0)}{y - y_0} \to \frac{1}{f'(x_0)} \text{ as } y \to y_0.$$

□

7.13. **Examples of inverse functions revisited.**

unpaid debt

We refer back to 5.19. Still assuming the Differentiation Theorem for power series and its consequences for the elementary functions, we deduce that the following are differentiable and have the expected derivatives

$$g(u) = \log u \quad \text{on } (0, \infty); \qquad g'(u) = \frac{1}{u},$$

$$g(u) = \arctan u \quad \text{on } (-\pi/2, \pi/2); \qquad g'(u) = \frac{1}{1+u^2}.$$

To confirm the result for $g(u) = \log u$, note that, for fixed $u_0 \in (0, \infty)$, Theorem 7.12 can be applied with $f = \exp$ and $[a, b]$ such that $a < \log u_0 < b$. Write $x_0 = \log u_0$ so $u_0 = \exp(x_0)$. The formula in the theorem gives

$$g'(u_0) = \frac{1}{f'(x_0)} = \frac{1}{f(x_0)} = \frac{1}{u_0}.$$

The derivative of arctan is handled similarly, making use of standard trigonometric formulae which can be derived with the aid of the Differentiation Theorem.

Chain Rule

Proof. Below, $x \in (a,b)$ and $y \in f((a,b)) \subseteq (c,d)$. Let $y_0 = f(x_0)$. We have

(1) $f(x) - f(x_0) = f'(x_0)(x - x_0) + (x - x_0)\varepsilon(x)$ where $\varepsilon(x) \to 0$ as $x \to x_0$

(since f differentiable at x_0).

(2) $g(y) - g(y_0) = g'(y_0)(y - y_0) + (y - y_0)\eta(y)$ where $\eta(y) \to 0$ as $y \to y_0$

(since g differentiable at $y_0 = f(x_0)$).

In (2), substitute $y = f(x)$ and $y_0 = f(x_0)$ to get

(3) $g(f(x)) - g(f(x_0)) = [g'(f(x_0)) + \eta(f(x))][f(x) - f(x_0)]$ where $\eta(f(x)) \to 0$ as $f(x) \to f(x_0)$.

In (3), substitute for $f(x) - f(x_0)$ from (1) to get

(4) $g(f(x)) - g(f(x_0)) = [g'(f(x_0)) + \eta(f(x))][(f'(x_0) + \varepsilon(x))(x - x_0)]$.

Hence

(5) $g(f(x)) - g(f(x_0)) = g'(f(x_0))f'(x_0)(x - x_0) + \rho(x)\,(x - x_0)$ where $\rho(x) := g'(f(x_0))\varepsilon(x) + f'(x_0)\eta(f(x)) + \varepsilon(x)\eta(f(x))$.

It remains to show that $\rho(x) \to 0$ as $x \to x_0$. Since f is differentiable at x_0 it is continuous at x_0 (by Proposition 7.5) so that $f(x) \to f(x_0)$ as $x \to x_0$. The result now follows from (AOL). $\square$

it suffices to show that F maps

8. Rolle's Theorem and the Mean Value Theorem

This section contains major theorems on differentiable functions which will find applications throughout the remainder of the course. We work with real-valued functions exclusively.

8.1. **Definitions (local maxima and minima).**

Let $E \subseteq \mathbb{R}$ and $f\colon E \to \mathbb{R}$.

(i) $x_0 \in E$ **is a local maximum** if for some $\delta > 0$,

$$f(x) \leqslant f(x_0) \text{ when } x \in (x_0 - \delta, x_0 + \delta) \cap E.$$

(ii) $x_0 \in E$ **is a local minimum** if for some $\delta > 0$,

$$f(x) \geqslant f(x_0) \text{ when } x \in (x_0 - \delta, x_0 + \delta) \cap E.$$

A local maximum or minimum is called **a local extremum**. If the inequality is strict (for $x \neq x_0$) we will say that the extremum is strict.

Here is the crucial property (which, of course, you have met before).

8.2. **Proposition (Fermat's theorem on extrema).** *Let $f\colon (a,b) \to \mathbb{R}$. Suppose that $x_0 \in (a,b)$ is a local extremum and f is differentiable at x_0. Then $f'(x_0) = 0$.*

Proof. If x_0 is a local maximum, then there exists $\delta > 0$ such that whenever $0 < x - x_0 < \delta$ and $x \in (a,b)$,

$$\frac{f(x) - f(x_0)}{x - x_0} \leqslant 0$$

so that

$$f'_+(x_0) = \lim_{x \to x_0+} \frac{f(x) - f(x_0)}{x - x_0} \leqslant 0.$$

On the other hand, whenever $-\delta < x - x_0 < 0$ and $x \in (a,b)$,

$$\frac{f(x) - f(x_0)}{x - x_0} \geqslant 0$$

so that

$$f'_-(x_0) = \lim_{x \to x_0-} \frac{f(x) - f(x_0)}{x - x_0} \geqslant 0.$$

Since f is differentiable at x_0, the left- and right-derivatives are equal so $f'(x_0) = f'_-(x_0) = f'_+(x_0)$. Hence $f'(x_0) = 0$.

Similarly if x_0 is a local minimum. □

[In Fermat's theorem it is essential that the interval (a,b) is open. Why?]

8.3. **Aside.**

Fermat's Theorem uses information about a function close to a local extremum x_0 in (a,b) to get information about $f'(x_0)$.

Let's turn this around. Recalling the result in 1.12 about local preservation of strict positivity we can expect get information about the local behaviour of a function f near x_0 from knowledge that $f'(x_0) > 0$ (or $f'(x_0) < 0$). See Problem sheet 5, Q. 2.

8.4. **ROLLE'S THEOREM (1691).** *Let $f\colon [a,b] \to \mathbb{R}$. Assume that*

(i) *f is continuous on $[a,b]$;*

(ii) *f is differentiable on (a,b);*

(iii) *$f(a) = f(b)$.*

Then there exists $\xi \in (a,b)$ such that $f'(\xi) = 0$.

Proof. If f is constant in $[a,b]$, then $f'(x) = 0$ for every $x \in (a,b)$, so that any point—say $\xi = \frac{1}{2}(a+b)$—will do.

As f is continuous on $[a,b]$ it is bounded and attains its maximum and minimum on $[a,b]$ (by the Boundedness Theorem 3.6. As $f(a) = f(b)$, either f is constant and we are done, or else the maximum or the minimum lies in the open interval (a,b). Suppose that $\xi \in (a,b)$ gives either the maximum or minimum. Then it is a local extremum, and by Fermat's result $f'(\xi) = 0$. □

We can express Rolle's Theorem informally by saying

'Between any two roots of f there is a root of f'.'

8.5. **Remarks on the conditions in Rolle's Theorem.**

Observe that we have needed exactly the conditions (i)–(iii) to carry through the proof.

Remember that f is differentiable implies that f is continuous. Thus the hypotheses (i) and (ii) would be satisfied if f was differentiable on $[a,b]$ (with one-sided derivatives. However, often it is important that Rolle holds under the given weaker conditions.

(When using the theorem remember to check ALL conditions including the continuity and differentiability conditions. For example $f\colon [-1,1] \to \mathbb{R}$ given by $f(x) = |x|$ satisfies all conditions of Rolle except that f is not differentiable at $x = 0$. But there is no ξ such that $f'(\xi) = 0$.

Often Rolle's Theorem is applied successively to $f, f', \ldots$ on $[a,b]$ and suitable subintervals of it. If, for example, f has derivatives of all orders on $\mathbb{R}$ then it is acceptable to make a blanket statement at the outset that Rolle's Theorem conditions (i) and (ii) will be satisfied on any closed bounded interval by f and by as many of its derivatives as may be required.

8.6. **Example on Rolle's Theorem.**

Assume that the real-valued function f is twice differentiable on an open interval containing $[0,1]$ and that f''' exists in $(0,1)$. Assume in addition that $f(0) = f'(0) = f(1) = f'(1) = 0$. To prove: that there exists a point $\xi \in (0,1)$ at which $f'''(\xi) = 0$.

The conditions are satisfied to apply Rolle's Theorem to f on $[0,1]$ and so there exists $\alpha \in (0,1)$ such that $f'(\alpha) = 0$. Now the conditions are satisfied to apply Rolle's Theorem to f' on each of $[0,\alpha]$ and $[\alpha,1]$ to obtain β_1 and β_2 with $0 < \beta_1 < \alpha < \beta_2 < 1$ and $f''(\beta_1) = f''(\beta_2) = 0$. Finally, since $\beta_1, \beta_2 \in (0,1)$ on which f''' is given to exist, we know f'' is continuous on $[\beta_1, \beta_2]$ and differentiable on (β_1, β_2), so we can apply Rolle's Theorem to f'' on $[\beta_1, \beta_2]$ to obtain the required point ξ.

The next theorem is one of the most important and useful in the course. We shall reveal many consequences and generalisations of it in this section and subsequent ones, and on problem sheets.

8.7. **MEAN VALUE THEOREM (MVT).** *Let $f\colon [a,b] \to \mathbb{R}$. Assume that*

(i) *f is continuous on $[a,b]$;*
(ii) *f is differentiable on (a,b).*

Then there exists $\xi \in (a,b)$ such that

$$f(b) - f(a) = f'(\xi)(b-a).$$

Proof. Apply Rolle's theorem to the function

$$F(x) = f(x) - k(x-a),$$

where k is a constant to be chosen. Certainly $F\colon [a,b] \to \mathbb{R}$ is continuous, and F is differentiable on (a,b). We choose k so that $F(a) = F(b)$, that is,

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

Thus Rolle's Theorem applies, so $F'(\xi) = 0$ for some $\xi \in (a,b)$. But $F'(x) = f'(x) - k$, so

$$f'(\xi) = k = \frac{f(b) - f(a)}{b-a}. \qquad \square$$

Here is one of the most useful corollaries of the MVT.

8.8. **Constancy Theorem ('A function with zero derivative is constant').** *Let $f\colon (a,b) \to \mathbb{R}$ be differentiable, and satisfy $f'(t) = 0$ for all $t \in (a,b)$. Then f is constant on (a,b).*

[Note that the interval (a,b) need not be bounded.]

Proof. For any $x, y \in (a,b)$ apply the MVT to f on $[x,y]$ where we assume $x < y$, without loss of generality. (Note that f is differentiable on (a,b) implies that f is continuous on (a,b) and hence f is continuous on $[x,y]$.) Then $f(x) - f(y) = f'(\xi)(x-y)$ for some $\xi \in (x,y)$. But $f'(\xi) = 0$, so that $f(x) = f(y)$. Therefore f is constant in (a,b). $\square$

Here's a very simple illustration of the use of the Constancy Theorem.

8.9. **Example.** Suppose that ϕ is a function whose derivative is x^2. Then we have, for all x, that $\phi(x) = \frac{1}{3}x^3 + A$ for some constant A.

Proof. Let $f(x) := \phi(x) - \frac{1}{3}x^3$; we aim to show f is constant. Then f is differentiable and $f'(x) = x^2 - \frac{1}{3} \cdot 3x^2 = 0$. By the Constancy Theorem $f(x) = A$ for some constant A. $\square$

We shall explore further consequences of the Constancy Theorem in the next section. In particular, we shall illustrate what the Constancy Theorem can tell us about the solutions of certain differential equations.

8.10. **Derivatives and monotonicity.**

Let $f\colon (a,b) \to \mathbb{R}$ be differentiable.

(i) If $f'(x) \geqslant 0$ for all $x \in (a,b)$ then f is increasing on (a,b).
To prove this, simply fix $x < y$ in (a,b) and apply the MVT to $[x,y] \subset (a,b)$ to get $f(y) - f(x) = f'(\xi)(y-x)$ for some $\xi \in (x,y)$. Then $f(y) - f(x)$ a product of non-negative numbers. Hence $f(y) \geqslant f(x)$ and we are done.

(ii) If $f'(x) \leqslant 0$ for all $x \in (a,b)$ then f is decreasing on (a,b).

(iii) If $f'(x) > 0$ for all $x \in (a,b)$ then f is strictly increasing on (a,b).

(iv) If $f'(x) < 0$ for all $x \in (a,b)$ then f is strictly decreasing on (a,b).

(Compare these results with what was said in 8.3. The conditions on f here are much stronger.)

8.11. **Examples: MVT, etc.**

(1) **Lipschitz continuous functions revisited**

Recall Example 4.3. Let E be a closed bounded interval $[a, b]$ in $\mathbb{R}$ and let $f\colon E \to \mathbb{R}$ be such that f' exists on E and is continuous. Then the MVT gives that, for $x, y \in E$, there exists $\xi \in E$ such that

$$|f(x) - f(y)| = |f'(\xi)|\,|x - y|.$$

Since we are assuming that f' is continuous on E, the Boundedness Theorem tells us that there exists a finite constant K such that $|f'| \leqslant K$. This means that

$$|f(x) - f(y)| \leqslant K|x - y| \quad \text{for all } x, y \in E.$$

We deduce that any continuously differentiable function on a closed bounded interval is Lipschitz continuous, so acquiring many examples of such functions.

