Real Analysis

S. P. Eveson
Department of Mathematics, University of York
simon.eveson@york.ac.uk

Spring/Summer terms 2022/23

Contents

1 Objectives

Principal questions to be addressed

What are the real numbers? We follow the *axiomatic* approach introduced in Mathematical Skills I, describing a minimal list of properties of the real numbers, from which all other properties follow.

What is a limit? We describe precisely ideas like "when x is very large, $1/x$ is very small, so $1/x \to 0$ as $x \to \infty$." Based on this, we can create rigorous definitions of important ideas such as infinite sums and derivatives.

What is an integral? Guided by the intuitive idea of the "area under the curve", we express the integral in terms of the basic properties of real numbers, and see why the Fundamental Theorem of Calculus (the connection between differentiation and integration) works.

How are the elementary functions of Calculus defined? We find rigorous definitions of the exponential, logarithm and trigonometric functions, from which their important properties can be proved.

Fundamental properties of the exponential and logarithmic functions

To show that our definitions match our expectations, we need to show that:

E1 There are bijective mappings $\exp : \mathbb{R} \to (0, \infty)$ and $\log : (0, \infty) \to \mathbb{R}$ which are inverses to each other: $\log(\exp(x)) = x$ for $x \in \mathbb{R}$ and $\exp(\log(y)) = y$ for $y \in (0, \infty)$. Both functions are strictly increasing:

$$x_1 < x_2 \implies \exp(x_1) < \exp(x_2); \qquad 0 < y_1 < y_2 \implies \log(y_1) < \log(y_2)$$

E2 We have the *functional equations*

$$\exp(x_1 + x_2) = \exp(x_1)\exp(x_2); \qquad \log(y_1 y_2) = \log(y_1) + \log(y_2)$$

for $x_1, x_2 \in \mathbb{R}$ and $y_1, y_2 > 0$, which imply the special values

$$\exp(0) = 1; \qquad \log(1) = 0$$

E3 The only solution to the Initial Value Problem

$$f'(x) = f(x); \qquad f(0) = 1$$

on $\mathbb{R}$ is $f(x) = \exp(x)$. The logarithm satisfies

$$\frac{d}{\mathrm{d}y} \log(y) = \frac{1}{y}$$

for $y > 0$.

E4 If $x \in \mathbb{R}$ and $y > 0$ then the relationship

$$y^x = \exp(x \log(y))$$

consistently extends the idea of integer powers (defined by repeated multiplication) and integer roots (defined as inverses of integer powers) to arbitrary real powers of positive real numbers, satisfying the identities

$$y^{x_1+x_2} = y^{x_1} y^{x_2}; \qquad (y^{x_1})^{x_2} = y^{x_1 x_2}$$

We define $\mathrm{e} = \exp(1)$, so $\mathrm{e}^x = \exp(x)$.

Fundamental properties of the trigonometric functions

To show that our definitions match our expectations, we need to show that:

T1 There are mappings $\sin : \mathbb{R} \to \mathbb{R}$ and $\cos : \mathbb{R} \to \mathbb{R}$

T2 We have the trigonometric addition formulae

$$\left.\begin{aligned} \sin(x+y) &= \sin(x)\cos(y) + \cos(x)\sin(y) \\ \cos(x+y) &= \cos(x)\cos(y) - \sin(x)\sin(y) \end{aligned}\right\} \quad (x, y \in \mathbb{R})$$

which, under the additional hypothesis that neither sin nor cos is identically zero, lead to the special values $\sin(0) = 0$ and $\cos(0) = 1$. They also give us the Pythagoras identity $\cos^2(x) + \sin^2(x) = 1$, from which it follows that $\sin(x)$ and $\cos(x)$ lie in the interval $[-1, 1]$, and the facts that sin is odd and cos is even.

T3 The only solution to the coupled Initial Value Problem

$$u'(x) = v(x), \quad v'(x) = -u(x); \qquad u(0) = 0, \quad v(0) = 1 \qquad (x, y \in \mathbb{R})$$

is $u(x) = \sin(x)$, $v(x) = \cos(x)$.

T4 There is a minimal strictly positive solution x to the equation $\sin(x) = 0$. We call this solution π. We have $\cos(\pi) = -1$, both sin and cos are 2π-periodic (and have no shorter period) and we can derive the usual simple values for the sines and cosines of $\pi/6$, $\pi/3$, $\pi/2$, etc.

Part I
Numbers

2 Axioms for the real numbers

The term "axiom" means, in common English, a self-evident, unarguable fact. In Mathematics it has a slightly different meaning: it means a fundamental assumption, a basis on which we build a mathematical theory.

We don't prove axioms: we assume they are true, and explore the consequences.

Typically, we find a small set of axioms that describe a system which has some important properties or applications. In this section, we introduce a set of axioms describing the real number system, denoted $\mathbb{R}$. Our geometric intuition is that real numbers describe points on a line, the *number line*. The system needs to be large enough to hold all the following numbers:

- the natural numbers $\mathbb{N} = \{1, 2, 3, \dots\}$;
- the integers $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$;
- the rational numbers $\{p/q : p \in \mathbb{Z}, q \in \mathbb{N}\}$;
- square, cube, etc. roots of all its positive elements;
- special numbers like e and π that arise in Calculus and other areas;
- exponentials, sines, cosines, ... of all its elements, logarithms of its positive elements;
- lots more!

2.1 Definition. The real number system $\mathbb{R}$ is a set of numbers with:

- Two special numbers, denoted 0 and 1 (with $0 \neq 1$), called the additive and multiplicative identities.
- Two basic algebraic operations of addition and multiplication, denoted by $+$ and $\cdot$. These are called *binary* operations, because they work with two numbers; for $x, y \in \mathbb{R}$ we write $x + y$ and $x \cdot y$ for the sum and product, with $x \cdot y$ usually abbreviated to xy as usual.
- Two inverse operations denoted by $-$ and $\cdot^{-1}$. These are called *unary* operations because the work with one number; for $x \in \mathbb{R}$ we write $-x$ and, if $x \neq 0$, x^{-1} for the outcomes.
- An order relation denoted $<$; for any $x, y \in \mathbb{R}$, the statement $x < y$ is either true or false (and not both!)

Apart from $x < y$, which is a true/false value, all the expressions above are real numbers, i.e. elements of $\mathbb{R}$: $0 \in \mathbb{R}$, $1 \in \mathbb{R}$, if $x, y \in \mathbb{R}$ then $x + y \in \mathbb{R}$, $xy \in \mathbb{R}$, $-x \in \mathbb{R}$ and $x^{-1} \in \mathbb{R}$; this property is called *closedness* of the operations.

From these operations, we can define a few others: for real numbers x and y

- Subtraction is defined by $x - y = x + (-y)$
- If $y \neq 0$, division is defined by $x/y = xy^{-1}$.
- The other order relations are defined in terms of $<$:

$$\begin{array}{ll} x \leq y & \text{means } x < y \text{ or } x = y \\ x > y & y < x \\ x \geq y & x > y \text{ or } x = y \end{array}$$

- For now, define $x^2 = x \cdot x$; we shall return to integer powers later (Definition 3.2).

This uses the relation of equality, which we assume is already defined as part of the statement that $\mathbb{R}$ is a set. Without going too deeply into mathematical logic, and indeed philosophy, the simplest interpretation of equality is the *substitution property*: $x = y$ means that, in any logical statement, x and y can be interchanged without affecting the truth of the statement. For the real number system, we can take the view that $x = y$ means that in any algebraic expression, x and y can be interchanged without affecting the numerical value of the expression. Look up *Leibniz's Law* for more on this, if it interests you.

We now begin to list the axioms describing how the operations and the order relation interact in the real number system.

2.2 Axiom (Axioms of arithmetic). *For all $x, y, z \in \mathbb{R}$*

A1 $x + y = y + x$ *(commutative law)*

A2 $x + (y + z) = (x + y) + z$ *(associative law)*

A3 $x + 0 = x$ *(additive identity element)*

A4 $x + (-x) = 0$ *(additive inverse)*

M1 $xy = yx$ *(commutative law)*

M2 $x(yz) = (xy)z$ *(associative law)*

M3 $1 \neq 0$ *and* $x \cdot 1 = x$ *(existence of a multiplicative identity element)*

M4 *for all $x \neq 0$ there exists an element x^{-1} such that $x \cdot x^{-1} = 1$ (existence of multiplicative inverses)*

D *for all x, y and z, $x(y + z) = xy + xz$ (distributive law)*

All the purely algebraic properties of the real numbers can be deduced from this list. Some examples:

2.3 Example. Suppose $x \in \mathbb{R}$.

1. $-(-x)$ is defined by the property (A4) that $(-x) + (-(-x)) = 0$. Add x to the left of both sides and cancel x with $-x$ to give $-(-x) = x$.

2. Start with the property $0 + 0 = 0$ (A3). Multiply by x and expand by the distributive law (D) to give $0x + 0x = 0x$. Add $-(0x)$ to both sides to give $0x = 0$.