By way of contrast, suppose we had a real-valued function on a bounded interval E which satisfied the condition

$$\forall x, y \in E \ \ |f(x) - f(y)| \leqslant |x - y|^{\alpha}$$

where α is a constant > 1. Then f is a Lipschitz continuous function:

$$|f(x) - f(y)| \leqslant K_{\alpha}|x - y| \quad \text{where } K_{\alpha} := \ell(E)^{\alpha-1},$$

where $\ell(E)$ is the length of the bounded interval E. So f is Lipschitz continuous. But let's fix x and consider what happens when $y \to x$. We see that $f'(x)$ exists and equals 0 (one-sided derivative exists at an endpoint of E if this belongs to E, but we don't need this fact). So the MVT then tells us that the only possibility for f is that it is a constant function.

(2) **Bernoulli's inequality revisited**.

In Analysis I you met the useful inequality

$$(1 + x)^r \geqslant 1 + rx \quad (x > -1,\ r \in \mathbb{N}).$$

This was proved by induction.

We can apply the MVT to $f(y) = (1 + y)^r$ on $[0, x]$ (or $[x, 0]$). There exists η strictly between 0 and x such that $f(x) - f(0) = x \cdot r(1 + \eta)^{r-1}$. Hence, since $\eta > -1$,

$$(1 + x)^r \geqslant rx.$$

Now we can remove the restriction that r be a natural number. For a a positive constant and y a positive real variable let $g(y) = y^a$ where $y^a := \mathrm{e}^{a \log y}$. Then from 7.13, making use of the Chain Rule, we see that y^a is differentiable on $(0, \infty)$ and

$$g'(y) = \frac{a}{y}\mathrm{e}^{a \log y} = a\mathrm{e}^{(a-1)\log y} = ay^{a-1},$$

by properties of exponentials. [*We could have introduced arbitrary powers and their derivatives at the end of Section 7, but have waited to mention them until we had a worthwhile application.*]

The MVT application now goes through just as before when we replace r by a constant $a \geqslant 1$. We obtain

$$(1 + x)^a \geqslant 1 + ax \quad \text{for } x > -1,\ a \in \mathbb{R},\ a \geqslant 1.$$

'bf Note. When getting the required inequality from the MVT result

$$\frac{(1 + x)^a - 1}{x} = a(1 + \xi)^{a-1} \quad \text{where } x \neq 0 \text{ and } \xi \text{ is between 0 and } x.$$

treat the cases $-1 < x < 0$ and $x > 0$ separately (don't fudge inequalities!).

8.12. **Inequalities involving trigonometric functions.**

There are various inequalities involving the trigonometric functions which should be part of every mathematician's toolkit. Assuming, as we have been doing hitherto, the derivatives of sin and cos, we can assemble these inequalities.

We don't dwell on the familiar facts that $|\sin x| \leqslant 1$ and $|\cos x| \leqslant 1$. These follow from $\cos^2 x + \sin^2 x = 1$, which in turn is proved by using the Differentiation Theorem for power series, and the Constancy theorem (see Analysis I notes and Section 9).

(1) $|\sin x| \leqslant |x|$ for all $x \in \mathbb{R}$.

Proof. Since both x and $\sin x$ are odd functions, we may wlog consider $x \geqslant 0$. Consider $f(x) = x - \sin x$ on $(0, \infty)$. Then f is differentiable and $f'(x) = 1 - \cos x \geqslant 0$. It follows from 8.3 that $f(x) - f(0) \geqslant 0$ for $x \geqslant 0$. Therefore $x - \sin x \geqslant 0$ for $x \geqslant 0$. □

Of course this inequality is most useful for values of x close to 0.

(2) **Jordan's inequality**:

$$\frac{2}{\pi} \leqslant \frac{\sin x}{x} \leqslant 1 \quad \text{for } x \in (0, \pi/2].$$

Proof. We have already proved the second inequality. To prove that

$$\sin x/x \geqslant \sin(\pi/2)/(\pi/2) \quad \text{for } 0 < x \leqslant \pi/2$$

it suffices to prove that $g(x) := \sin x/x$ is decreasing on $(0, \pi/2]$. Here we assume that $\pi/2$ is the smallest positive zero of $\cos x$ (recall Problem sheet 2, Q. 4).

Differentiation gives

$$g'(x) = \frac{x\cos x - \sin x}{x^2} = \frac{x - \tan x}{x^2 \cos x}.$$

So let's consider the derivative of $h'(x) := \tan x - x$ on $(0, \pi/2)$. By the Chain Rule, $h'(x) = \sec^2 x - 1$ so $h'(x) > 0$ on $(0, \pi/2)$ and hence h is strictly increasing. □

We conclude thus section with a generalisation of the Mean Value Theorem.

8.13. **Cauchy's Mean Value Theorem.** *Let $f, g \colon [a, b] \to \mathbb{R}$. Assume that*

(i) *f, g are continuous on $[a, b]$;*
(ii) *f, g are differentiable on (a, b).*

Then there exists $\xi \in (a, b)$ such that

$$f'(\xi)(g(b) - g(a)) = g'(\xi)(f(b) - f(a)).$$

If in addition $g'(x) \neq 0$ for all $x \in (a, b)$, then $g(b) \neq g(a)$ and the conclusion can be written

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

Note. We cannot obtain this result by applying the MVT to f and g individually since that way we'd obtain two 'ξ's, one for f and one for g, and these would in general not be equal.

Proof. Define F on $[a, b]$ by

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

that is,

$$F(x) = (f(a)g(b) - f(b)g(a)) + f(x)\,(g(a) - g(b)) + g(x)\,(f(b) - f(a))\,.$$

The function F is a linear combination of f and g. Hence it is continuous on $[a,b]$ and differentiable on (a,b). Clearly $F(a) = F(b) = 0$. By Rolle's Theorem there exists $\xi \in (a,b)$ such that $F'(\xi) = 0$. But

$$0 = F'(\xi) = 0 + f'(\xi)\,(g(a) - g(b)) + g'(\xi)\,(f(b) - f(a)).$$

For the last part note that Rolle's Theorem implies that $g(b) \neq g(a)$ if g' is never zero on (a,b) and we can rearrange to express the conclusion in the required fractional form. □

---

9. Elementary functions: debts paid

Our objective in this section is to prove—at last!—the Differentiation Theorem for real power series. In conjunction with the Constancy Theorem (which you took for granted in Analysis I) this leads to the derivation, now with complete justification, of many well-known properties of the elementary functions. With functions defined by power series discussed in Section 5 and differentiability defined in terms of function limits in Section 7, this is an opportune point at which to revisit the Differentiation Theorem, whose proof is a longstanding unpaid debt. We are in luck: there is a relatively simple proof now available for real power series—making use of the Mean Value Theorem!

You were introduced to the Differentiation Theorem for power series in Analysis I. It asserts that, for a (real) power series $\sum c_k x^k$ with radius of convergence $R > 0$,

$$f(x) = \sum_{k=0}^{\infty} c_k x^k$$

is differentiable on $(-R, R)$ with derivative given by term-by-term differentiation:

$$f'(x) = g(x), \text{ where } g(x) := \sum_{k=1}^{\infty} k c_k x^{k-1}.$$

We shall obtain this as a corollary of a more general result on differentiating a real series term by term, Theorem 9.2. The general result is proved with the aid of the MVT. In accordance with the syllabus for Analysis II we shall also formulate and prove a Differentiation Theorem for complex power series; we relegate this very technical proof to an Appendix to the webnotes. Neither of the proofs we give is examinable material for Prelims. [A third proof, restricted to real power series, will appear in Analysis III. This relies on justifying term-by-term integration of the candidate derivative $g(x)$; it is easy—and examinable.]

9.1. **Preparation: recap on series and the Cauchy and uniform Cauchy conditions.**

Suppose (v_k) is a sequence of functions with $v_k \colon E \to \mathbb{R}$ ($E \subseteq \mathbb{R}$, $k \geqslant 1$). Let (g_n) be the associated sequence of partial sums:

$$g_n := v_1 + \cdots + v_n \quad (n \geqslant 1).$$

Then

- (p) $\sum v_k$ converges pointwise on E, meaning that (f_n) converges pointwise on E, $\iff$ for each $x \in E$, the real sequence $(f_n(x))$ satisfies the Cauchy condition.
- (u) $\sum v_k$ converges uniformly on E, meaning that (f_n) converges uniformly on E, $\iff$ the sequence (f_n) satisfies the uniform Cauchy condition:

  $$\forall \varepsilon > 0\ \exists N \in \mathbb{N}\ \forall m, n > N\ \forall x \in E\ \ |f_n(x) - f_m(x)| < \varepsilon.$$

  (See 5.11, 5.13.)

9.2. **Theorem (term-by-term differentiation of a series).**[2] *Let (u_k) be a sequence of real-valued functions on an open interval $E = (a, b)$. Assume that*

(i) *each u_k is differentiable on E;*

(ii) $\sum_{k=1}^{\infty} u_k(x_0)$ *converges for some $x_0 \in E$;*

(iii) $\sum u'_k$ *converges uniformly on E.*

Then

(a) $f(x) := \sum_{k=1}^{\infty} u_k(x)$ *converges for each $x \in E$;*

(b) *f is differentiable in E with*

$$f'(x) = \sum_{k=1}^{\infty} u'_k(x).$$

Proof. Let $f_n = \sum_{k=1}^{n} u_k$ $(n \geqslant 1)$. Then each f_n is differentiable with $f'_n = u'_1 + \cdots + u'_n$. For each $x, y \in E$ and each m, n, the MVT applies to $(f_n - f_m)$ on $[x, y]$ (or $[y, x]$) to supply $\xi \in E$ such that

$$(f_n(x) - f_m(x)) - (f_n(y) - f_m(y)) = (x - y)(f'_n(\xi) - f'_m(\xi)); \qquad (*)$$

here ξ depends on x, y, m, n. This relation (*) is the key to the proof of both parts of the theorem.

Consider (a). Put $y = x_0$. Given $\varepsilon > 0$ and $x \in E$ with $x \neq x_0$, choose $N_0 \in \mathbb{N}$ such that, for $m, n \geqslant N_0$ and all $z \in E$,

$$|f_n(x_0) - f_m(x_0)| < \tfrac{1}{2}\varepsilon,$$

$$|f'_n(z) - f'_m(z)| < \tfrac{1}{2}\varepsilon |x - x_0|^{-1}.$$

The first of these exploits (ii) and the second comes from (iii) and 9.1. Now (*) and the Triangle Inequality give

$$|f_n(x) - f_m(x)| < \varepsilon \qquad (m, n \geqslant N_0,\ x \in E)$$

(here $x = x_0$ can be included, by (*)). This tells us that (f_n) satisfies the uniform Cauchy condition and so is both pointwise and uniformly convergent on E, and its pointwise limit is $f = \sum_{k=1}^{\infty} u_k$ (by definition of the RHS and uniqueness of limits). Therefore (a) holds.

Now consider (b). Fix $x \in E$ and let $y \neq x$. Let

$$h_n(y) := \frac{f_n(x) - f_n(y)}{x - y} \quad \text{and} \quad h(y) := \frac{f(x) - f(y)}{x - y}.$$

We know that $h_n(y) \to f'_n(x)$ as $y \to x$, for each fixed n. We want to show that $h(y)$ tends to a limit as $y \to x$ (so that $f'(x)$ exists) and that $f'(x) = \lim_{n\to\infty} f'_n(x)$. This is exactly like what we needed to do to prove that uniform convergence preserves continuity, except that now we're considering differentiability; [Expressing this another way, we need to prove that the limits as $y \to x$ and as $n \to \infty$ commute.] We want an $\varepsilon/3$ argument.