3. Now take $1 + (-1) = 0$ (A4) and multiply by x to give $1 \cdot x + (-1) \cdot x = 0$ (D), hence $x + (-1) \cdot x = 0$ (M3). Add $(-x)$ to both sides to give $(-1) \cdot x = -x$.

4. Now combine the first and third items above: $(-1)^2 = (-1)(-1) = -(-1) = 1$. Then $(-x)^2 = [(-1)x]^2 = (-1)^2 x^2 = x^2$ using M1 and M2 to reorder the product and $(-1)^2 = 1$

Associativity was used several times above without being mentioned.

Many other mathematical systems, for example the rational numbers and the complex numbers, share these properties. Such systems are called *fields*.

Compare with the definition of a *group* in Mathematical Skills I: A1–A4 say that $(\mathbb{R}, +)$ is an Abelian (commutative) group with identity element 0. Similarly, if we let $\mathbb{R}^* = \mathbb{R} \setminus \{0\}$, then M1–M4 say that $(\mathbb{R}^*, \cdot)$ is an Abelian group with identity element 1. D ties together the additive and multiplicative structures.

The way the order relation works can also be captured described by a short list of properties.

2.4 Axiom (Axioms of order)**.** *For any $x, y, z \in \mathbb{R}$:*

O1 *exactly one of the three statements $x < y$, $x = y$, $y < x$ is true (law of trichotomy)*

O2 *if $x < y$ and $y < z$ then $x < z$ (transitive law)*

O3 *if $x < y$ then $x + z < y + z$ (compatibility with addition)*

O4 *if $x < y$ and $z > 0$ then $xz < yz$ (compatibility with multiplication)*

Again, there are other mathematical systems, such as the rational numbers, which share these axioms. Some fields do not, though: for example, there is no way to put an order relation satisfying these axioms on the complex numbers (see Example 2.5 below).

Once again, all the basic properties of the order relation follow from these. A few examples:

2.5 Example. Suppose $x \in \mathbb{R}$ with $x \neq 0$. Then (O1) we have $x < 0$ or $x > 0$.

if $x > 0$ multiply both sides by x to give $x^2 > 0$

if $x < 0$ add $-x$ to both sides to give $0 < -x$. Multiply both sides by $(-x)$ to give $0 < (-x)^2$. But $(-x)^2 = x^2$ (Example 2.3(4)), so $x^2 > 0$.

So, we have shown that all non-zero numbers have strictly positive squares. This shows that the complex numbers cannot be ordered: if they were, we would have $-1 = \mathrm{i}^2 > 0$ and $1 = 1^2 > 0$ which add to give the contradiction $0 > 0$.

For the sake of brevity, we will not prove absolutely everything in this level of detail!

From now on, we assume that we can do algebra (*finite* sequences of algebraic operations) in the usual way, but with the understanding that everything we do can be deduced from the axioms of arithmetic and order.

3 Properties of the real numbers

3.1 Definition. The familiar set $\mathbb{N}$ of *natural numbers* in $\mathbb{R}$ is characterised by the following three facts:

1. $1 \in \mathbb{N}$
2. if $n \in \mathbb{N}$ then $n+1 \in \mathbb{N}$;
3. if $S \subseteq \mathbb{N}$ satisfies (1) and (2) (i.e. $1 \in S$ and $n \in S \implies n+1 \in S$) then $S = \mathbb{N}$ (principle of induction)

(compare with the Peano axioms, mentioned in Mathematical Skills I).

That is, the natural numbers include $1, 2, 3, \ldots$ and nothing in between. Some authors include 0 but in this course we use the convention that 1 is the smallest natural number. We occasionally use the notation

$$\mathbb{N}_0 = \mathbb{N} \cup \{0\} = \{0, 1, 2, 3, \ldots\}$$

The *integers* $\mathbb{Z}$ are defined by

$$\mathbb{Z} = \{n \in \mathbb{R} : n = 0 \text{ or } n \in \mathbb{N} \text{ or } -n \in \mathbb{N}\}.$$

The *rational numbers* $\mathbb{Q}$ are defined by

$$\mathbb{Q} = \left\{ \frac{p}{q} : p \in \mathbb{Z}, q \in \mathbb{N} \right\}$$

The natural numbers are closed under addition and multiplication; the integers are closed under addition, subtraction and multiplication; the rational numbers are closed under addition, subtraction, multiplication and division (not by zero).

Having set up the integers, we can define the integer powers of real numbers.

3.2 Definition. Given $x \in \mathbb{R}$, define $x^0 = 1$. Then for $n \in \mathbb{N}$ define x^n inductively by

$$x^n = x x^{n-1}$$

If $x \neq 0$ and $n \in \mathbb{N}$ then define $x^{-n} = (x^{-1})^n$ (where x^{-1} is the multiplicative inverse in Axiom 2.2, the Axioms of Artithmetic).

Remark.

- Note that this definition states that $0^0 = 1$. This is consistent with e.g. the usual way that polynomials are described as $\sum_{n=0}^{N} a_n x^n$, with x^0 meanng 1 even for $x = 0$.
- Checking that the algebraic properties

$$(ab)^p = a^p b^p; \qquad a^{p_1+p_2} = a^{p_1} a^{p_2}; \qquad (a^{p_1})^{p_2} = a^{p_1 p_2}$$

  ($a, b \in \mathbb{R}$, $p, p_1, p_2 \in \mathbb{Z}$) is a routine exercise.
- Technically, the validity of this definition has to proved by induction; look up the *Recursion Theorem* if interested.

Non-integer powers of real numbers are considerably more problematical, and will not be fully resolved until near the end of the course. We begin with a, probably familar, results that demonstrates some of the difficulties: in the rational number system, we cannot find fractional powers.

3.3 Theorem. *There is no rational number x such that $x^2 = 2$.*

Proof. The proof is based on the fact that even numbers have even squares, and odd numbers have odd squares. Underlying this is the fact that every integer $n \in \mathbb{Z}$ is either even (of the form $n = 2k$ for some $k \in \mathbb{Z}$) or odd (of the form $n = 2k + 1$ for some $k \in \mathbb{Z}$), and no integer is both even and odd. This is proved in Theorem A.15, for anyone who would like to get right down to the foundations!

With this understanding, the property about squares is easy to check:

$$(2k)^2 = 4k^2 = 2(2k^2); \qquad (2k+1)^2 = 4k^2 + 4k + 1 = 2(2k^2 + 2k) + 1$$

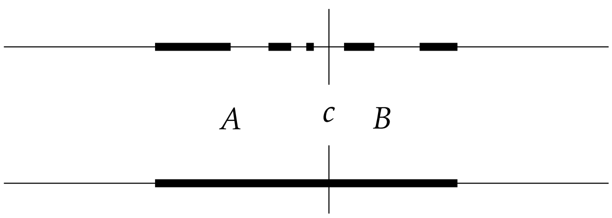
Now, we suppose for a contradiction that $x \in \mathbb{Q}$ and that $x^2 = 2$. We can write $x = p/q$ where $p \in \mathbb{Z}$, $q \in \mathbb{N}$ and p and q have no common factors. The equation $x^2 = 2$ now reads $(p/q)^2 = 2$, or $p^2 = 2q^2$.

We see that p^2 is even, so p must be even. If we write $p = 2k$ then we have $(2k)^2 = 2q^2$, so $2k^2 = q^2$. Now, since q^2 is even, q must be even.

We have now shown that p and q are both even, so they have a common factor of 2. But they were chosen to have no common factor: contradiction. □

Both the real numbers and the rational numbers satisfy all the axioms of arithmetic and order introduced so far. But in the rational numbers, $x^2 = 2$ has no solution, whereas in the real numbers we want a solution. What extra property do the real numbers have, which guarantees the existence of square roots?

3.4 Axiom (Axiom of Completeness)**.** *Suppose A and B are non-empty subsets of $\mathbb{R}$ with the property that if $a \in A$ and $b \in B$ then $a \le b$. Then there exists $c \in \mathbb{R}$ such that for all $a \in A$ and $b \in B$, $a \le c \le b$.*



The idea is that there are no gaps in the real line; if you separate the line into a left-hand part (the set A) and a right-hand part (the set B), then there is either a space between A and B (first diagram) or at least a point of the line (first diagram, the point c) identifying where the break happened.

We shall now construct the real square root of 2, using the Axiom of Completeness. We begin with a lemma that is useful in this and many other contexts.

3.5 Lemma. *Suppose $u, v \in \mathbb{R}$ and $c, h_0 > 0$. Then*

$$(1) \qquad (\text{for all } h \text{ with } 0 < h < h_0,\ u \le v + ch) \implies (u \le v)$$
$$(2) \qquad (\text{for all } h \text{ with } 0 < h < h_0,\ u \ge v - ch) \implies (u \ge v)$$

Proof. Later in the course, we will be able to interpret this as a limit: letting $h \to 0+$ on the LHS leads to the RHS. For now, though, the proof is purely about numbers: consider the equivalent contrapositive version of (1):

$$u > v \implies \text{there exists } h \text{ with } 0 < h < h_0 \text{ and } u > v + ch$$

Such an h must satisfy $h > 0$, $h < h_0$ and $h < (u - v)/c$. We can choose e.g. $h = \min(h_0/2, (u-v)/(2c))$, showing that the contrapositive and therefore the original statement is true. The proof of (2) is similar. □

3.6 Example. We can use the Axiom of Completeness to construct square roots. For simplicity, we construct $\sqrt{2}$. We do this by letting

$$A = \{a \in \mathbb{R} : a > 0 \text{ and } a^2 < 2\}$$
$$B = \{b \in \mathbb{R} : b > 0 \text{ and } b^2 > 2\}$$

We shall show firstly that A and B satisfy the hypotheses of the Axiom of Completeness, so there a real number x such that $a \le x \le b$ for all $a, b > 0$ with $a^2 < 2$ and $b^2 > 2$. Then (this is where the work comes) we shall show that $x^2 = 2$.

Note first the following fact about squares: if $a, b > 0$ then we have

$$a^2 < b^2 \iff a^2 - b^2 < 0 \iff (a-b)\underbrace{(a+b)}_{>0} < 0 \iff a - b < 0 \iff a < b$$

Now, if $a \in A$ and $b \in B$ then $a, b > 0$ and $a^2 < 2 < b^2$, so $a < b$. This verifies the main hypothesis for the Axiom of Completeness. We also need to check that $A \neq \emptyset$ and $B \neq \emptyset$, which we do by exhibiting explicit elements of A and B: we can e.g. note that $1 > 0$ and $1^2 < 2$, so $1 \in A$, and $2 > 0$ and $2^2 > 2$, so $2 \in B$.

Now, the Axiom of Completeness shows that there exists $x \in \mathbb{R}$ such that for all $a \in A$ and $b \in B$, $a \le x \le b$. Note that, because $1 \in A$ and $2 \in B$, we have $1 \le x \le 2$.

Consider $x + h$ where $0 < h < 1$. For any $a \in A$, $x \ge a$ so $x + h > a$; it follows that $x + h \notin A$ and hence $(x+h)^2 \ge 2$, so $x^2 + 2xh + h^2 \ge 2$; but $h^2 < h$ so $x^2 + (2x+1)h > 2$. Because this is true for all $h \in (0,1)$, we can conclude from Lemma 3.5 that $x^2 \ge 2$.

Similarly, consider $x - h$ where $0 < h < 1$ (so $x - h > 0$). For any $b \in B$, $x \le b$ so $x - h < b$; it follows that $x - h \notin B$ and hence $(x-h)^2 \le 2$, so $x^2 - 2xh + h^2 \le 2$; but $h^2 > 0$

so $x^2 - 2xh < 2$. Because this is true for all $h \in (0,1)$, we can conclude from Lemma 3.5 that $x^2 \leq 2$.

We now have $x^2 \leq 2$ and $x^2 \geq 2$, so $x^2 = 2$. Finally, we note that this is the unique positive square root of 2: if also $y > 0$ and $y^2 = 2$ then $0 = x^2 - y^2 = (x-y)(x+y)$ so (because $x + y > 0$) $x - y = 0$.

Remark. As we have seen, the Axiom of Completeness allows us to construct $\sqrt{2}$ in $\mathbb{R}$. We know (Theorem 3.3) that no such number exists in $\mathbb{Q}$; it follows that the Axiom of Completeness does not hold in $\mathbb{Q}$. Specifically, if

$$A = \{a \in \mathbb{Q} : a > 0 \text{ or } a^2 < 2\}; \qquad B = \{q \in \mathbb{Q} : b > 0 \text{ and } b^2 > 2\}$$

Then there is no number $x \in Q$ such that $a < x < b$ for every $a \in A$ and $b \in B$: if there were such a number, then the same argument as above would gives $x^2 = 2$, and no such rational number exists.

3.7 Definition. Arguments similar to the above, but more complicated, show that if $c > 0$ and $q \in \mathbb{N}$ then the equation $x^q = c$ has a unique positive real solution x, which we denote $c^{1/q}$. As usual, we often write $\sqrt{c}$ for $c^{1/2}$. If $p \in \mathbb{Z}$ and p, q have no common factors (so p/q is a fraction is lowest terms) then we define $c^{p/q}$ by $(c^p)^{1/q}$ (recall that c^p was defined above (Definition 3.2) by repeated multiplication).

Checking that the algebraic properties

$$(cd)^r = c^r d^r; \qquad c^{r_1+r_2} = c^{r_1} c^{r_2}; \qquad (c^{r_1})^{r_2} = c^{r_1 r_2}$$

($c, d > 0$, $r, r_1, r_2 \in \mathbb{Q}$) hold is a routine exercise.

We also define $0^{p/q} = 0$ for positive $p/q \in \mathbb{Q}$. Negative powers of 0 cannot be defined, of course.

Defining fractional powers of negative numbers, while possible to some extent, is problematical and we shall not go down that route.

We cannot yet assign a meaning to c^y where y is irrational; for this, we need the exponential and logarthmic functions.

We now introduce some notation and terminology that allows us to express the Axiom of Completeness in a particularly useful way.

3.8 Definition. A subset S of $\mathbb{R}$ is said to be *bounded above* if there is a constant $b \in \mathbb{R}$ such that $x \leq b$ for all $x \in S$. Such a constant is called an *upper bound* for S.

S is said to be *bounded below* if there is a constant $a \in \mathbb{R}$ such that $a \leq x$ for all $x \in S$. Such a constant is called a *lower bound* for S.

S is called *bounded* if it is bounded both above and below, i.e. there exist a and b such that $a \leq x \leq b$ for all $x \in S$.

A somewhat related definition, which should be familiar from Calculus or elsewhere, is that of an *interval*:

3.9 Definition. Suppose $a, b \in \mathbb{R}$. We define sets (*intervals*) as follows:

$(a,b) = \{x \in \mathbb{R} : a < x < b\}$	open, bounded interval
$(a,b] = \{x \in \mathbb{R} : a < x \leq b\}$	half-open, bounded interval
$[a,b) = \{x \in \mathbb{R} : a \leq x < b\}$	half-open, bounded interval
$[a,b] = \{x \in \mathbb{R} : a \leq x \leq b\}$	closed, bounded interval, a.k.a. compact interval
$(-\infty,b) = \{x \in \mathbb{R} : x < b\}$	open half-line, a.k.a. open infinite interval
$(-\infty,b] = \{x \in \mathbb{R} : x \leq b\}$	closed half-line, a.k.a. closed infinite interval
$(a,\infty) = \{x \in \mathbb{R} : x > a\}$	open half-line, a.k.a. open infinite interval
$[a,\infty) = \{x \in \mathbb{R} : x \geq a\}$	closed half-line, a.k.a. closed infinite interval
$(-\infty,\infty) = \mathbb{R}$	whole line

The first four are mostly of use if $a < b$. If $b < a$, they all just give the empty set. If $a = b$, the first three formulae give the empty set, but the fourth gives a singleton (a set with exactly one element): $[a,a] = \{x \in \mathbb{R} : a \leq x \leq a\} = \{a\}$. Authors are not entirely consistent (and some are quite unclear!) on whether they consider these special cases as intervals. A few authors only include the first four, bounded, cases as intervals.

For the purposes of this course, an *interval* is any one of the sets described above, including singletons $[a,a] = \{a\}$ but excluding the empty set.

3.10 Example. Fix $a, b \in \mathbb{R}$ with $a < b$.

$$(a,b) = \{x \in \mathbb{R} : a < x < b\}$$

is bounded below, with lower bound a (or anything less than a) and bounded above, with upper bound b (or anything bigger than b).

$$(a,\infty) = \{x \in \mathbb{R} : a < x\}$$

is bounded below, with lower bound a (or anything smaller), but is not bounded above.

$$(-\infty,b) = \{x \in \mathbb{R} : x < b\}$$

is bounded above, with upper bound b (or anything bigger), but is not bounded below.

$\mathbb{R}$ itself is bounded neither above nor below.

As these examples emphasise, upper and lower bounds are not unique: if you have one, you have lots!

For a final example, the set $\{1/n : n \in \mathbb{N}\}$ is bounded. This is because if $n \in \mathbb{N}$ then $n > 0$, so $1/n > 0$, and $n \geq 1$, so $1/n \leq 1$. We thus have a lower bound of 0 and an upper bound of 1.

3.11 Definition. Suppose $S \subseteq \mathbb{R}$. We say that S has a *maximum/maximal element*, denoted by $\max(S)$, if it contains an element larger than all other elements; that is, an upper bound b which is an element of S. Similarly, we say that S has a *mimimum/minimal element*, denoted by $\min(S)$, if it contains an element smaller than all others; that is, if it has a lower bound a which is an element of S.