Fix $\varepsilon > 0$. Use (iii) and 9.1 with $v_k = u'_k$ to choose $N \in \mathbb{N}$ such that

$$|f'_n(z) - f'_m(z)| < \tfrac{1}{3}\varepsilon \qquad (m, n \geqslant N,\ z \in E).$$

Then (*) gives

$$\left| \frac{f_n(x) - f_n(y)}{x - y} - \frac{f_m(x) - f_m(y)}{x - y} \right| < \tfrac{1}{3}\varepsilon, \qquad (**)$$

[2]From an unpublished note by J.F.C. Kingman.

for $x, y \in E$ $(x \neq y)$ and $m, n \geqslant N$. Fix x, y and $n \geqslant N$ and let $m \to \infty$ to get

$$\left| \frac{f_n(x) - f_n(y)}{x - y} - \frac{f(x) - f(y)}{x - y} \right| \leqslant \tfrac{1}{3}\varepsilon.$$

So $|h_N(y) - h(y)| \leqslant \frac{1}{3}\varepsilon$ for all $y \neq x$. From (**) we know that (h_n) is uniformly Cauchy and so has a limit ℓ, say, as $y \to x$. Now

$$\begin{aligned} |h(y) - \ell| &\leqslant |h(y) - h_N(y)| + |h_N(y) - f'_N(x)| + |f'_N(y) - \ell| \\ &< \tfrac{1}{3}\varepsilon + \tfrac{1}{3}\varepsilon + \tfrac{1}{3}\varepsilon \end{aligned}$$

if $y \neq x$ is such that $|x - y|$ is sufficiently small for the middle term to be $< \frac{1}{3}\varepsilon$, which can be done by differentiability of f_N at x. □

9.3. **DIFFERENTIATION THEOREM FOR REAL POWER SERIES.** *Let the real power series $\sum c_k x^k$ have radius of convergence $R > 0$ and define f in $(-R, R)$ by $f(x) := \sum_{k=0}^{\infty} c_k x^k$. Then f is differentiable in $(-R, R)$ and f' is given by term-by-term differentiation:*

$$f'(x) = \sum_{k=1}^{\infty} k c_k x^{k-1} \quad (|x| < R).$$

Furthermore f has derivatives of all orders in $(-R, R)$ and $f^{(n)}(0) = n! c_n$ for $n \geqslant 1$.

DEBT PAID!

Proof. We wish to show that the theorem is a corollary of Theorem 9.2. Assume $R \in \mathbb{R}$. We fix η with $0 < \eta < R$ and let $E = (-R + \eta, R - \eta)$. Take $u_k(x) = c_k x^k$ (we may work with $(u_k)_{k \geqslant 0}$ with only trivial changes to the theorem we wish to apply). Certainly $\sum u_k(x)$ converges for $x = 0$ and each u_k is differentiable on E. We claim that $\sum u'_k$ converges uniformly on E. For $x \in E$ with $x \neq 0$, Then, for $x \neq 0$, $x \in E$, corrected/simplified

$$|k c_k x^{k-1}| \leqslant M_k := \frac{k}{(R - \eta)} |c_k (R - \eta)^k|$$

To show $\sum M_k$ converges, choose ρ such that $R - \eta < \rho < R$. Then

$$|k(R - \eta)^k| = k((R - \eta)/\rho)^k \cdot |c_k \rho^k|.$$

The series $\sum k((R - \eta)/R)^k$ converges by the Ratio test. Since the terms of a convergent series tend to 0, there exists a constant M such that $k((R - \eta)/R)^k \leqslant M$ for all k. Also $\sum |c_k \rho^k|$ converges by 5.17 since $0 < \rho < R$. Hence $\sum M_k$ converges by comparison. So $\sum u'_k$ converges uniformly on E by the M-test. Therefore the conditions of Theorem 9.2 are satisfied. We conclude that term-by-term differentiation of $\sum c_k u^k$ is justified for each point in E. Since this is true for each $\eta > 0$, this completes the proof for the case that R is finite.

The proof when $R = \infty$ is very similar: prove uniform convergence of the differentiated series, and hence the validity of term-by-term differentiation, in any bounded interval $(-C, C)$.

The final statement is proved by induction on n. □

Note. Uniform convergence, on suitable subintervals of $(-R, R)$, of the differentiated power series can be seen as consequence of a general result asserting that differentiating term by term does not reduce the radius of convergence (in fact it does not change it). We have opted in our proof above to take the most direct proof possible to our immediate goal. (See the Appendix to the webnotes for a full account.)

9.4. **Exploiting the Differentiation Theorem for power series.**

At the end of Analysis I you saw how, assuming the Differentiation Theorem for real power series, you could obtain the familiar formulae for the derivatives of the real exponential, trigonometric and hyperbolic functions which are *defined* by power series with infinite radius of convergence.

The series defining exp, cos, sin, cosh, sinh all have infinite radius of convergence. The Differentiation Theorem gives, for $x \in R$,

$$\frac{\mathrm{d}}{\mathrm{d}x}\exp(x) = \frac{\mathrm{d}}{\mathrm{d}x}\sum_{k=0}^{\infty}\frac{x^k}{k!} \overset{\star}{=} \frac{\mathrm{d}}{\mathrm{d}x}\frac{x^k}{k!} = \sum_{k=1}^{\infty}\frac{x^{k-1}}{(k-1)!} = \exp(x);$$

$$\frac{\mathrm{d}}{\mathrm{d}x}\cos x = \frac{\mathrm{d}}{\mathrm{d}x}\sum_{k=0}^{\infty}(-1)^k\frac{x^{2k}}{(2k)!} \overset{\star}{=} \frac{\mathrm{d}}{\mathrm{d}x}(-1)^k\frac{x^{2k}}{(2k)!} = \sum_{k=1}^{\infty}(-1)^k\frac{x^{2k-1}}{(2k-1)!} = -\sin x;$$

$$\frac{\mathrm{d}}{\mathrm{d}x}\sin x = \frac{\mathrm{d}}{\mathrm{d}x}\sum_{k=0}^{\infty}(-1)^k\frac{x^{2k+1}}{(2k+1)!} \overset{\star}{=} \frac{\mathrm{d}}{\mathrm{d}x}(-1)^k\frac{x^{2k+1}}{(2k+1)!} = \sum_{k=1}^{\infty}(-1)^k\frac{x^{2k}}{(2k)!} = \cos x;$$

and likewise for $\cosh x$ and $\sinh x$. The occurrences of $\overset{\star}{=}$ show the points at which we have differentiated term by term, as the Differentiation Theorem tells us we may.

In general, if you are given a function *defined* by a power series and asked to investigate its properties, *your first task is to calculate the radius of convergence R.*

9.5. **Elementary functions: stocktake on properties.**

Finally, we have proved all the general theorems on which the key properties of the elementary functions rest. This is an appropriate point at which to sum up what we have discovered.

Table 1 lists our principal tools. Note that other key theorems are used in proving certain of these results: Boundedness Theorem, Rolle's Theorem and MVT.

	Theorem	ref. number
(1)	Intermediate Value Theorem (IVT)	3.9
(2)	Inverse Function Theorem (IFT)	3.14
(3)	Continuity Theorem for power series	5.17
(4)	Differentiation: basic properties	Sec. 7
(5)	IFT (differentiability)	7.12
(6)	Constancy Theorem	8.8
(7)	Monotonicity test, via derivative	8.9
(8)	Differentiation Theorem for (real) power series (validates term-by--term differentiation & gives a formula for derivative)	9.3

Table 1. Elementary functions: the principal tools

Table 2 pulls together how the tools are used to derive properties of the elementary functions.

	proof uses
exp (see 7.19)	
continuity	(3)
differentiability & what the derivative is	(8)
$\forall x \forall y\ \mathrm{e}^x \mathrm{e}^y = \mathrm{e}^{x+y}$ (†)	(8), (6)
$\mathrm{e}^x > 0$, $e^0 = 1$ & $\mathrm{e}^{-x} = 1/\mathrm{e}^x$	defn ($x \geqslant 0$) & (†)
limiting behaviour as $x \to \pm\infty$	defn & (†)
strictly increasing	(8), (7) (or via (†))
exp maps $\mathbb{R}$ onto $(0, \infty)$	(3), (1) and limiting behaviour
sin, cos; sinh, cosh	
continuity	(3), (1)
differentiability & what the derivatives are	(8)
$\cos^2 x + \sin^2 x = 1$ (‡)	(8), (6)
$\cosh^2 x - \sinh^2 x = 1$	(8), (6)
addn formulae	(8), (6)
sin, cos	
$\lvert \cos x \rvert \leqslant 1$, $\lvert \sin x \rvert \leqslant 1$	from (‡)
$\exists$ smallest zero > 0 ($\pi, \pi/2$ resp.), leading to	(3)
periodicity properties	(1) & addition formulae (Problem sheets: 2, Q4; 6, Q3)
log	
exists on $(0, \infty)$ as inverse to exp & is continuous	properties of exp and (2)
differentiability & what the derivative is	(5)
$\forall u \forall v\ \log uv = \log u + \log v \quad (u, v > 0)$	formula (†) for exp

Table 2: The tools exploited

We now open up a gateway to the solution of differential equations, and uniqueness theorems. Last term you learned methods for guessing solutions of first and second order linear odes. You were then told that these solutions could be used to get the general solution. The Constancy Theorem gives us a tool to prove the uniqueness of solutions of DEs and to justify that you did indeed have general solutions last term, as would have been claimed.

9.6. **Example.**

Here is a very fundamental example of how we use the Constancy Theorem to find the general solution of a differential equation. We shall show that the general solution for $f'(x) = f(x)$ for all $x \in \mathbb{R}$, is $f(x) = A\exp(x)$ where A is a constant. (That is, every solution is of this form.)

The 'trick' for solving differential equations is to manipulate them so that they look like $\frac{\mathrm{d}}{\mathrm{d}x}F = 0$ for some F, and then 'integrate'. This can often be achieved by multiplying by 'integrating factors'. The same 'trick' lets us apply the MVT (or the Constancy Theorem) to *prove* that the solution must be of this form.

Last term you learnt that to solve the differential equation $\dfrac{\mathrm{d}f}{\mathrm{d}x} - f = 0$ you multiply it by $\mathrm{e}^{-\int 1 \mathrm{d}x}$, rewrite it as $\dfrac{\mathrm{d}}{\mathrm{d}x}\left(\mathrm{e}^{-x} f(x)\right) = 0$ and deduce that $\mathrm{e}^{-x} f(x) = A$.

Now let's write this as a piece of pure mathematics! Consider $F(x) := f(x)\exp(-x)$. Then $F'(x) = f'(x)\exp(-x) - f(x)\exp(-x) = 0$. Hence, by the Constancy Theorem $F(x)$ is constant; that is $f(x)\exp(-x) = A$ say, and so $f(x) = A\exp(x)$ and all solutions are of this form.

---

10. Useful limits and L'Hôpital's Rule

Analysis II provides an in-depth exploration of function limits, continuity and differentiability and along the way sets the elementary functions—the backbone of everyday mathematics—on a firm footing, with their important properties revealed. From here one can go on to work with these functions with confidence in a variety of areas of pure and applied mathematics.

We have already indicated how the MVT leads to useful inequalities involving the elementary functions. Alongside these as part of a mathematician's toolkit should be a catalogue of standard limits, and methods for evaluating limits of compound functions which cannot be obtained by simple (AOL) arguments.

We start off with some useful limits which can be obtained from first principles, from results about function limits, or by combining basic inequalities with sandwiching arguments.

Often we are interested in limits as $x \to 0$ or as $x \to \infty$ (or $x \to -\infty$). We record an ancillary lemma. Part (a) helps us toggle between these; its proof is a definition chase and parts (b) and (c) appeared on Problem sheet 5 (Q.5).

10.1. **Lemma (limits at infinity).**

(a) *Let f be a real-valued function defined on some interval $(0, \delta)$, where $\delta > 0$. Then $\lim_{x\to\infty} f(x)$ exists if and only if $\lim_{x\to 0+} f(1/x)$ exists and when this is true the values of the two limits are the same.*

(b) *Suppose $f(x) \to \infty$ and that $g(x) \to 1$, both as $x \to \infty$. Then $f(x)g(x) \to \infty$ as $x \to \infty$.*

(c) (i) *Let $f, g\colon \mathbb{R} \to \mathbb{R}$ be functions. Suppose that $f(x) \to k$ as $x \to \infty$, and that g is continuous at k. Then $g(f(x)) \to g(k)$ as $x \to \infty$.*

Let $f, g\colon \mathbb{R} \to \mathbb{R}$ be functions. Suppose that $f(x) \to \infty$ as $x \to \infty$, and that $g(x) \to \ell$ as $x \to \infty$. Then $g(f(x)) \to \ell$ as $x \to \infty$.

Many of the examples which follow are function limit versions of results you encountered for sequence limits in Analysis I, some with proofs, others (such as Euler's limit), you may not have seen proved there. The $x \to \infty$ results subsume the corresponding $n \to \infty$ results and the proofs are generally slicker.

10.2. **Some old friends: polynomials and exponentials.**

(1) Let m be a positive integer. Then, as $x \to \infty$, the power $x^m \to \infty$ and as $x \to -\infty$,

$$x^m \to \begin{cases} \infty & \text{if } m \text{ is even,} \\ -\infty & \text{if } m \text{ is odd.} \end{cases}$$

Moreover $x^{-m} \to 0$ as $x \to \pm\infty$.

Proof. Go back to first principles or use an argument based on inequalities. $\square$

(2) Let $p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0$ be a non-constant monic real polynomial. Then $p(x) \to \infty$ as $x \to \infty$.

Proof. Write

$$p(x) = x^n q(x) \quad \text{where } q(x) = 1 + a_{n-1}x^{-1} + \cdots + a_0 x^{-n}.$$

By (1) above and (AOL), $q(x) \to 1$ as $x \to \infty$. Now invoke Lemma 10.1(b). □

(3) Let $\alpha \in \mathbb{R}$ and $\beta > 0$ be constants. Then $\lim_{x\to\infty} x^\alpha \mathrm{e}^{-\beta x} = 0$.

Proof. We may restrict attention to $x > 0$. Then, by properties of exp,

$$\begin{aligned} 0 \leqslant x^\alpha \mathrm{e}^{-\beta x} &= \frac{x^\alpha}{1 + \beta x + \beta^2 x^2/2! + \cdots + \beta^n x^n/n! + \cdots} \\ &\leqslant n!\beta^{-n} x^{\alpha - n}, \end{aligned}$$

for any fixed n. Fix a value of $n > 2\alpha$. Then

$$0 \leqslant x^\alpha \mathrm{e}^{-\beta x} \leqslant n!\beta^{-n} x^{-n}.$$

The result now follows by sandwiching. □

[**Note**. When $x > 0$, working with the power series for e^x, which has all terms positive, is preferable to working with the power series for e^{-x} which has terms of alternating sign. Inequalities, as we need for sandwiching, interact badly with expressions with mixed signs.]

10.3. **Limits involving logarithms.**

(1) Consider $f(x) = \dfrac{\log x}{x^p}$, where $p > 0$ is a constant. We claim $\lim_{x\to\infty} f(x)$ exists and equals 0.

Proof. Write $y = \log x$. Then $x = \mathrm{e}^y$ and $x^p = \mathrm{e}^{p\log x} = \mathrm{e}^{py}$. Moreover $x \to \infty$ implies $y \to \infty$ (a consequence of the results in 3.20). So we look at $y\mathrm{e}^{-py}$ as $y \to \infty$. By 10.2(3),

$$\lim_{y\to\infty} y\mathrm{e}^{-py} = 0.$$

Hence

$$\lim_{x\to\infty} \frac{\log x}{x^p} = 0. \quad \square$$

(2) Consider $g(x) = x^q \log x$, where $q > 0$ is a constant. We claim $\lim_{x\to 0} g(x)$ exists and equals 0.

Proof. The result is equivalent to $\lim_{x\to\infty} (1/x)^q \log(1/x) = 0$, by Lemma 10.1(a). The required result follows from the preceding one. □

10.4. **Limits involving powers.**

Here we give easy proofs of some famous limits. Their analogues for sequences were useful in Analysis I.

(1) We claim that $x^{1/x} \to 1$ as $x \to \infty$.

Proof. Let $x > 0$. By definition, $x^{1/x} = \mathrm{e}^{(\log x)/x}$. By Example **??**(2), $x^{-1}\log x \to 0$ as $x \to \infty$. Since exp is continuous, Lemma 10.1(c) gives

$$x^{1/x} = \mathrm{e}^{(\log x)/x} \to \mathrm{e}^0 = 1.$$

□

(2) $\lim_{x\to 0} \dfrac{\log(1+x)}{x} = \lim_{x\to 0} \dfrac{\log(1+x) - \log 1}{x-0} = 1$, because the limit is just the derivative of $\log(1+x)$ evaluated at $x = 0$.

Hence $(1+x)^{1/x} = \mathrm{e}^{\log(1+x)/x} \to \mathrm{e}^1 = \mathrm{e}$, by continuity of exp at 0. So

$$\lim_{x\to 0}(1+x)^{1/x} = 1.$$

(3) **Euler's limit.** Now let $x > 0$ and consider $u = 1/x$ and $v = -1/x$. Consider $u \to 0+$ and $v \to 0-$. Then (2) together with Lemma 10.1(a) gives

$$\lim_{x\to\infty}\left(1+\frac{1}{x}\right)^x = \lim_{u\to 0+}(1+u)^{1/u} = \mathrm{e}.$$

In a similar way,

$$\lim_{x\to\infty}\left(1-\frac{1}{x}\right)^{-x} = \lim_{v\to 0-}(1+v)^{v} = \mathrm{e},$$

from which we obtain

$$\lim_{x\to\infty}\left(1-\frac{1}{x}\right)^{x} = \frac{1}{\mathrm{e}}.$$

10.5. **Limits involving trigonometric functions.**

(1) Given that the Differentiation Theorem for power series tells us that $\sin x$ is differentiable with derivative $\cos x$, we can immediately see that

$$\frac{\sin x}{x} = \frac{\sin x - \sin 0}{x - 0} \to \cos 0 = 1 \quad \text{as } x \to 0,$$

by definition of the derivative at 0 and continuity of cos.

(2) (AOL) gives that

$$\lim_{x\to 0}\frac{\sin^{11}(37x)}{\sin^{11}(19x)} = \lim_{x\to 0}\left(\frac{37}{19}\right)^{11}\cdot\lim_{x\to 0}\frac{\sin^{11}(37x)}{(37x)^{11}}\cdot\lim_{x\to 0}\frac{\sin^{11}(19x)}{(19x)^{11}} = \left(\frac{37}{19}\right)^{11}.$$

10.6. **Introduction to L'Hôpital's Rule(s).**

It should be apparent that in some of the above examples what prevented us using (AOL) directly to find a limit of a quotient $f(x)/g(x)$ as $x \to a$, say, is that the individual limits, $\lim_{x\to a} f(x) = 0$ and $\lim_{x\to a} g(x)$ are both 0, so the quotient rule for function limits did not apply. Here f and g may be defined at a but need not be. We sometimes also encounter one-sided limits when the domain of f or g forces this: as when $a = 0$ and a log function is involved. We have seen too that issues of a similar kind arise for limits as $x \to \infty$ and/or where $f(x)$ and $g(x)$ both tend to ∞.

What we are contending with here are limits which involve what are known generically as **indeterminate forms**. They come in a variety of flavours, and our examples so far illustrate how to deal, albeit in a somewhat ad hoc way, with many of the limits that crop up frequently in practice. Can we more systematic and can we invoke theoretical tools to extend our catalogue of examples? The answer to both questions is a qualified 'yes'.

In the remainder of this section we discuss a technique known as **L'Hôpital's Rule** (or maybe it should be referred to as L'Hôpital's Rules). The key to the method lies in the material in Section 8. We stressed that the MVT allows us to relate the behaviour of a differentiable function to the behaviour of its derivative, and this idea extends to pairs of functions via the Cauchy MVT. Our objectives here are:

- to prove a 'vanilla' version of L'Hôpital's Rule (10.10) and to demonstrate how to apply it rigorously;
- to draw attention, briefly, to some of the most commonly used variants;

- to add to our catalogue of useful limits, in some cases providing alternative derivations of ones we know already.

It is not our intention to provide a comprehensive handbook of the various scenarios to which the L'Hôpital technique can be adapted. In any case, certain indeterminate limits arising in applications may require special treatment and call for ingenuity.

10.7. **Differentiability conditions.**

We shall begin with some technical comments. Suppose we wish to evaluate a function limit

$$\lim_{x\to a}\frac{f(x)}{g(x)},$$

where f and g are defined in some interval $(a-\delta, a+\delta)$ and $g(x)\neq 0$ in this interval. Either of the two following scenarios may apply:

(I) $f'(x)$ and $g'(x)$ may exist for $x\neq a$ (x near a will do), but one or both of $f'(a)$ and $g'(a)$ may fail to exist;

(II) f and g may both be differentiable *at* a, but one (or both) may fail to be differentiable in any set $(a-\eta, a+\eta)\setminus\{a\}$, where $\eta > 0$.

Of course for well-behaved functions neither of these possibilities will occur.

To illustrate that (I) can arise, recall that $f(x)=\sin(1/x)$ is differentiable on $\mathbb{R}\setminus\{0\}$ but not at 0. To illustrate (II) It is easy to construct a function f on $(-1,1)$ for which $f'(0)$ exists and equals 0 but which fails to be differentiable at each point $1/n$ ($n=2,3,\dots$). One way to do this is to construct a continuous piecewise linear function f with the properties:

$$|f(x)|\leqslant x^2,\quad f(1/n)=(-1)^n\frac{1}{n^2}.$$

With this warning in mind, we now give a baby version of L'Hôpital's Rule.

10.8. **L'Hôpital's Rule (A), for functions differentiable at** a**).** *Let* f, g *be functions defined in some interval* $(a-\delta, a+\delta)$*. Assume that*

(i) $f'(a)$ *and* $g'(a)$ *exist;*
(ii) $f(a)=g(a)=0$*;*
(iii) $g'(a)\neq 0$.

Then

$$\lim_{x\to a}\frac{f(x)}{g(x)} \text{ exists and equals } \frac{f'(a)}{g'(a)}.$$

One-sided versions, with the appropriate one-sided derivatives, also work.

Proof. First note that $g'(a)\neq 0$ implies that $g(x)=g(x)-g(a)\neq 0$ for $0<|x-a|<\eta$, for some η with $0<\eta\leqslant\delta$. For such x,

$$\frac{f(x)}{g(x)}=\frac{f(x)-f(a)}{g(x)-g(a)}=\frac{f(x)-f(a)}{x-a}\cdot\frac{x-a}{g(x)-g(a)}\to\frac{f'(a)}{g'(a)},$$

by (AOL). □

10.9. **Applications of L'Hôpital's Rule (A).**

L'Hôpital's Rule (A) is not powerful. It does no more than combine (AOL) with the definition of derivative (and when $g(x) = x$ it reduces to the definition of the derivative of f at a). We can see that we dealt with the rather artificial example in 10.5(2) by the L'Hôpital's Rule (A) strategy. Here are some other examples that work the same way. Note that all the required conditions are satisfied!

(1) $\lim_{x\to 0} \dfrac{\log(1+x)}{\sin x} = 1$;

(2) $\lim_{x\to 0} \dfrac{x^{3/2}}{\tan x} = \dfrac{3/2 \cdot 0^{1/2}}{\sec^2(0)} = 0$.

Now a grown-up version of L'Hôpital's Rule.

10.10. **L'Hôpital's Rule (B).**

For a right-hand limit: *Suppose f and g are real-valued functions defined in some closed interval $[a, a+\delta]$ $(\delta > 0)$. Assume that*

(i) *f and g are continuous in $[a, a+\delta]$;*

(ii) *f and g are differentiable in $(a, a+\delta)$;*

(iii) *$f(a) = g(a) = 0$;*

(iv) $\lim_{x\to a+} \dfrac{f'(x)}{g'(x)}$ *exists.*

Then

$$\lim_{x\to a+} \frac{f(x)}{g(x)} \text{ exists and equals } \lim_{x\to a+} \frac{f'(x)}{g'(x)}.$$

For the case of a **left-hand limit** *replace $[a, a+\delta]$ by $[a-\delta, a]$ and $(a, a+\delta)$ by $(a-\delta, a)$.*

For the case of a **two-sided limit** *replace $[a, a+\delta]$ by $[a-\delta, a+\delta]$ and replace $(a, a+\delta)$ by $(a-\delta, a+\delta) \setminus \{a\}$.*

Proof. We have opted to prove the version for a right-hand limit since we can do this without the distraction of having to bother about the sign of $x - a$ when working with the Cauchy MVT. The left-hand limit version is proved likewise and the two-sided version then follows from 1.15.

So assume conditions (i)–(iv) hold as set out for the right-hand limit version. Tacitly, (iv) tells us that $g'(x) \neq 0$ in some interval $(a, a+\delta')$ so we may assume that δ is chosen small enough that $g'(x) \neq 0$ in the given interval $(a, a+\delta)$. We also know by Rolle's Theorem for g that $g(x) - g(a) \neq 0$ for $x \in (a, a+\delta)$. Apply the Cauchy MVT to obtain $\xi_x \in (a, x)$ such that

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

Since $a < \xi_x < x$, necessarily $x \to a$ forces $\xi_x \to a$, the result now follows from (iv). $\square$

10.11. **L'Hôpital's Rule (B) in practice.**

Rarely in practice do we encounter instances where L'Hôpital's Rule (B) applies as stated, other than in cases where alternative methods for obtaining the limit (if it exists) of $f(x)/g(x)$ are as good or better. Where L'Hôpital's Rule (B) comes into its own is when it is used *recursively*. (The word 'recursive' means 'running backwards'.)

Let us illustrate with a simple example. Suppose we want to investigate the limiting behaviour, as $x \to 0$, of $f(x)/g(x)$, where $f(x) = 1 - \cos x$ and $g(x) = x^2$. Here $f'(x) = \sin x$ and

$g'(x) = 2x$. Certainly $f(0) = g(0) = 0$ so the rule applies and tells us we now need to know $\lim_{x\to 0} f'(x)/g'(x)$. We have $f'(0) = g'(0) = 0$ so it is tempting just to apply L'Hôpital's Rule again, this time with f and g replaced by f' and g'. But we need to be careful in order that we don't assume a limit exists before we have proved that it does. (In fact in this case we don't need to invoke L'Hôpital's Rule (B) a second time because we already know that $\lim_{x\to 0} \sin x/x$ exists and equals 1 (Example 10.5(1)).