Note that, unlike upper and lower bounds, maximum and minimum elements are, if they exist at all, unique. To see this, suppose that b_1 and b_2 are both maximal elements of S. Then they are both elements of S, and both upper bounds for S. This means that

$$\underset{\text{element of } S}{b_1} \leq \underset{\text{upper bound for } S}{b_2}$$

$$\underset{\text{element of } S}{b_2} \leq \underset{\text{upper bound for } S}{b_1}$$

and so $b_1 = b_2$. The argument for minimum elements is similar.

3.12 Example. Consider the set

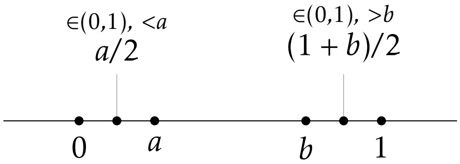
$$S = (0,1) = \{x \in \mathbb{R} : 0 < x < 1\}$$

Plainly, 0 is a lower bound for S; it follows that any number less than 0 is also a lower bound for S. Any number greater than 0, however, is not. To see this, suppose $a > 0$. If $a \geq 2$ then a is certainly not a lower bound. If $0 < a < 2$ then $0 < a/2 < a < 1$ so $a/2 < a$ and $a/2 \in S$, showing that a is not a lower bound.

The lower bounds of S in $\mathbb{R}$ are thus $\{a \in \mathbb{R} : a \leq 0\}$. None of these are elements of S, so S has no minimal element.

Similarly, 1 and all numbers larger than 1 are upper bounds for S, but any number smaller than 1 is not: any $b \leq 0$ is plainly not an upper bound for S and if $0 < b < 1$ then $(1+b)/2 \in S$ and $(1+b)/2 > b$, showing that B is not a lower bound.

The upper bounds of S in $\mathbb{R}$ are thus $\{b \in \mathbb{R} : b \geq 1\}$. None of these are elements of S, so S has no maximal element.



Even though S does not have a maximum or a minimum, its lower bounds have a maximum, 0, and its upper bounds have a minimum, 1. These coincide with the intuitive idea that 0 and 1 mark the extremities of the set S.

The set $S' = [0,1] = \{x \in \mathbb{R} : 0 \leq x \leq 1\}$ has exactly the same lower and upper bounds as the set S. In this case, 0 is both a lower bound and an element of S', and 1 is both an upper bound and an element of S'. We thus have $\min(S') = 0$ and $\max(S') = 1$.

There is, of course, nothing special about $(0,1)$ and $[0,1]$: if $\alpha, \beta \in \mathbb{R}$ with $\alpha < \beta$ then the set of upper bounds of the open interval (α, β) is the closed half-line $[\beta, \infty)$, and the set of lower bounds is the closed half-line $(-\infty, \alpha]$. The upper and lower bounds of the closed interval $[\alpha, \beta]$ are exactly the same as those of the open interval (α, β). We can show that this is typical of all bounded sets of real numbers using the Axiom of Completeness. Recall the statement:

Suppose A and B are non-empty subsets of $\mathbb{R}$ with the property that if $a \in A$ and $b \in B$ then $a \leq b$. Then there exists $c \in \mathbb{R}$ such that for all $a \in A$ and $b \in B$, $a \leq c \leq b$.

In terms of bounds, the hypotheses are that every element of A is a lower bound for B and every element of B is an upper bound for A; the conclusion is that c is an upper bound for A and a lower bound for B. This connection between bounds and the Axiom of Completeness allows us to construct bounds with special properties.

3.13 Theorem. *Suppose S is a non-empty subset of $\mathbb{R}$. Then:*

1. *If S is bounded above, then it has a minimal / least upper bound, called the* supremum *of S and written* $\sup(S)$.
2. *If S is bounded below, then it has a maximal / greatest lower bound, called the* infimum *of S and written* $\inf(S)$.

Proof.

1. Suppose S is bounded above. Let B be the (non-empty) set of all upper bounds of S; if $x \in S$ and $b \in B$ then $x \leq b$ by the definition of an upper bound. By the Axiom of Completeness, there is a real number c such that for all $x \in S$ and $b \in B$, $x \leq c \leq b$. The first inequality, $x \leq c$ for all $x \in S$, says that c is an upper bound for S; the second inequality, $c \leq b$ for all $b \in B$, says that c is the least upper bound for S.
2. Almost identical: if S is bounded below, let A be the set of lower bounds for S, then by the Axiom of Completeness there is a real number c such that for all $a \in A$ and $x \in S$, $a \leq c \leq x$. This c is a lower bound for S greater than all other lower bounds of S, i.e. the greatest lower bound of S.

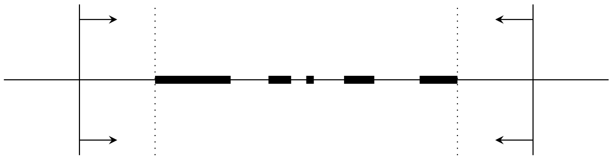
□

3.14 Remark.

1. The supremum of a non-empty set that is bounded above is, by definition, the minimum of the upper bounds. Since the minimum, when it exists, is unique (Definition 3.11), the supremum is also unique. Similarly, the infimum of a non-empty set that is bounded below is unique.
2. Looking back at Example 3.12, we can see that 0 is the greatest lower bound, i.e. the infimum, of both $(0,1)$ and $[0,1]$ and that 1 is the least upper bound, i.e. the supremum, of both $(0,1)$ and $[0,1]$.
3. The supremum can be thought of as a substitute for the maximum, when the maximum does not exist. If a set S has a maximum, b, then by definition b is an upper bound for S. If $b' < b$ then b' is not an upper bound, since $b \in S$. This shows that the maximum, if it exists, is also the supremum. However, as Example 3.12 shows, some sets have a supremum but do not have a maximum.

   Similarly, if a set has a minimum then that is also its infimum, but a set can have an infimum but not minimum.

3.15 Remark. The supremum and infimum of a set can be pictured on the number line.



Start off with a lower and an upper bound for a set, illustrated here by the solid vertical lines with arrows on them. Imagine sliding them towards the set until they just touch it: if you were to move them any more, they would overlap the set and not longer be bounds. At this point, you have have found the greatest lower and the least upper bounds, i.e. the infimum and supremum, shown here are dotted vertical lines. If the point where you stopped lies in the set, you have a minimum or maximum, but it might not.

3.16 Remark. Suppose S is a non-empty subset of $\mathbb{R}$ and b is an upper bound for S. Here are three equivalent ways of expressing the extra property that b is the *least* upper bound.

1. if $b' < b$ then b' is not an upper bound (the original definition);
2. if $b' < b$ then there exists $x \in S$ such that $x > b'$ (expanding the statement 'is not an upper bound');
3. if $h > 0$ then there exists $x \in S$ such that $x > b - h$ (changing notation from b' to $b - h$; $b' < b$ is equivalent to $h > 0$).

In working with suprema, we can use whichever of these we like; in practice, the third one is often (but not always) the most useful. There are corresponding results for infima; e.g., if a is a lower bound of S, then it is the greatest lower bound if and only if for every $h > 0$ there exists $x \in S$ such that $x < a + h$.

The ideas of spremum and infimum allow us to establish a useful characterisation of intervals, sometimes (e.g. Thomas Calculus(used as th

3.17 Theorem. *A non-empty set $I \subseteq \mathbb{R}$ is an interval if and only if it has the* intermediate value property*: if $x \in I$, $z \in I$ and $y \in \mathbb{R}$ with $x < y < z$ then $y \in I$; that is, if I contains two numbers x and z then it contains every number y between x and z.*

Proof. It is clear that intervals, as defined above, satisfy the intermediate value property.

The proof of the reverse implication is straightforward but repetitive: the left-hand endpoint, when finite, is $\inf(I)$ and the right-hand endpoint, where finite, is $\sup(I)$. Here is one reasonably efficient presentation. Suppose $I \neq \emptyset$ has the intermediate value property and fix $x_0 \in I$. We show first:

If I is unbounded above then for any $x > x_0$ there exists $y \in I$ with $y > x$ (otherwise x would be an upper bound for I). Since $x_0 < x < y$ and $y \in I$, we have $x \in I$. So, $[x_0, \infty) \subseteq I$.

If I is unbounded below then a similar argument gives $(-\infty, x_0] \subseteq I$.

If I is bounded above then for any $x < \sup(I)$ there exists $y \in I$ with $y > x$ (otherwise x would be an upper bound for I). Since $x_0 < x < y$ and $y \in I$, we have $x \in I$. So, $[x_0, \sup(I)) \subseteq I$.

If I is bounded below then a similar argument gives $(\inf(I), x_0] \subseteq I$.

Now, we can consider the four possibilities of I being bounded or unbounded above and below; in each case, the arguments above give us two subsets of I, whose union is therefore contained in I.

If I is unbounded below and unbounded above then I contains $(-\infty, x_0) \cup [x_0, \infty) = \mathbb{R}$, so $I = \mathbb{R}$.