In what follows let us restrict to a special case, which often arises when we are working with specific functions. (For comments on the general case see 10.14.) We shall assume that f and g are infinitely differentiable in some open interval containing a, so conditions (i) and (ii) are automatically satisfied. This allows us to focus attention as we proceed on conditions (iii) (indeteminacy condition) and (iv) (candidate limit).

10.12. **Example (L'Hôpital's Rule (B)).**

We look at a contrived example, for illustrative purposes. We claim that

$$\lim_{x\to 0} \frac{\sin x - x}{\sinh^3 x} \text{ exists and equals } -\frac{1}{6}.$$

Proof. Take $f(x) = \sin x - x$ and $g(x) = \sinh^3 x$. The blanket conditions of infinite differentiability hold. Now we have

$f(x) = \sin x - x$	$g(x) = \sinh^3 x$
$f(0) = 0$	$g(0) = 0$
$f'(x) = \cos x - 1$	$g'(x) = 3\sinh^2 x \cosh x$
$f'(0) = 0,$	$g'(0) = 0$
$f''(x) = -\sin x$	$g''(x) = 6\sinh x \cosh^2 x + 3\sinh^3 x$
$f''(0) = 0$	$g''(0) = 0$
$f'''(x) = -\cos x$	$g'''(x) = 6\cosh^3 x + \text{ [terms in } \sinh x]$.

By (AOL),

$$\lim_{x\to 0} \frac{f'''(x)}{g'''(x)} \text{ exists and equals } -\frac{1}{6}.$$

We now see that we can apply L'Hôpital's Rule (B), in turn, to f'', to f' and to f (running backwards!). To hammer home the point, the argument goes like this:

$$\ell := \lim \frac{f'''(x)}{g'''(x)} \text{ exists, whence L'HR implies}$$

$$\lim \frac{f''(x)}{g''(x)} \text{ exists and equals } \ell;$$

$$\text{SO } \lim \frac{f''(x)}{g''(x)} \text{ does exist, whence L'HR implies}$$

$$\lim \frac{f'(x)}{g'(x)} \text{ exists and equals } \ell;$$

$$\text{SO } \ell := \lim \frac{f'(x)}{g'(x)} \text{ does exist, whence L'HR implies}$$

$$\lim \frac{f(x)}{g(x)} \text{ exists and equals } \ell.$$

□

This is rather laborious, and we can without loss of rigour present the argument more compactly. Start as before by defining f and g and observing that they are infinitely differentiable. Then proceed as follows:

$$\lim_{x\to 0}\frac{\sin x - x}{\sinh^3 x} = \lim_{x\to 0}\frac{\cos x - 1}{3\sinh^2 x\cosh x} \quad \text{[Note } f(0)=g(0)=0]$$
if limit on RHS exists

$$= \lim_{x\to 0}\frac{-\sin x}{3\sinh^2 x\cosh x} \quad \text{[Note } f'(0)=g'(0)=0]$$
if limit on RHS exists

$$= \lim_{x\to 0}\frac{-\cos x}{6\cosh^3 x + [\text{terms in } \sinh x]} \quad \text{[Note } f''(0)=g''(0)=0]$$

and this exists and equals $-\dfrac{1}{6}$ by (AOL)

Recursive procedure for L'Hôpital's Rule (B).

We now set out algorithmically (but in an informal style), the process for applying L'Hôpital's Rule (B). We don't presume conditions (i) and (ii) are satisfied automatically; checking these at each step needs to be part of the procedure.

0. `Let` $h := f$ `and` $k := g$, `where` $f(a) = g(a) = 0$.
1. `IF` $k'(a) \neq 0$ `THEN put` $\lim \dfrac{h((x)}{k(x)} = \dfrac{f'(a)}{k'(a)}$ `(call it` ℓ`) and STOP`
   `ELSE`
2. `IF` $\lim \dfrac{h'(x)}{k'(x)} = \ell$ `exists THEN put` $\lim \dfrac{h(x)}{k(x)} = \ell$ `and STOP`
   `ELSE`
3. `IF` $\dfrac{h'(x)}{k'(x)} \to \pm\infty$ `THEN put` $h := g$ `and` $k := f$ `and GOTO Step 1.`
   `ELSE`
4. `IF` $h'(a) = k'(a) = 0$ `THEN put` $h := h'$ `and` $k := k'$ `and GOTO Step 1.`
   `ELSE`
5. `EXIT`

A few explanatory comments are in order.

STOP indicates that the algorithm terminates. The required limit of $f(x)/g(x)$ is ℓ (or its reciprocal in the case Step 3., which swaps the functions, was used). Note that Step 3. provides a work around for one 'bad' case.

EXIT means that L'Hôpital's Rule (B) has failed to deliver an answer, perhaps because the limit sought fails to exist. To take a silly example, suppose we tried to apply L'Hôpital's Rule (B) to $\lim_{x\to 0}(x^3\sin(1/x))/x^3$. The limit clearly fails to exist.

10.13. **Variants of L'Hôpital's Rule.** Problem sheet 7, Q.3 supplies a version of L'Hôpital's Rule (B) for limits of the form

$$\lim_{x\to\infty}\frac{f(x)}{g(x)} \quad \text{where } f(x)\to 0 \text{ and } g(x)\to 0.$$

This variant is obtained from standard L'Hôpital's Rule (B) for a right-hand limit by considering $y = 1/x$, taking care over the technical conditions. Observe that the proof of L'Hôpital's Rule (B) uses the Cauchy MVT, for which we need continuity on a closed interval containing the point a. When we use $x \mapsto 1/x$ to convert a limit at ∞ to a limit at 0, we have to engineer continuity at 0, since our original functions are not defined 'at infinity'. Limits at infinity in concrete cases are usually best treated by conversion to limits at 0 and then treated directly, rather than by treating L'Hôpital's Rule at infinity as a quotable result.

Another scenario that we have not considered is that in which we have ∞/∞ indeterminacy, rather than 0/0 indeterminacy. This is trickier: See optional scaffolded question 7 on Problem sheet 7. Some of our examples in 10.2 & **??** can be viewed as ∞/∞ indeterminacy and be handled by this version of L'Hôpital's Rule. but we do not recommend this.

10.14. **Summing up and some tactical tips.**

Don't get into the way of thinking that L'Hôpital's Rule is the universal answer to evaluating limits in which indeterminacy arises. It would be devious to invoke it, in either (A) or (B) form, to prove that $\sin x/x \to 1$ as $x \to 0$. Also, it is often well worth while to pre-process an indeterminate limit to get it into a more amenable form before bringing in L'Hôpital's Rule: for example, by simplifying the given quotient by making use of trigonometric formulae, for example, or by using an 'it would be enough to consider' argument taking advantage of known limits and (AOL). An examples of the latter would be

$$\lim_{x\to 0} \frac{x^2}{\tan^2 x} = \lim_{x\to 0} \cos^2 x \cdot \left(\frac{x}{\sin x}\right)^2 :$$

Just use continuity of cos, the standard $\lim_{x\to 0} \sin x/x = 1$ and (AOL).

For infinitely differentiable functions f and g, you can expect that the existence of the limit as $x \to a$ you finally need to confirm exists will follow from (AOL), from L'Hôpital's Rule (A) or just because you've reached a limit which defines the derivative at a of a function whose derivative is known. Don't do an extra application of L'Hôpital's Rule you don't need.

Suppose you are asked to evaluate an indeterminate limit $\lim_{x\to 0} f(x)/g(x)$ where f and g are well-behaved, familiar functions for which you could write down the first few terms of Maclaurin expansions. This will give you an idea of the orders of magnitude of f and g near 0, and hence what the limit of the quotient is likely to be 0, a finite constant, or infinity. As it stands, this is not a proof because we do not know how terms embraced by $\cdots$'s behave: you would need to estimate these in order to get a watertight proof. (Further comments on this to follow in connection with Taylor's Theorem.)

Finally, a comment on a situation in which you meed to be careful not to apply L'Hôpital's Rule (B) more times than is allowed. This is likely to crop up when you are given an indeterminate limit involving a general function with restricted differentiability properties. See Problem sheet 7, Q.5 for an example.

.

.

11. Taylor's Theorem

Our objective in this section to investigate how a real-valued function may be approximated by a polynomial. We begin by showing how some of our earlier results may be viewed as providing such approximations.

We emphasise that our methods rely on Rolle's Theorem and the Mean Value Theorem. This means that the t results of this section do not extend to complex-valued functions.

Let's review the context, based on what we know already. Let $f \colon E \to \mathbb{R}$ be a function defined on some interval $E = [-\delta, \delta]$, where $\delta > 0$. (Later we'll consider other domains such as $[a - \delta, a + \delta]$ but we take $a = 0$ for now, for notational simplicity.)

We begin by looking at two extreme cases.

11.1. **Real power series.**

Suppose we have a function $f\colon (-R,R) \to \mathbb{R}$ defined by

$$f(x) = \sum_{k=0}^{\infty} c_k x^k \quad x \in (-R,R)$$

where $R > 0$ is the radius of convergence of $\sum c_k x^k$. Then the Differentiation Theorem for power series tells us that f has derivatives of all orders and, for all n,

$$c_n = \frac{f^{(n)}(0)}{n!}$$

where by convention we write $f^{(0)} = f$. Therefore

$$f(x) = f(0) + xf'(0) + \frac{x^2}{2!}f''(0) + \cdots + \frac{x^k}{k!}f^{(k)}(0) + \cdots \quad (x \in (-R,R)).$$

This allows is to write, for any n,

$$\begin{aligned} f(x) = P_n(x) + E_n(x), \quad \text{where}& \\ P_n(x) &:= f(0) + xf'(0) + \cdots + \frac{x^n}{n!}f^{(n)}(0), \\ E_n(x) &:= f(x) - P_n(x). \end{aligned}$$

Here $P_n(x)$ is a polynomial of degree n and can be viewed as an approximation to $f(x)$, and $E_n(x)$ as an error term. Moreover we know that $E_n(x) \to 0$ for each $x \in (-R,R)$ and even, by properties of power series, that $E_n \xrightarrow{\text{u}} 0$ on any closed subinterval $[-R+\eta, R-\eta]$ of $(-R,R)$.

But what we don't know in general is how fast the tail of the power series, that is,

$$E_n(x) = \sum_{k=n+1}^{\infty} \frac{f^{(k)}}{k!} x^k,$$

tends to 0, at individual points or on closed subintervals of $(-R,R)$. We do not so far know a method which allows us to express $E_n(x)$ in closed form and hence to estimate its magnitude.

11.2. **Crude approximations.**

(1) Assume only that $f'(0)$ exists. Then Lemma 7.4 tells us that

$$f(x) = f(0) + x[f'(0) + x\varepsilon(x) \quad \text{where } \varepsilon(x) \to 0 \text{ as } x \to 0.$$

Then we may view $P_1(x) := f(0) + xf'(0)$ as providing a linear approximation to $f(x)$, where P_1 is the polynomial we introduced above, in the special case that $n = 1$.

(2) Assume now that f is continuous on $[-\delta, \delta]$ and differentiable on $(-\delta, \delta) \setminus \{0\}$. Then the Mean Value Theorem gives

$$f(x) = f(0) + xf'(\theta x) \quad \text{where } 0 < \theta < 1, \text{ with } \theta \text{ depending on } x.$$

This time our approximation to $f(x)$ matches up with that we proposed for power series only for $n = 0$. But we do have an explicit formula for the error $f(x) - f(0)$ when f is approximated by the constant $f(0)$: it is expressible as $xf'(\theta x)$, where $\theta(x)$ lies between x and 0.

In general, suppose that we have a real-valued function f defined in some interval $(a-\delta, a+\delta)$ and such that the first n derivatives of f exist in this interval. (We are principally interested in $f'(a)$, …, $f^{(n)}(a)$ but recall that in order to define $f^{(k)}(a)$ we need $f^{(k-1)}(a)$ defined in an open interval containing a. Hence the conditions we have imposed on f are appropriate.) Then we define the **Taylor polynomial of degree** n (at a given a) to be

$$P_n(x) := f(a) + f'(0)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n.$$

We have so far considered functions which lie at opposite ends of the spectrum as regards to their differentiability properties. It is tempting to ask whether, by assuming the existence of finitely many of f'', f''', ..., we can improve on what we can get on the basis of information about f' alone.

Above we cited the MVT to arrive at a very crude approximation to a differentiable function, Remember that the proof of the MVT involved a single application of Rolle's Theorem. Examples on repeated applications of Rolle's Theorem (see for example 8.6 and Problem sheet 5) lead us to be optimistic that this idea may be fruitful in the present context.

Without further ado we go straight to Taylor's Theorem. Our initial formulation matches the notation we used when introducing the Mean Value Theorem, but shall quickly thereafter encourage adaptability in choice of notation.

11.3. **Taylor's Theorem with Lagrange form of remainder.** *Let $f\colon [a,b]$. Let $n \geqslant 0$ be such that*

(i) *$f, f', f'', \ldots, f^{(n)}$ exist (one-sided derivatives at endpoints) and are continuous on $[a,b]$;*

(ii) *$f^{(n+1)}$ exists on (a,b).*

Then there exists $\xi \in (a,b)$ such that

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

Proof. Define $F\colon [a,b] \to \mathbb{R}$ by

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

where K is a constant to be chosen. We have $F(a) = 0$. Choose K so that $F(b) = 0$. By (i) F is continuous on $[a,b]$. By (ii) F is differentiable on (a,b) and

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

We can apply Rolle's Theorem to F on $[a,b]$. Hence there exists $\xi_1 \in (a,b)$ such that $F'(\xi_1) = 0$.

Assume $n \geqslant 1$, otherwise we're done. Observe that we have $F'(a) = 0 = F'(\xi)$ and that F' is continuous on $[a,\xi]$ and differentiable on (a,ξ). So we can apply Rolle's Theorem to F' on $[a,\xi_1]$ to obtain $\xi_2 \in (a,\xi_1)$ for which $F'(\xi_2) = 0$.

Likewise for any k with $2 \leqslant k \leqslant n$ we have $F^{(k)}$ continuous on any interval $[a,c]$ with $a < c \leqslant b$, and differentiable on (a,c), and also $F^{(k)}(a) = 0$. Hence we can apply Rolle's Theorem successively ($n+1$ times in total) to obtain $\xi_1, \xi_2, \ldots, \xi_{n+1}$ with

$$a < \xi_{n+1} < \cdots < \xi_2 < \xi_1 < b$$

and $F^{(k)}(\xi_k) = 0$ for $k = 1, \ldots, n+1$. [A diagram is recommended.]

Set $\xi = \xi_{n+1}$. Notice that

$$f^{(n+1)}(x) = f^{(n+1)}(x) - K(n+1)!\,.$$

Recalling that we chose K so that $F(b) = 0$, the required result drops out. □

11.4. **Remarks: Taylor's Theorem.**

As we presented it in 11.3, Taylor's Theorem gave a formula for $f(b)$ in terms of the Taylor polynomial of degree n at a, where tacitly we assumed $a < b$. But we can equally well have $b < a$—only some fine detail in the proof changes. We are dealing here with what may be seen as one-sided versions of the theorem. In view of its importance as also formulate a two-sided version (11.5). Note that the conditions are a little stronger in terms of the technical assumptions.

11.5. **Taylor's Theorem (symmetric form).** *Let* $f\colon (a-\delta, a+\delta) \to \mathbb{R}$, *where* $\delta > 0$. *Let* $n \geqslant 0$. *Assume that* $f', f'', \ldots, f^{(n+1)}$ *exist on* $(a-\delta, a+\delta)$. *Let* $x \in (a-\delta, a+\delta)$. *Then there is a number* ξ *between* a *and* x *such that*

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k + \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1}.$$

Proof. If $x > a$ this comes from Taylor's Theorem as in 11.3. If $x < a$ we just use that theorem on the function $f(-x)$. If $x = a$ then take $\xi = a$ (we define $0^0 := 1$). □

11.6. **A notational variant.**

Consider the situation in which the assumptions of Theorem 11.3 hold for f on $[a, b]$. Then they hold too for f on $[a, a+h]$, where $0 < h \leqslant (b-a)$. Then the theorem gives

$$f(a+h) = f(a) + hf'(a) + \frac{h^2}{2}f''(a) + \cdots + \frac{h^n}{n!}f^{(n)}(a) + \frac{h^{n+1}}{(n+1)!}f^{(n+1)}(a+\theta h) \quad \text{where } 0 < \theta < 1.$$

A left-sided version can also be formulated.

It is important to realise that the number θ here *depends on* h (and on a, which we regard as fixed). We have in general no information on how θ varies with h, though it may sometimes be possible to get information in the limit as $h \to 0$ (see Problem sheet 7, Q. 6).

The further $a+h$ is from a the less likely the Lagrange polynomial is to give a good approximation to $f(a+h)$. Moreover it may be hard in specific cases to find a tight estimate of the size of the error term $\frac{h^{n+1}}{(n+1)!}f^{(n+1)}(a+\theta h)$ especially since the value of θ is not known, so that we need a global upper bound covering all possible values of $a + \theta h$.

11.7. **Variations on a theme.**

There are different versions of Taylor's Theorem valid under different technical assumptions and with the remainder term expressible in different ways. Two illustrations can be found on Problem sheet 8.

Not a different theorem, just a warning. We have presented Taylor's Theorem with Lagrange remainder for a function which is $(n+1)$-times differentiable on (a, b), so that the associated polynomial approximation has degree n; here n may be any integer $\geqslant 0$. The result is also often seen formulated with $n+1$ replaced by n; now $n \geqslant 1$, and consequential notational changes made. The theorem itself is unchanged.

A brief aside. We have focused exclusively on use of the Taylor polynomials as polynomial approximations to a given function f on some closed interval, with suitable assumptions on existence of derivatives $f', f'', \ldots$. There are other possibilities that may be appropriate in certain contexts. For example, one might want to construct a polynomial approximation which agrees with f at some specified finite set of n points (a curve-fitting problem). This requires **Lagrange interpolation** to obtain N approximating polynomial of degree $(n-1)$. Then one can use repeated applications of Rolle's Theorem on a suitably defined function—a strategy akin to that we used to prove Taylor's Theorem. This and other similar problems are taken up in courses on Numerical Analysis.

11.8. **Example: the function** $\log(1+x)$**.**

By way of an example of the use of Taylor's Theorem we investigate the series expansion of $\log(1+x)$ for $x \in (-1, 1]$.

This expansion is

$$\log(1+x) = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{x^n}{n} \quad \text{for all } x \in (-1, 1].$$

Note that the series has radius of convergence 1 and that for $x = -1$ it becomes the divergent harmonic series, so we cannot extend the result beyond $(-1, 1]$.

The proof of the expansion was previewed in Analysis I; see Analysis I notes, 12.13 and 13.12. For $|x| < 1$ the expansion can be derived from the Differentiation Theorem. For $x = 1$ it is a spin-off from the use of Euler's constant. But these methods do not supply an estimate of how fast the series converges to its pointwise limit. So let's see what Taylor's Theorem can tell us.

Consider $f(x) = \log(1+x)$. We have already proved that on $(-1, \infty)$ the function f is differentiable with $f'(x) = \frac{1}{1+x}$; so, by induction,

$$f^{(n)}(x) = \frac{(-1)^{n-1}(n-1)!}{(1+x)^n} \quad \text{for all } n \geqslant 1.$$

Hence, by Taylor's Theorem (the symmetric version), for $|x| < 1$,

$$\log(1+x) - \sum_{k=1}^{n-1} \frac{(-1)^{k-1} x^k}{k} = (-1)^{n-1} \frac{1}{n} \left(\frac{x}{1+\xi_n} \right)^n$$

for some ξ_n between 0 and x, with ξ_n depending on x.

To get our required result from this it would be enough to show that

$$\left| \frac{x}{1+\xi_n} \right| \leqslant 1$$

for every n and $x \in (-1, 1]$.

For $x > 0$ this is no problem, $0 < \xi_n < 1$ and so $1 + \xi_n > 1$; hence $\frac{x}{1+\xi_n} < x \leqslant 1$.

For negative x it is not so easy; the nearer x is to -1 the nearer $1+\xi_n$ may get to 0. However, if $x \geqslant -\frac{1}{2}$ we have

$$-\tfrac{1}{2} \leqslant x \leqslant \xi_n \leqslant 0$$

and so

$$\tfrac{1}{2} \leqslant 1 + \xi_n \leqslant 1$$

which implies

$$2x \leqslant \frac{x}{1+\xi_n} \leqslant x.$$

Now $2x \geqslant -1$ and $x \leqslant 1$ so we have

$$\left| \frac{x}{1+\xi_n} \right| \leqslant 1$$

as required.

That is, the functions $\log(1+x)$ and $\displaystyle\sum_{k=1}^{\infty} (-1)^{k-1} \frac{x^k}{k}$ are equal on $[-\frac{1}{2}, 1]$.

What about $(-1, -\frac{1}{2})$? We must use a very different argument.

Consider the functions $f(x) = \log(1+x)$ and $g(x) = \displaystyle\sum_{k=1}^{\infty}(-1)^{k-1}\frac{x^k}{k}$ on $(-1,1)$. Both are differentiable there; we have proved $f'(x) = \dfrac{1}{1+x}$, and by the Differentiation Theorem,

$$g'(x) = \sum_{k=1}^{\infty}(-1)^{k-1}k\frac{x^{k-1}}{k} = \sum_{k=1}^{\infty}(-1)^{k-1}x^{k-1} = \frac{1}{1+x}.$$

Hence $f'(x) - g'(x) = 0$, so by the Constancy Theorem,

$$f(x) - g(x) = f(0) - g(0) = 0.$$

That is, on the whole of $(-1,1]$ we have the required series expansion.

The last part has actually proved the result for $x \in (-1,1)$. It is only at $x = 1$ that we have to prove that the error tends to zero.

11.9. **Cautionary tales.**

(1) The technical conditions in Taylor's Theorem can seem irksome. But they are necessary. We have seen that there are functions which fail to have derivatives of all orders at some given point a. In such cases there will be restrictions on how large a value of n can be used in a Taylor expansion about a. Occurrence of a rapidly oscillating term in a function, such as $\sin(1/x)$ near 0, often constrains the number of times a function can be differentiated. But in a worst case scenario there are functions for which Taylor's Theorem is useless. These include the continuous functions which are **nowhere differentiable**, and they are much less rare than you might think. See eg Bartle & Sherbert for a discussion of how to construct one such function (don't expect an accurate picture of the graph: such functions cannot be 'drawn').

(2) Let us now reveal a quite different type of bad behaviour. Our intuition, built on experience of polynomials, trigonometric and exponential functions, is misleading. The following example shows us that there are functions $f(x)$, with derivatives of all orders at every point of $\mathbb{R}$, such that $\sum \frac{f^k(0)}{k!}x^k$ is convergent for every x—but for which $E_n(x) \not\to 0$.

Consider $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} \exp(-\frac{1}{x^2}) & \text{whenever } x \neq 0, \\ 0 & \text{for } x = 0. \end{cases}$$

Some experimentation shows that we expect

$$f^{(k)}(x) = \begin{cases} Q_k(\frac{1}{x})\exp(-\frac{1}{x^2}) & \text{whenever } x \neq 0, \\ 0 & \text{for } x = 0, \end{cases}$$

for some polynomial Q_k of degree $3k$. We can prove this by induction: At points $x \neq 0$ this is routine use of linearity, the product rule and the chain rule. But at $x = 0$ we need to take more care, and use the definition:

$$\frac{f^{(k)}(x) - f^{(k)}(0)}{x - 0} = \frac{1}{x}Q_k\left(\frac{1}{x}\right)\exp\left(-\frac{1}{x^2}\right) = \sum_{s=1}^{3k+1} a_s\frac{\exp(-\frac{1}{x^2})}{x^s}$$

which we must prove tends to zero as $x \to 0$. Change the variable to $t = \frac{1}{x}$ then we have a finite sum of terms like $t^s \exp -t^2$ which we know tend to zero as $|t|$ tends to infinity.

So for this function f the series $\sum \frac{f^{(k)}(0}{k!}x^k = 0$ so converges to 0 at every x. But the error term $E_n(x)$ is the same for all n (it equals $f(x)$) and so does not tend to 0 at any point except 0.

Note that we can add this function to $\exp x$ and $\sin x$ and so on, and get functions with the same set of derivatives at 0 as these functions, so that they will have the same Taylor polynomials—but are different functions.

On the positive side we record that the picture changes radically when one considers complex valued functions of a complex variable. Then condition of differentiability is much stronger, and any complex-valued function differentiable on an open disc in $\mathbb{C}$ is infinitely differentiable and is representable as a power series. This is a story for Part A, not Prelims.

---

12. The binomial expansion

We have in Analysis II defined arbitrary powers and found their derivatives. This, combined with use of the Differentiation Theorem and Constancy Theorem, validates, for any real number p, the expansion

$$(1+x)^p = \sum_{k=0}^{\infty} \binom{p}{k} x^k \qquad \text{for all } |x| < 1.$$

See Analysis I notes, Example 13.11 for the calculations.

What follows is a self-contained account of the binomial theorem for arbitrary index This is taken, unedited and with acknowledgement, from notes for Analysis II from an earlier year. The earlier material repeats the example presented in Analysis I, but in more detail; it is included for completeness. The later material was not covered in Analysis II lectures previously, and it will not be this year either. It is included for the benefit of those who are interested in, or might later find useful, the analysis of the behaviours of the series expansion at the endpoints. (The special case of $(1+x)^{-1/2}$ appears in the final, optional, question on Problem sheet 8.)