If I is unbounded below and bounded above then I contains $(-\infty, x_0) \cup [x_0, \sup(I)) = (-\infty, \sup(I))$. Moreover, because $\sup(I)$ is an upper bound for I, we have $I \subseteq (-\infty, \sup(I)]$. There is only one point in $(-\infty, \sup(I)]$ that is not in $(-\infty, \sup(I))$, namely $\sup(I)$ itself. This determines all the elements of I, except $\sup(I)$, which might be an element of I or not. These two possibilities correspond to $I = (-\infty, \sup(I)]$ and $I = (-\infty, \sup(I))$, respectively.

If I is bounded below and unbounded above then a similar argument gives either $I = [\inf(I), \infty)$ or $I = (\inf(I), \infty)$.

If I is bounded below and bounded above then I contains $(\inf(I), x_0] \cup [x_0, \sup(I)) = (\inf(I), \sup(I))$. Moreover, because $\inf(I)$ and $\sup(I)$ are respectively lower and upper bounds for I, we have $I \subseteq [\inf(I), \sup(I)]$. This determines all the elements of I, except for $\inf(I)$ and $\sup(I)$, which might be elements of I or not. The four possibilities of these two points being elements or not give rise to four possible cases: $I = (\inf(I), \sup(I))$, $I = (\inf(I), \sup(I)]$, $I = [\inf(I), \sup(I))$ and $I = [\inf(I), \sup(I)]$.

□

The next property seems innocuous but is actually very important.

3.18 Axiom (Axiom of Archimedes/Archimedean Property)**.** *$\mathbb{N}$ is not bounded above in $\mathbb{R}$.*

Intuitively, this probably falls into the "self-evidently true" category. Perhaps surprisingly, the proof requires the axiom of completeness: it does not follow from the other axioms (look up "p-adic numbers" or "non-standard real numbers" if you'd like to know more).

Proof. Suppose for a contradiction that $\mathbb{N}$ is bounded above. Then it has a least upper bound b. For any $n \in \mathbb{N}$, $n + 1 \in \mathbb{N}$, so $n + 1 \leq b$; but now we have $n \leq b - 1$ for all $n \in \mathbb{N}$, showing that $b - 1$ is an upper bound for $\mathbb{N}$, smaller than the least upper bound b. Contradiction. □

3.19 Corollary (Equivalent formulations of the Archimedean property)**.**

1. *If $x \in \mathbb{R}$ then there exists $n \in \mathbb{N}$ such that $n > x$. [This just says that any $x \in \mathbb{R}$ is not an upper bound for $\mathbb{N}$.]*

2. *If $x \in \mathbb{R}$ and $x \leq 1/n$ for all $n \in \mathbb{N}$ then $x \leq 0$ (cf. Lemma 3.5; $1/n$ plays the role of h). [If $x > 0$ and $x \leq 1/n$ for all $n \in \mathbb{N}$, then $1/x \geq n$ for all $n \in \mathbb{N}$. This is impossible by the previous part, so any such x must satisfy $x \leq 0$.]*

3. $\inf\{1/n : n \in \mathbb{N}\} = 0$. *[Certainly 0 is a lower bound for $\{1/n : n \in \mathbb{N}\}$. If a is any other lower bound then $a \leq 1/n$ for all n, so $a \leq 0$ by the previous part; this shows that 0 is the greatest lower bound.]*

4. *For any $h > 0$ there exists $n \in \mathbb{N}$ such that $1/n < h$. [This is another way of writing the previous infimum statement: see Remark 3.16.]*

5. *If $x \in \mathbb{R}$ and $|x| \leq 1/n$ for all $n \in \mathbb{N}$ then $x = 0$. [If $|x| \leq 1/n$ for all $n \in \mathbb{N}$ then $|x| \leq 0$ by (2); it follows that $|x| = 0$, so $x = 0$.]*

Where sets are defined in terms of an integer parameter, we can sometimes determine boundedness and the supremum or infimum using Archimedes.

3.20 Example. The set

$$\{n/2 - 3 : n \in \mathbb{N}\}$$

is unbounded above. To see this, suppose for a contradiction that there is an upper bound, so we have $n/2 - 3 \leq b$ for all $n \in \mathbb{N}$. Rearranging gives $n \leq 2b + 6$ for all $n \in \mathbb{N}$, contradicting Archimedes.

Let

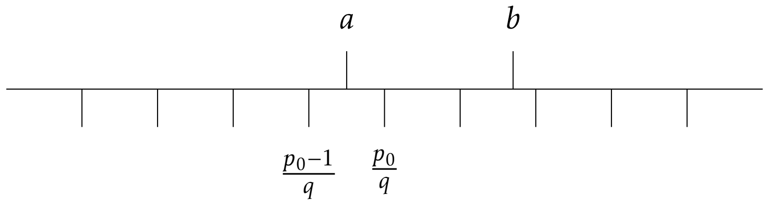
$$S = \left\{ \frac{n+2}{n+1} : n \in \mathbb{N} \right\}.$$

Clearly, 1 is a lower bound. To show that 1 is the greatest lower bound, we have to show (Remark 3.16) that for any $h > 0$ there exists $n \in \mathbb{N}$ such that $(n+2)/(n+1) < 1 + h$. Rearranging this gives the logically equivalent inequality $n > 1/h - 1$; such $n \in \mathbb{N}$ exists by Archimedes.

Another useful consequence of Archimedes is that, given any two real numbers, we can find a rational number in between them. This property is called *density*.

3.21 Theorem (Density of the rationals)**.** *Suppose $a, b \in \mathbb{R}$ with $a < b$. Then there exists $p/q \in \mathbb{Q}$ such that $a < p/q < b$.*

Proof. Since $b > a$, by the Archimedean property there exists $q \in \mathbb{N}$ such than $1/q < b - a$. The idea is that the numbers $\{p/q : p \in \mathbb{Z}\}$ are equally spaced, each at a distance $1/q$ from its neighbours; since $1/q < b - a$, one or more of the p/q must lie between a and b.



To make this precise, consider the set

$$S = \{p \in \mathbb{Z} : p/q > a\}$$

Note that $p \in S$ if and only if $p > qa$; this shows that S is bounded below and (by Archimedes) non-empty; S therefore has an infimum and we can (Remark 3.16) choose $p_0 \in S$ such that $p_0 < \inf(S) + 1$; note that $p_0 - 1 < \inf(S)$ so $p_0 - 1 \notin S$.

Now, $p_0 \in S$, so $p_0/q > a$, and $p_0 - 1 \notin S$, so $(p_0 - 1)/q \leq a$, or equivalently $p_0/q \leq a + 1/q$. But $1/q < b - a$ so $p_0/q < a + (b - a) = b$; we now have $a < p_0/q < b$, as claimed. $\square$

3.22 Remark. Density can also be interpreted in terms of approximation. Suppose x is some fixed real number (think of x as being irrational like $\sqrt{2}$), and $h > 0$. Putting $a = x - h$ and $b = x + h$ in the above result shows that we can find a rational p/q such that $x - h < p/q < x + h$. Rearranging, this gives us $-h < p/q - x < h$, or equivalently $|x - p/q| < h$. So, however small we make h, we can find a rational number p/q closer to x than h.

This idea is a familiar one in disguise: we are used to approximating real numbers by finite decimal expansions, which are rational numbers: e.g., we can approximate $\sqrt{2}$ by 1.414 in the sense that

$$1.414 < \sqrt{2} < 1.415; \qquad |\sqrt{2} - 1.414| < 0.001$$

Note that approximating by truncated decimals, while similar in spirit to inequalities like these, is not exactly the same thing. In general, if two numbers agree to n decimal places then they differ by no more than 10^{-n}, but the converse is not true: for example, 0.999 and 1.000 differ by 10^{-3} but have no decimal places in common.

4 Part I: key points

The purpose of Part I is to write down an unambiguous description of the real number system, from which all properties of real numbers can be deduced.

This description consists of a relatively small list of properties (axioms), to do with its basic algebraic operations ($+$, $\cdot$, $-$, $\div$) and its order relation ($<$). Most of these are very simple (Axioms 2.2, 2.4), but there is one much more sophisticated one: the *Axiom of Completeness* (Axiom 3.4), which says that if A and B are non-empty sets of real numbers such that $a \leq b$ for every $a \in A$ and $b \in B$ then there is a real number c that fits between the two sets, i.e. $a \leq c \leq b$ for every $a \in A$ and $b \in B$.

The Axiom of Completeness encapsulates the fundamental difference between the rational and real numbers: for example, the equation $x^2 = 2$ has no solution in $\mathbb{Q}$ (Theorem 3.3), but it does have a solution in $\mathbb{R}$, which can be expressed as the unique real number in between the sets $\{x \in \mathbb{R} : x > 0, x^2 < 2\}$ and $\{x \in \mathbb{R} : x > 0, x^2 < 2\}$ (Example 3.6). More generally, we can find rational powers of real numbers: if $a > 0$, $p \in \mathbb{Z}$ and $q \in \mathbb{N}$ then $a^{p/q}$ is well-defined (Definition 3.7). The important question of the meaning of a^x when x is irrational remains unresolved at this point, but will be addressed (again using completeness) in Part V.