Supplement to notes, from 2016 lecture notes.

We shall use many of the theorems we have proved about uniform convergence and continuity, power series, monotonicity as well as Taylor's Theorem. As well as proving an important result we are showing off the techniques we now have available to us.

12.1. **Motivation and preliminary algebra.** By simple induction we can prove that for any natural number n (including 0) we have for all real or complex x that

$$\begin{aligned}(1+x)^n &= \sum_{k=0}^{n} \binom{n}{k} x^k \\ &= \sum_{k=0}^{\infty} \binom{n}{k} x^k;\end{aligned}$$

where the coefficient $\binom{n}{k}$ of x^k can be proved to be

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

or

$$\binom{n}{k} = \frac{n\cdot(n-1)\cdot\cdots\cdot(n-k+1)}{k\cdot(k-1)\cdot\cdots\cdot 1} \qquad (= 0 \text{ if } k > n).$$

We want to extend this result. We have also seen in our work on sequences and series that

$$(1+x)^{-1} = \sum_{k=0}^{\infty} (-1)^k x^k \qquad \text{for all } \ |x| < 1$$

and here the coefficient of x^k can be written as

$$(-1)^k = \frac{(-1)\cdot(-2)\cdot\dots\cdot(-k)}{k\cdot(k-1)\cdot\dots\cdot 1};$$

and we can prove by induction (for example using differentiation term by term) that for all natural numbers $n \geqslant 1$ we have that

$$(1+x)^{-n} = \sum_{k=0}^{\infty} \frac{(-n)\cdot(-n-1)\cdot\dots\cdot(-n-k+1))}{k\cdot(k-1)\cdot\dots\cdot 1} x^k \qquad \text{for all } \ |x| < 1,$$

so the binomial theorem above holds for all integers n if we define

$$\binom{n}{k} := \frac{n\cdot(n-1)\cdot\dots\cdot(n-k+1)}{k\cdot(k-1)\cdot\dots\cdot 1}.$$

We are going to generalise these—in the case of some real values of x—to all values of n, not just integers. Note that this is altogether deeper: $(1+x)^p$ is defined for non-integral p, and for (real) $x > -1$, to be the function $\exp(p\log(1+x))$.

12.2. **Definition.**

For all $p \in \mathbb{R}$ and all $k \in \mathbb{N}$ we extend the definition of **binomial coefficient** as follows:

$$\binom{p}{0} := 1; \qquad \text{and} \qquad \binom{p}{k} := \frac{p(p-1)\dots(p-k+1)}{k!}.$$

We now make sure that the key properties of binomial coefficients are still true in this more general setting.

12.3. **Lemma.**

$$k\binom{p}{k} = p\binom{p-1}{k-1}, \qquad \textit{for all } k \geqslant 1.$$

Proof. If $k = 1$ then by the definition we must see $1 \cdot \frac{p}{1} = p \cdot 1$ which is clear. Otherwise

$$k\binom{p}{k} = k\frac{p(p-1)\dots(p-k+1)}{k!} = p\frac{(p-1)\dots(p-k+1)}{(k-1)!} = p\binom{p-1}{k-1}.$$

□

12.4. **Lemma.**

$$\binom{p}{k} + \binom{p}{k-1} = \binom{p+1}{k}, \qquad \textit{for all } k \geqslant 1.$$

Proof. When $k = 1$ we must prove $\frac{p}{1} + 1 = \frac{p+1}{1}$ which is clear. Otherwise

$$\begin{aligned}\binom{p}{k} + \binom{p}{k-1} &= \frac{p(p-1)\dots(p-k+1)}{k!} + \frac{p(p-1)\dots(p-k+2)}{(k-1)!} \\ &= \frac{p(p-1)\dots(p-k+2)}{k!}\left[(p-k+1)+k\right] \\ &= \frac{(p+1)p(p-1)\dots(p-k+2)}{k!} \\ &= \binom{p+1}{k}.\end{aligned}$$

□

12.5. **Theorem: the real Binomial Expansion.** *Let p be a real number. Then*

$$(1+x)^p = \sum_{k=0}^{\infty} \binom{p}{k} x^k \qquad \textit{for all } |x| < 1. \qquad (B).$$

Note that the coefficients are all non-zero provided p is not a natural number or zero; as we have a proof of the expansion in that case we may assume that $p \notin \mathbb{N} \cup \{0\}$.

12.6. **Lemma.** *The function f defined on $(-1,1)$ by $f(x) := (1+x)^p$ is differentiable, and satisfies $(1+x)f'(x) = pf(x)$. Also, $f(0) = 1$.*

Proof. The derivative is easily got by the chain rule from the definition of f; it is $f'(x) = p(1+x)^{p-1}$. Multiply by $(1+x)$ and get the required relationship. The value at 0 is clear. □

12.7. **Lemma.** *The radius of convergence of $\sum_{k=0}^{\infty} \binom{p}{k} x^k$ is $R = 1$.*

Proof. Use the ratio test; we have that "$|a_{k+1}x^{k+1}/a_k x^k|$" is

$$\left| \frac{p \cdot (p-1) \cdot \dots \cdot (p-k)}{(k+1) \cdot k \cdot (k-1) \cdot \dots \cdot 1} \cdot \frac{k \cdot (k-1) \cdot \dots \cdot 1}{p \cdot (p-1) \cdot \dots \cdot (p-k+1)} x \right| = \left| \frac{p-k}{k+1} x \right| \to |x|$$

as $k \to \infty$. □

12.8. **Lemma.** *The function g defined on $(-1,1)$ by $g(x) = \sum_{k=0}^{\infty} \binom{p}{k} x^k$ is differentiable, with derivative satisfying $(1+x)g'(x) = pg(x)$. Also, $g(0) = 1$.*

Proof. We have

$$\begin{aligned}
(1+x)g'(x) &= (1+x)\sum_{k=0}^{\infty}\binom{p}{k}kx^{k-1}, \text{ differentiation term by term valid for } |x|<1,\\
&= (1+x)\sum_{k=1}^{\infty}\binom{p}{k}kx^{k-1}\\
&= p(1+x)\sum_{k=1}^{\infty}\binom{p-1}{k-1}x^{k-1} \quad \text{by Lemma 12.3}\\
&= p\left\{\sum_{k=1}^{\infty}\binom{p-1}{k-1}x^{k-1}+\sum_{k=1}^{\infty}\binom{p-1}{k-1}x^{k}\right\}\\
&= p\left\{\sum_{m=0}^{\infty}\binom{p-1}{m}x^{m}+\sum_{m=1}^{\infty}\binom{p-1}{m-1}x^{m}\right\}\\
&= p\left\{1+\sum_{m=1}^{\infty}\binom{p-1}{m}x^{m}+\sum_{m=1}^{\infty}\binom{p-1}{m-1}x^{m}\right\}\\
&= p\left\{1+\sum_{m=1}^{\infty}\left[\binom{p-1}{m}+\binom{p-1}{m-1}\right]x^{m}\right\}\\
&= p\left\{1+\sum_{m=1}^{\infty}\binom{p}{m}x^{m}\right\} \quad \text{by Lemma 12.4}\\
&= p\sum_{m=0}^{\infty}\binom{p}{m}x^{m}\\
&= pg(x).
\end{aligned}$$

□

12.9. **Proof of the Binomial Theorem.** Consider $\phi(x) = \dfrac{g(x)}{f(x)}$, which is well-defined on $(-1,1)$ as $f(x) > 0$. By the Quotient Rule we can calculate $\phi'(x)$, and then use the lemmas:

$$\phi'(x) = \frac{f(x)g'(x) - f'(x)g(x)}{f(x)^2} = \frac{p}{1+x}\,\frac{f(x)g(x) - f(x)g(x)}{f(x)^2} = 0.$$

Hence by the Constancy Theorem, $\phi(x)$ is constant, $\phi(x) = \phi(0) = 1$. This implies that $f(x) = g(x)$ on $(-1,1)$. □

12.10. **The end points: preliminary issue.**

The existence of these functions and their equality at the end points requires more sophisticated argument. *The following subsections should be viewed as illustrations of the way Taylor's Theorem can be exploited, rather than theorems to be learnt.*

The cases $x = 1$ or $x = -1$ need to be considered separately. But there is a difference between these!

For $x = -1$ we have not yet defined $(1+x)^p$.

For $p \in \mathbb{N}$ we have the usual algebraic definition, so $0^p = 0$. Can we define 0^p sensibly for any other values of p?

For $p > 0$: If $x > -1$ we defined $(1+x)^p := \exp p \log(1+x)$. As $\log(1+x) \to -\infty$ as $x \to -1$, we have $\exp p\log(1+x) \to 0$ as $x \to -1$. Thus to make $(1+x)^p$ continuous at $x = -1$ we should define $0^p = 0$. This we now do.

If $p = 0$: How one defines 0^0 depends on the context. (Sometimes $0^0 := 1$ sometimes $0^0 := 0$.) If (B) is to hold for $x = 0$ then we must define $0^0 = 1$. But if we do this , then to preserve the rule of exponents $A^p A^q = A^{p+q}$ we cannot define negative powers; if $p > 0$ then 0^{-p} makes no sense.

So let us extend out definition of $(1+x)^p$ in this way, in the case when $p > 0$.

But we need to take care.

12.11. **Lemma.** *If $p > 0$ then the function $(1+x)^p$ is continuous on $[-1, \infty)$.*

12.12. **Lemma.** *If $p > 1$ then the function $(1+x)^p$ is differentiable on $[-1, \infty)$ with derivative $p(1+x)^{p-1}$.*

Proofs. Exercises. □

12.13. **The end points: $p \leqslant -1$.**

Let $p \leqslant -1$. Then as remarked above, the function $(1+x)^p$ is not defined at $x = -1$. Further the expansion does not converge at $x = 1$:

12.14. **Proposition.** *The series $\sum_{k=0}^{\infty} \dfrac{p \cdot (p-1) \cdot \dots \cdot (p-k+1)}{k \cdot (k-1) \cdot \dots \cdot 1}$ is divergent.*

Proof. Write $q = -p \geqslant 1$; then the modulus of the k-th term

$$\left| \frac{p \cdot (p-1) \cdot \dots \cdot (p-k+1)}{k \cdot (k-1) \cdot \dots \cdot 1} \right| = \left| (-1)^k \frac{q}{1} \cdots \frac{q+s}{s+1} \cdots \frac{q+k-1}{k} \right| \geqslant 1;$$

the terms alternate in sign but as they do not tend to 0 the series diverges. □

12.15. **The end points: $-1 < p < 0$.**

Let $-1 < p < 0$; note that $p + 1 > 0$. Again the function $(1+x)^p$ is not defined at $x = -1$. However, now the expansion converges at $x = 1$:

12.16. **Proposition.** *The series $\sum_{k=0}^{\infty} \dfrac{p \cdot (p-1) \cdot \dots \cdot (p-k+1)}{k \cdot (k-1) \cdot \dots \cdot 1}$ is convergent with sum 2^p.*

Proof. We apply Taylor's Theorem to $(1+x)^p$ on the interval $[0, 1]$ and find, for each $n \geqslant 1$, a point $\xi_n \in (0, 1)$ such that

$$2^p = \sum_{k=0}^{n-1} \frac{p \cdot (p-1) \cdot \dots \cdot (p-k+1)}{k \cdot (k-1) \cdot \dots \cdot 1} + E_n$$

where

$$E_n = \frac{p \cdot (p-1) \cdot \dots \cdot (p-n+1)}{n \cdot (n-1) \cdot \dots \cdot 1} (1+\xi_n)^{p-n}.$$

We have then that

$$|E_n| \leqslant \left| \frac{p \cdot (p-1) \cdot \dots \cdot (p-n+1)}{n \cdot (n-1) \cdot \dots \cdot 1} \right|;$$

and we will have the result if we prove that this tends to 0 as $n \to \infty$. We rewrite $|E_n|$ as

$$\left| \frac{[(p+1)-1] \cdot \dots \cdot [(p+1)-s] \cdot \dots \cdot [(p+1)-n]}{1 \cdot 2 \cdot \dots \cdot n} \right|$$
$$= \left(1 - \frac{p+1}{1}\right) \cdot \left(1 - \frac{p+1}{2}\right) \cdots \left(1 - \frac{p+1}{s}\right) \cdots \left(1 - \frac{p+1}{n}\right).$$

Now $\exp(-x)+x-1$ has positive derivative on $(0,1)$ so by the MVT we have that

$$\left(1-\frac{p+1}{s}\right)\leqslant \exp\left(-\frac{p+1}{s}\right)$$

so that

$$|E_n|\leqslant \exp\left(-(p+1)\sum_{s=1}^{n}\frac{1}{s}\right).$$

As the harmonic series diverges and $(p+1)>0$, we get that $E_n\to 0$ as $n\to\infty$.