The Axiom of Completeness also leads to a number of other significant order-related properties:

- every non-empty set of real numbers which is bounded below has a least upper bound or *supremum* and a greatest lower bound or *infimum* (Theorem 3.13);
- the natural numbers $\mathbb{N}$ are not bounded above in $\mathbb{R}$ (property/axiom of *Archimedes*, Axiom 3.18);
- if $x \in \mathbb{R}$ and $x \leq 1/n$ for all $n \in \mathbb{N}$ then $x \leq 0$ (property of Archimedes again, expressed in a different way, Corollary 3.19);
- between any two distinct real numbers there is a rational number (*rational density*, Theorem 3.21);
- if $x \in \mathbb{R}$ and $h > 0$ then there exists $p/q \in \mathbb{Q}$ such that $|x - p/q| < h$ (rational density expressed in a different way, Remark 3.22).

Literature:

1. F. Mary Hart *Guide to Analysis* (2nd Edition, Palgrave Macmillan 2001, JBM classmark S7 Har) Chapter 1.
2. Keith E. Hirst, *Numbers, Sequences and Series* (Edward Arnold, 1995, JBM classmark S2.81 Hir) Chapter 5 (principally; also parts of Chapters 2 and 3).
3. any other introduction to Real Analysis; browse classmark S7 in the library.

A The natural numbers, integers and rationals

This is a technical appendix, for anyone who would like a deeper understanding of the way the number system works; in particular, none of this is examinable!

As stated in Definition 3.1, within the real number system, the natural numbers can be characterised by three properties:

1. $1 \in \mathbb{N}$
2. if $n \in \mathbb{N}$ then $n+1 \in \mathbb{N}$;
3. if $S \subseteq \mathbb{N}$ satisfies (1) and (2) (i.e. $1 \in S$ and $n \in S \implies n+1 \in S$) then $S = \mathbb{N}$ (principle of induction)

Here we describe a construction of $\mathbb{N}$ within $\mathbb{R}$.

A.1 Definition. If $x \in \mathbb{R}$ then $x+1$ is known as the *successor* of x and $x-1$ as the *predecessor* of x. A subset S of $\mathbb{R}$ is called an *inductive set* if:

1. $1 \in S$
2. if $x \in S$ then $x+1 \in S$ (successor property).

A.2 Example. $\mathbb{R}$ is an inductive set, as is $\{x \in \mathbb{R} : x \geq 1\}$ and $\{x \in \mathbb{R} : x = 1 \text{ or } x \geq 2\}$. The interval $[0,1]$ is not an inductive set.

As the example shows, there are many inductive sets, most of which do not satisfy the third property of $\mathbb{N}$ mentioned above: if S is an inductive set and $x \in S$ then there could be many elements $y \in S$ with $x < y < x+1$. To isolate the natural numbers, we need to find the *smallest* inductive set.

A.3 Definition. The set $\mathbb{N}$ of natural numbers is defined to be the intersection of all inductive sets; that is, $n \in \mathbb{N}$ iff n lies in *every* inductive set. This makes $\mathbb{N}$ the smallest inductive set, in the sense that every inductive set contains $\mathbb{N}$.

A.4 Example. By definition, 1 lies in every inductive set, hence in $\mathbb{N}$. By the successor property, $1+1=2$ lies in every inductive set, hence in $\mathbb{N}$. Similarly, $3 \in \mathbb{N}$. However, $3/2$ does not lies in the inductive set $\{x \in \mathbb{R} : x = 1 \text{ or } x \geq 2\}$, so $3/2 \notin \mathbb{N}$.

We need to extend the ideas in the example to show that $\mathbb{N}$ contains $1, 2, 3, \ldots$ and nothing in between. The $\ldots$ here needs clarification, of course; this is where the principle of mathematical induction comes in.

A.5 Theorem. *$\mathbb{N}$ is an inductive set (i.e. the first two properties from Definition 3.1 hold).*

Proof. By definition, 1 lies in every inductive set, so $1 \in \mathbb{N}$. Suppose $n \in \mathbb{N}$ and S is an inductive set. Then $n \in S$ (because $n \in \mathbb{N}$) and $n+1 \in S$ (because S is inductive). This shows that $n+1$ lies in every inductive set, so $n+1 \in \mathbb{N}$. □

A.6 Theorem. *1 is the minimum element of* $\mathbb{N}$.

Proof. The set $\{x \in \mathbb{R} : x \geq 1\}$ is inductive, so by definition contains the natural numbers; that is, $n \geq 1$ for all $n \in \mathbb{N}$. Since $1 \in \mathbb{N}$, $1 = \min(\mathbb{N})$. □

A.7 Theorem (Principle of mathematical induction)**.** *If $S \subseteq \mathbb{N}$ is an inductive set, then $S = \mathbb{N}$ (i.e. the third property from Definition 3.1 holds).*

Proof. Since S is an inductive set, $\mathbb{N} \subseteq S$ by definition. In combination with the hypothesis $S \subseteq \mathbb{N}$, it follows that $S = \mathbb{N}$. □

This is the set-theoretic version of mathematical induction. To relate it to the more common form, suppose we have a proposition P about natural numbers, and we know that $P(1)$ is true and that $P(n) \implies P(n+1)$.Let $S = \{n \in \mathbb{N} : P(n) \text{ is true}\}$. The hypotheses about P show that $1 \in S$ and that if $n \in S$ then $n+1 \in S$; that is, that S is an inductive subset of $\mathbb{N}$. We now conclude from the set-theoretic induction principle above that $S = \mathbb{N}$, so, by definition of S, $P(n)$ is true for all $n \in \mathbb{N}$.

A.8 Theorem (Existence of predecessors)**.** *If $n \in \mathbb{N}$ then either $n = 1$ or $n - 1 \in \mathbb{N}$.*

Proof. We prove this by induction. Let

$$S = \{n \in \mathbb{N} : \underbrace{n = 1}_{\text{(a)}} \text{ or } \underbrace{n - 1 \in \mathbb{N}}_{\text{(b)}}\}$$

Then certainly $1 \in S$ (condition (a)). Now suppose $n \in S$. Then $n+1$ satisfies condition (b) because $(n+1) - 1 = n \in \mathbb{N}$ (because $S \subseteq \mathbb{N}$). It follows that $S = \mathbb{N}$. □

Another way to express the minimality of $\mathbb{N}$ is the property that if $n \in \mathbb{N}$ then there is no natural number between n and $n+1$.

A.9 Theorem. *If $n \in \mathbb{N}$, then there is no natural number x such that $n < x < n+1$.*

Proof. Let

$$S = \{n \in \mathbb{N} : \text{ there is no natural number } x \text{ such that } n < x < n+1 \}.$$

If $x \in \mathbb{N}$ with $1 < x < 2$ then $x-1 \in \mathbb{N}$ by the previous result. But $x-1 < 1$ and $1 = \min(\mathbb{N})$; contradiction. This shows that there is no natural number between 1 and 2, so $1 \in S$. Now suppose $n \in S$. If $x \in \mathbb{N}$ and $n+1 < x < n+2$ then $x-1 \in \mathbb{N}$ (previous result) and $n < x-1 < n+1$. This contradicts $n \in S$, so no such x exists; we therefore have $n+1 \in S$. By the principle of induction (above), $S = \mathbb{N}$, so the result holds for all $n \in \mathbb{N}$. □

A.10 Theorem (Well-ordering principle)**.** *Every non-empty subset of $\mathbb{N}$ has a minimal element.*

Proof. Again, we shall prove this by induction; for variety, this time we use the more familiar approach of statements, rather than sets. Let P_n be the statement "every subset of $\mathbb{N}$ containing an element less than or equal to n has a minimal element".

P_1 is true because 1 is the minimal natural number: if S contains an element less than or equal to 1 then it must contain 1 itself, which must be the minimal element of S.

Now suppose P_n is true for some particular n and that $S \subseteq \mathbb{N}$ contains an element less than or equal to $n+1$. We now consider two cases:

1. S contains an element less than or equal to n. In this case the induction hypothesis applies directly to S, which therefore has a minimal element.

2. S contains no element less than or equal to n. Since we are assuming that S contains an element less than or equal to $n+1$, and there are no natural numbers between n and $n+1$, it must contain $n+1$. It has no elements less than $n+1$, so $n+1$ is the minimal element of S.

□

A.11 Remark. We have now seen that the principle of mathematical induction implies that the natural numbers are well-ordered. In fact, the two properties are equivalent. Imagine that we know that the well-ordering principle holds but we do not know about induction. If $S \subseteq \mathbb{N}$ has the properties that $1 \in S$ and that if $n \in S$ then $n+1 \in S$; we can show that $S = \mathbb{N}$ using the well-ordering principle. We do this by considering the complement $S' = \{n \in \mathbb{N} : n \notin S\}$. If $S \neq \mathbb{N}$ then $S' \neq \emptyset$, so S' has a minimal element, n. Since $1 \in S$, $n \neq 1$. We now have $n-1 \in \mathbb{N}$ and $n-1 \notin S'$ because $n = \min(S')$. We therefore have $n-1 \in S$, so from the inductive hypothesis $n \in S$; but $n \in S'$ which is a contradiction. We must therefore have $S' = \emptyset$ and $S = \mathbb{N}$, which establishes the principle of mathematical induction.

The Archimedean property, that $\mathbb{N}$ is not bounded above, cannot be proved without using the Axiom of Completeness (or some other axiom). This is because it is not a property of the natural numbers themselves, but of the way they sit inside the real number system.

So far we have looked at the order properties of the natural numbers. We can also show that:

A.12 Theorem.

1. *If $m, n \in \mathbb{N}$ then $m+n \in \mathbb{N}$*

2. *If $m, n \in \mathbb{N}$ then $mn \in \mathbb{N}$*

3. *If $m, n \in \mathbb{N}$ and $m > n$ then $m-n \in \mathbb{N}$*

Proof.

1. Straightforward exercise for the reader: fix an arbitrary m and use induction on n

2. Straightforward exercise for the reader: fix an arbitrary m and use induction on n; you'll need to use part (a)

3. A bit more delicate! For $m \in \mathbb{N}$, let P_m be the statement "if $n \in \mathbb{N}$ and $n < m$ then $m-n \in \mathbb{N}$."

   Start with the induction base: P_1 is vacuously true because there is no $n \in \mathbb{N}$ with $n < 1$, so the statement has no opportunity to be false. If you don't like that, start with P_2: if $n \in \mathbb{N}$ and $n < 2$ then $n = 1$ (as we have seen above, there is no natural number between 0 and 1), so $m-n = 2-1 = 1 \in \mathbb{N}$.

   Now suppose P_m is true for some particular $m \in \mathbb{N}$. To prove P_{m+1}, suppose $n \in \mathbb{N}$ is such that $n < m+1$. We now need to consider two cases:

(a) If $n \neq 1$ then we can argue that $(m+1)-n = m-(n-1)$, $n-1 < m$ (because $n < m+1$) and $n-1 \in \mathbb{N}$ (existence of predecessors, above); it now follows from P_m that $(m+1-n) \in \mathbb{N}$.

(b) If $n = 1$ then the previous argument does not work, but the result is true for simpler reasons: $(m+1)-1 = m \in \mathbb{N}$

This proves P_{m+1}, so P_m is true for all $m \in \mathbb{N}$ by induction.

□

All the hard work went into constructing $\mathbb{N}$; it's now quite easy to build the integers. Two plausible definitions are:

$$\mathbb{Z}_1 = \{x \in \mathbb{R} : x \in \mathbb{N} \text{ or } x = 0 \text{ or } -x \in \mathbb{N}\}$$
$$\mathbb{Z}_2 = \{m-n : m, n \in \mathbb{N}\}$$

The first emphasises the order relation, the second the additive structure.

A.13 Lemma. *In the above notation,* $\mathbb{Z}_1 = \mathbb{Z}_2$

Proof. As usual, we show that two sets are equal by showing that each is a subset of the other. Suppose $x \in \mathbb{Z}_1$. Then either $x \in \mathbb{N}$, in which case $x = (n+1)-1$; or $x = 0$, in which case $x = 1-1$; or $-x \in \mathbb{N}$, in which case $x = 1-(-x+1)$. In each case we have represented x in the form $x = m-n$ for $m, n \in \mathbb{N}$, showing that $x \in \mathbb{Z}_2$.

Now suppose $x \in \mathbb{Z}_2$, so $x = m-n$ for $m, n \in \mathbb{N}$. We now proceed by trichotomy. If $m > n$ then $x \in \mathbb{N}$ by the subtraction property of $\mathbb{N}$; if $m = n$ then $x = 0$; if $m < n$ then $-x = n-m \in \mathbb{N}$, again by the subtraction property of $\mathbb{N}$ □

A.14 Definition. The integers $\mathbb{Z}$ are defined to be the set given by $\mathbb{Z}_1 = \mathbb{Z}_2$ in the above notation.

To establish the properties of $\mathbb{Z}$, we can now use whichever formulation is more convenient. For example, to show that if $x, y \in \mathbb{Z}$ then $x+y \in \mathbb{Z}$, $x-y \in \mathbb{Z}$ and $xy \in \mathbb{Z}$, the $\mathbb{Z}_2$ formulation is much easier to use. On the other hand, to show that if $x \in \mathbb{Z}$ then there is no integer between x and $x+1$, the $\mathbb{Z}_1$ formulation is easier. All these properties are left as a straightforward exercise for the reader.

Finally, we define the rationals:

Definition.
$$\mathbb{Q} = \left\{ \frac{x}{y} : x \in \mathbb{Z}, y \in \mathbb{Z}, y \neq 0 \right\}$$

It is straightforward to check that $\mathbb{Q}$ satisfes all the same axioms of arithmetic and order that $\mathbb{R}$ satisfies, except the Axiom of Completeness.

Finally, we return to the integers in support of the Theorem 3.3:

A.15 Theorem. *Every integer $n \in \mathbb{Z}$ is either even (of the form $n = 2k$ for some $k \in \mathbb{Z}$) or odd (of the form $n = 2k+1$ for some $k \in \mathbb{Z}$), and no integer is both even and odd.*

Proof. We show first that every natural number is either even or odd. Firstly, 1 is odd because $1 = 2 \cdot 0 + 1$. Now, suppose some particular $n \in \mathbb{Z}$ is either even or odd, so for some integer k we have either $n = 2k$ or $n = 2k + 1$. Then either $n + 1 = 2k + 1$ or $n + 1 = 2k + 2 = 2(k + 1)$, so $n + 1$ is either odd or even. It now follows by induction that every $n \in \mathbb{N}$ is either even or odd.

Next, if $n \in \mathbb{Z}$ then either $n = 0$, $n \in \mathbb{N}$ or $-n \in \mathbb{N}$. If $n = 0$ then $2 = 2 \cdot 0$, so n is even. If $n \in \mathbb{N}$ then n is either even or odd from the previous paragraph. If $-n \in \mathbb{N}$ then, again from the previous paragraph, $-n$ is either even or odd, so we have either $-n = 2k \implies n = 2(-k)$ or $-n = 2k + 1 \implies n = -2k - 1 = 2(-k - 1) + 1$, showing that n is either even or odd.

Finally, suppose for a contradiction that $n \in \mathbb{Z}$ is both even and odd. Then $n = 2k$ and $n = 2j + 1$, where $j, k \in \mathbb{Z}$. Subtracting, we get $2(k - j) = 1$, so $k - j = 1/2$. Now, $k - j$ is a positive integer, i.e. a natural number; but $1/2$ is not a natural number because 1 is the smallest natural number. This contradiction shows that n cannot be both even and odd. $\square$

Part II
Sequences and Series

5 Sequences, Series and Convergence

5.1 Definition. A (real) *sequence* is an infinite list of real numbers. We write

$$(a_n)_{n\geq 1} \text{ or } (a_n)_{n\in\mathbb{N}}$$

to mean the sequence

$$a_1,\ a_2,\ a_3,\ldots$$

whose nth *term* is a_n. Sometimes it is natural or convenient to consider sequences starting with, e.g., $n = 0$ or $n = 2$, and we write

$$(a_n)_{n\geq 0}, \qquad (a_n)_{n\geq 2}$$

etc. for these. If the starting point is not specified, we assume it is 1. We generally develop the theory for sequences starting at 1, and typically it makes little difference if they start at some other value.

More formally, the sequence $(a_n)_{n\geq n_0}$ is a function from the set $\{n \in \mathbb{Z} : n \geq n_0\}$ to $\mathbb{R}$.

5.2 Example. In theory, any list of numbers qualifies as a sequence. In practice, sequences are usually defined in one of a few ways:

1. as explicit functions of n; e.g.,

$$a_n = 1/n\ (n \in \mathbb{N});\ b_n = 1/(n-1)^2\ (n \geq 2);\ c_n = n\ (n \geq 0)$$

2. *recursively* or *inductively*, by specifying the first term(s) and a rule for determining subsequent terms from their predecessor(s); e.g.,

$$a_1 = 1,\ a_{n+1} = (n+1)a_n \quad (n \in \mathbb{N}) \tag{1}$$
$$b_1 = 0,\ b_2 = 1,\ b_{n+2} = (b_n + b_{n+1})/2 \quad (n \in \mathbb{N}) \tag{2}$$

3. As a *partial sum* of an existing sequence, e.g.

$$s_n = \sum_{j=1}^{n} a_j$$

Here we have two sequences: the original sequence $(a_j)_{j\in\mathbb{N}}$ and a new sequence $(s_n)_{n\in\mathbb{N}}$ formed by adding up the terms of the original sequence. See Definition 5.9 below for more on this.