□

12.17. **The end points:** $0<p$**.**

Let $0<p$. In this case the expansion is valid at $x=1$ and $x=-1$.

12.18. **Proposition.** *The series* $\sum_{k=0}^{\infty}\dfrac{p\cdot(p-1)\cdot\dots\cdot(p-k+1)}{k\cdot(k-1)\cdot\dots\cdot 1}$ *is convergent with sum* 2^p*; and the series* $\sum_{k=0}^{\infty}\dfrac{p\cdot(p-1)\cdot\dots\cdot(p-k+1)}{k\cdot(k-1)\cdot\dots\cdot 1}(-1)^k$ *is convergent with sum* 0*.*

Proof. The end point $x=+1$ is straightforward; use Taylor's Theorem as before and consider the error estimate

$$E_n=\frac{p\cdot(p-1)\cdot\dots\cdot(p-n+1)}{n\cdot(n-1)\cdot\dots\cdot 1}(1+\xi_n)^{p-n}$$

for some $\xi_n\in(0,1)$. Then

$$|E_n|\leqslant\frac{p}{n}\left|\frac{(p-1)\cdot\dots\cdot(p-n+1)}{1\cdot 2\cdot\dots\cdot(n-1)}\right|\frac{2^p}{1^n}.$$

Now $\left|\dfrac{p-s}{s}\right|\leqslant 1$ whenever $2s\geqslant p$; so we get that

$$|E_n|\leqslant\frac{p}{n}\left|\frac{(p-1)\cdot\dots\cdot(p-\left[\frac{p}{2}\right])}{1\cdot 2\cdot\dots\cdot(\left[\frac{p}{2}\right])}\right|\frac{2^p}{1^n}\to 0\quad\text{as}\quad n\to\infty$$

as required. □

The end point $x=-1$ is more difficult. What we do is prove that the sum converges. Noting that as soon as $k\geqslant p+1$ all the terms have the same sign, we see that this means we have proved that the series is absolutely convergent. Now by the properties of power series $\sum_{k=0}^{\infty}\dfrac{p\cdot(p-1)\cdot\dots\cdot(p-k+1)}{k\cdot(k-1)\cdot\dots\cdot 1}x^k$ is absolutely convergent on $(-1,1)$. In particular we have that the series is absolutely convergent on the closed interval $[-1,0]$. Hence the series is uniformly convergent on that interval; and so the series is continuous on $[-1,0]$. As the series is equal to $(1+x)^p$ on $(-1,0]$ we have by continuity that there is equality at -1 as well.

So we must prove that the series converges. We claim that if we can prove this for any p then we can prove it for $(p+1)$. This is because for all $n\geqslant 2p+2$ we have that $\left|\dfrac{p+1}{p-n+1}\right|\leqslant 1$; this allows us to compare the n-th terms and see that those for $(p+1)$ are smaller in modulus. As both series are ultimately the series of terms of constant sign, the comparison test will yield that convergence for p yields convergence for $(p+1)$. So assume from now that $0<p<1$; it will suffice to deal with this case.

The modulus of the n-th term can then be written

$$|u_n|=\frac{p}{n}\left(1-\frac{p}{1}\right)\dots\left(1-\frac{p}{s}\right)\dots\left(1-\frac{p}{n-1}\right)$$

and so, using again $(1-t) \leqslant \exp(-t)$, we have that

$$\begin{aligned}
|u_n| \leqslant & \frac{p}{n} \exp\left(-p \sum_{s=1}^{n-1} \frac{1}{s}\right) \\
= & \frac{p}{n} \exp\left(-p\left(\sum_{s=1}^{n-1} \frac{1}{s} - \log n\right)\right) \exp\left(-p(\log n)\right) \\
= & \frac{p}{n} \frac{1}{n^p} \exp\left(-p\left(\sum_{s=1}^{n-1} \frac{1}{s} - \log n\right)\right).
\end{aligned}$$

Now we have (Integral Test argument) that

$$\sum_{s=1}^{n-1} \frac{1}{s} - \log n \to \gamma \quad \text{as } n \to \infty \qquad (\gamma \text{ is Euler's constant}).$$

Hence we have a constant C such that

$$|u_n| \leqslant C \frac{1}{n \cdot n^p} \quad \text{for sufficiently large } n,$$

and so, by the Comparison Test, $\sum |u_n|$ converges.[3]

[3] $\sum \frac{1}{n^s}$ is convergent for $s > 1$ by the Integral Test.

Appendix: The Differentiation Theorem for complex power series

This appendix to the Analysis II webnotes provides a proof, applicable to complex power series, of the theorem which validates term-by-term differentiation of such series. This is non-examinable material. This proof is different from, and presented independently of, that given for real power series in Theorem 9.2 of the notes, which is also non-examinable. [A much simpler proof applicable to *real* power series is given in Analysis III.]

Extending concepts in Analysis II from $\mathbb{R}$ to $\mathbb{C}$.

We have mentioned several times that many of the concepts and much of the theory in Analysis II extends without more than notational change to complex-valued functions on subsets of the complex plane. The arithmetic and modulus properties in $\mathbb{C}$ are the same as those for $\mathbb{R}$; only order properties are excluded. In particular $\varepsilon - \delta$ and $\varepsilon - N$ definitions carry over to the complex case so long as they are phrased in terms of modulus.

As regards complex power series, the notions which concern us are:

- Open (closed) discs take the place of open (closed) intervals.
- Radius of convergence of a power series and its basic properties (as in 5.17).
- Uniform convergence of s sequence or series of complex-valued functions on a subset of $\mathbb{C}$—the uniform Cauchy condition, and so also the M-test, remain available.
- Continuity of a function defined on a disc in $\mathbb{C}$; uniform convergence preserves continuity.
- Function limits for complex-valued functions on open discs and the definition of differentiability; (AOL) and Chain Rule remain available.

For a complex power series $\sum_{k=0}^{\infty} c_k z^k$ there exists R, where

$$R := \begin{cases} \sup\{\, |z| \in \mathbb{R} \mid \sum |c_k z^k| \text{ converges} \,\} & \text{if the sup exists,} \\ \infty & \text{if } \sum |c_k z^k| \text{ converges for all } z. \end{cases}$$

The series converges absolutely, and so also converges, in its **disc of convergence**

$$\mathrm{D}(0; R) := \{\, z \in \mathbb{C} \mid |z| < R \,\}.$$

By the M-test, it converges uniformly on each closed disc

$$\overline{\mathrm{D}}(0, R - \delta) := \{\, z \in \mathbb{C} \mid |z| \leqslant R - \delta \,\},$$

where $\delta > 0$ is constant, if $R \in \mathbb{R}$, and on each closed bounded disc if $R = \infty$. Convergence is seldom uniform on the whole disc of convergence.

The function

$$f(z) := \sum_{k=0}^{\infty} c_k z^k$$

is defined on $\mathrm{D}(0; R)$ and is continuous at each point of this disc (just as in Theorem 5.18).

Proving differentiability is harder. It will be convenient to use the following notation in defining the derivative. We say that f is **differentiable at** $z \in \mathrm{D}(0; R)$ if

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

exists. That is,

$$\forall \varepsilon > 0 \ \exists \delta > 0 \ \left(0 < |h| < \delta \implies \left| \frac{f(z+h) = f(z)}{h} - f'(z) \right| < \varepsilon \right).$$

Our objective is to show that $f'(z)$ exists for $|z| < R$ and is given by differentiating term by term. We first need a technical lemma to prove that term-by-term differentiation of a power series doesn't alter the radius of convergence. (Strictly, we shall only need to know that R does not decrease, but the other direction is easy to prove so we include it too.)

Lemma. *The power series* $\sum c_k z^k$ *and* $\sum k c_k z^{k-1}$ *have the same radius of convergence.*

Proof. We first prove that, if $\sum |c_k z^k|$ converges for $|z| < R$, then $\sum |k c_k z^{k-1}|$ also converges for $|z| < R$. Choose ρ such that $|z| < \rho < R$ and assume $z \neq 0$. Then

$$|k c_k z^{k-1}| = \frac{k}{|z|} \left(\frac{|z|}{\rho} \right)^k |c_k \rho|^k.$$

Since $|z|/\rho < 1$, the series $\sum k(|z|/\rho)^k$ converges by the Ratio Test. Since the terms of a convergent series tend to 0, there exists a constant M such that $k(|z|/\rho)^k \leqslant M$ for all k. Hence

$$|k c_k z^{k-1}| \leqslant \frac{M}{|z|} |c_k \rho^k|.$$

The result now follows from the Comparison Test and 5.17 (complex version).

Conversely, suppose that $\sum |k c_k z^{k-1}|$ converges. Then

$$|c_k z^k| \leqslant |z| |k c_k z^{k-1}| \quad (k \geqslant 1),$$

so $\sum |c_k z^k|$ converges by the Comparison Test. □

The Differentiation Theorem for complex power series. *Let* $\sum c_k z^k$ *have radius of convergence* $R > 0$ *and define* f *in* $\mathrm{D}(0; R)$ *by* $f(z) := \sum_{k=0}^{\infty} c_k z^k$. *Then* f *is differentiable in* $\mathrm{D}(0; R)$ *and* f' *is given by term-by-term differentiation:*

$$f'(z) = \sum_{k=1}^{\infty} k c_k z^{k-1} \quad (|z| < R).$$

Furthermore f *has derivatives of all orders in* $\mathrm{D}(0; R)$ *and* $f^{(n)}(0) = n! c_n$ *for* $n \geqslant 1$.

Note: Saying that we can obtain $f'(z)$ by differentiating term by term is the same as saying that $\mathrm{d}/\mathrm{d}z$ commutes with the infinite sum $\sum$; an instance of two limiting processes commuting.

Proof. The lemma allows us to define DEBT PAID!

$$g(z) := \sum_{k=1}^{\infty} k c_k z^{k-1} \quad (|z| < R).$$

We want to show that $f'(z)$ exists and equals $g(z)$ for $z \in \mathrm{D}(0; R)$. For $z, z + h \in \mathrm{D}(0; R)$,

$$\frac{f(z+h) - f(z)}{h} - g(z) = \sum_{k=1}^{\infty} \left(\frac{(z+h)^k - z^k}{h} - k z^{k-1} \right),$$

and we must prove that this tends to 0 as $h \to 0$. We do this by estimating the terms in the series on the right-hand side. We shall need the binomial expansion

$$(z+h)^k = \sum_{r=0}^{k} \binom{k}{r} z^{k-r} h^r \qquad (k = 2, 3, \ldots).$$

This expansion, valid for all z and h in $\mathbb{C}$, is proved by induction, just as the real version is: it relies solely on the arithmetic properties of $\mathbb{C}$.

So (notice the cancellation)

$$\begin{aligned}
\frac{(z+h)^k - z^k}{h} - kz^{k-1} &= \frac{khz^{k-1} + \cdots + \binom{k}{r} h^r z^{k-r} + \cdots + h^k}{h} - kz^{k-1} \\
&= h\left(\binom{k}{2} z^{k-2} + \cdots + \binom{k}{r} h^{r-2} z^{k-r} + \cdots + h^{k-2}\right) \\
&= h \sum_{r=2}^{k} \binom{k}{r} h^{r-2} z^{k-r} \\
&= h \sum_{s=0}^{k-2} \frac{k!}{(k-(s+2))!(s+2)!} h^s z^{k-2-s}
\end{aligned}$$

(writing $s = r - 2$ for the last step). Hence, invoking the infinite version of the triangle inequality (from Analysis I),

$$\begin{aligned}
\left|\frac{f(z+h) - f(z)}{h} - g(z)\right| &= \left|\sum_{k=2}^{\infty} c_k \left(h \sum_{s=0}^{k-2} \frac{k!}{(k-s-2)!(s+2)!} h^s z^{k-2-s}\right)\right| \\
&\leqslant |h| \sum_{k=2}^{\infty} \sum_{s=0}^{k-2} |c_k| \frac{k!}{(k-s-2)!(s+2)!} |h|^s |z|^{k-2-s} \\
&\leqslant |h| \sum_{k=2}^{\infty} k(k-1)|c_k| \left(\sum_{s=0}^{k-2} \frac{(k-2)!}{(k-2-s)!s!} |h|^s |z|^{k-2-s}\right) \\
&= |h| \sum_{k=2}^{\infty} k(k-1)|c_k|(|z| + |h|)^{k-2}.
\end{aligned}$$

Fix z and choose ρ with $|z| < \rho < R$, so that $|z| + |h| < \rho$ whenever $|h| < \rho - |z|$. By the Lemma, used twice over, $\sum_{k=2}^{\infty} k(k-1)|c_k|\rho^{k-2}$ converges, to a finite constant independent of h. We conclude that $f'(z)$ does indeed exist and equal $g(z)$.

To prove the final claim, use induction on n. □