5.3 Question. *What does it mean to say that a sequence (a_n) converges to a limit a as $n \to \infty$? We want to capture the intuitive idea that as n grows larger and larger, a_n becomes closer and closer to a, without ever necessarily reaching a.*

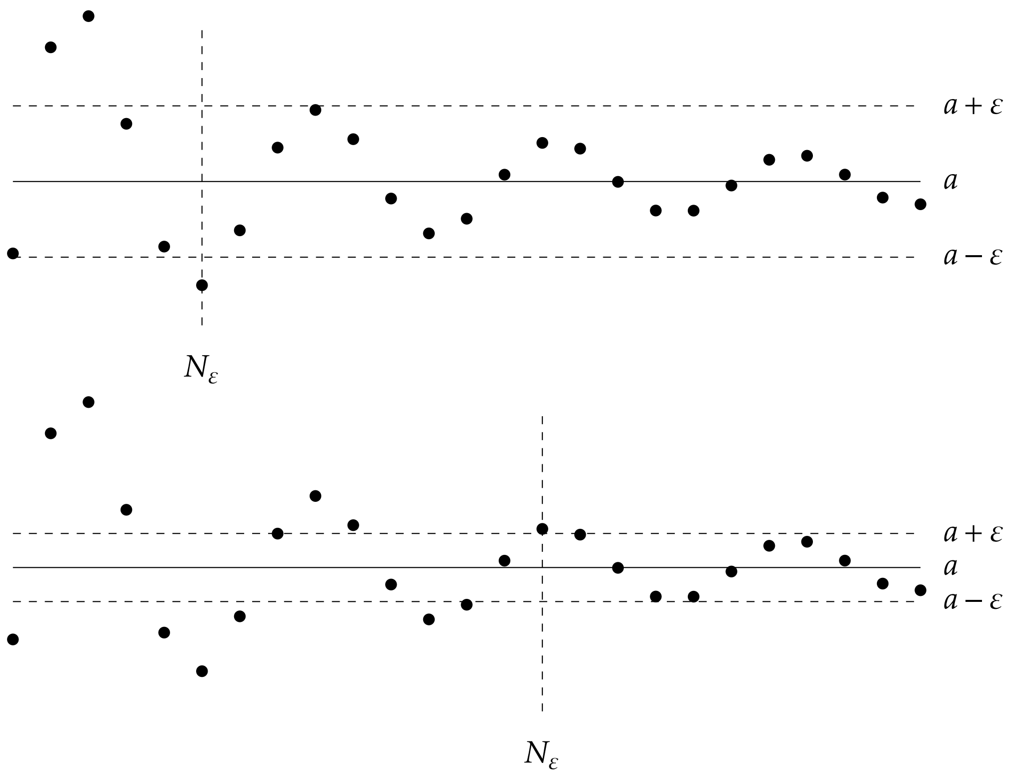


Figure 1: Two ε-N_ε relationships (Definition 5.4) for the same sequence. Terms after N_ε lie within ε of a, i.e. between $a-\varepsilon$ and $a+\varepsilon$. When ε is made smaller, N_ε becomes larger.

5.4 Definition. Consider a sequence (a_n) and a real number a. We say that a_n *converges* to a as n tends to infinity, written $a_n \to a$ as $n \to \infty$, if for any $\varepsilon > 0$ there exists a natural number N_ε (depending on ε) such that if $n > N_\varepsilon$ then $|a_n - a| < \varepsilon$. We also write $\lim_{n\to\infty} a_n$ to denote the limit, when it exists.

The idea is that for any tolerance (ε), if we go far enough along the sequence ($n > N_\varepsilon$) then a_n and a will agree to the specified tolerance ($|a_n - a| < \varepsilon$). See Figure 1 for an illustration.

Roughly speaking, if we put $\varepsilon = 10^{-k}$ then we are asking a_n and a to agree to k decimal places.

5.5 Example. Suppose $c \in \mathbb{R}$, $c \neq 0$. We would expect c/n to tend to zero as $n \to \infty$. How do we make this fit with the above definition? We need to lead up to the line

$$\left|\frac{c}{n} - 0\right| < \varepsilon$$

which is equivalent to $|c|/n < \varepsilon$, or $n > |c|/\varepsilon$. This is almost the definition of convergence to zero, with $|c|/\varepsilon$ playing the role of N_ε: if $n > |c|/\varepsilon$ then $|c/n - 0| < \varepsilon$. The only (minor) difference is that the N_ε in the definition is supposed to be a natural number, whereas $|c|/\varepsilon$ is not typically a natural number. The way out is to choose N_ε to be a natural number larger than $|c|/\varepsilon$. Now we can check it works:

Given $\varepsilon > 0$, let N_ε be some natural number such that $N_\varepsilon > |c|/\varepsilon$. If $n > N_\varepsilon$ then $n > |c|/\varepsilon$, and

$$\left|\frac{c}{n} - 0\right| = \frac{|c|}{n} < \frac{|c|\varepsilon}{|c|} = \varepsilon$$

This shows that $c/n \to 0$ as $n \to \infty$.

Note the use of the Archimedes property, which guarantees that there is a natural number larger than $|c|/\varepsilon$.

5.6 Example. A sequence $(a_n)_{n\in\mathbb{N}}$ defined by a simple formula can often be handled by the following technique (we shall see better methods later).

First, guess the limit. Then write down the inequality $|a_n - a| < \varepsilon$. Finally, manipulate it (taking care of the way the implications flow; "if and only if" is good) until it looks like $n >$(some function of ε), and read off N_ε.

For example, let

$$a_n = \frac{2n^2 + 1}{n^2 - 2}$$

We make the guess that this converges to 2, because when n is large the constants are much smaller than the n^2 terms. We then write down

$$\begin{aligned}
\left|\frac{2n^2+1}{n^2-2} - 2\right| < \varepsilon &\iff \left|\frac{5}{n^2-2}\right| < \varepsilon \\
&\iff \frac{5}{n^2-2} < \varepsilon \text{ (provided } n > 1) \\
&\iff n^2 > 2 + 5/\varepsilon \\
&\iff n > \sqrt{2 + 5/\varepsilon}
\end{aligned}$$

This gives us a suitable N_ε: any integer larger than $\sqrt{2+5/\varepsilon}$. Note that the "provided $n>1$" restriction above is harmless: if $n > N_\varepsilon$ then $n > 1$.

Now, we write the solution down formally:

Given $\varepsilon > 0$, choose $N_\varepsilon \in \mathbb{N}$ such that $N_\varepsilon \geq \sqrt{2+5/\varepsilon}$. If $n > N_\varepsilon$ then $n > \sqrt{2+5/\varepsilon}$ and hence, by the above calculation (note that each step is "if and only if"), $|a_n - 2| < \varepsilon$. This shows that $a_n \to 2$ as $n \to \infty$.

If our inequality had been true only for $n > M$ instead of $n > 1$, we could have chosen $N_\varepsilon \geq \max\{M, \sqrt{2+5/\varepsilon}\}$.

5.7 Example. Sometimes, it is better to estimate. Consider the example

$$a_n = \frac{1}{n+\sqrt{n}}$$

Here we guess that the limit is 0. We could follow the above technique and consider the inequality

$$\left|\frac{1}{n+\sqrt{n}} - 0\right| < \varepsilon$$

or equivalently

$$\frac{1}{n+\sqrt{n}} < \varepsilon$$

but this would be rather messy. It's easier to observe that

$$\frac{1}{n+\sqrt{n}} < \frac{1}{n}$$

and $1/n < \varepsilon$ if $n > 1/\varepsilon$.

The solution is written down formally as:

Given $\varepsilon > 0$, choose $N_\varepsilon \in \mathbb{N}$ such that $N_\varepsilon > 1/\varepsilon$. If $n > N_\varepsilon$ then $n > 1/\varepsilon$, so

$$|a_n - 0| = \left|\frac{1}{n+\sqrt{n}} - 0\right| < \frac{1}{n} < \varepsilon$$

which shows that $a_n \to 0$ as $n \to \infty$.

5.8 Example. We should also know how to prove that sequences do not converge. Let

$$a_n = (-1)^n = \begin{cases} -1 & \text{if } n \text{ is odd} \\ 1 & \text{if } n \text{ is even} \end{cases}$$

and suppose for a contradiction that $a_n \to a$ as $n \to \infty$. Then we can put $\varepsilon = 1$ into the definition of convergence to see that there exists $N \in \mathbb{N}$ such that if $n > N$ then

$$|(-1)^n - a| < 1$$

If n is even, this translates into $0 < a < 2$; if n is odd, it translates into $-2 < a < 0$. These cannot both be true; contradiction.

5.9 Definition. If $(a_j)_{j\in\mathbb{N}}$ is a real sequence, we define its nth *partial sum* to be the finite sum

$$s_n = a_1 + a_2 + \cdots + a_n = \sum_{j=1}^{n} a_j$$

The series $\sum a_j$ is said to converge to a if the sequence of partial sums converges to a limit, which we call the *infinite sum*. In symbols:

$$\sum_{j=1}^{\infty} a_j = \lim_{n\to\infty} \sum_{j=1}^{n} a_j$$

if this limit exists. Note the potentially confusing terminology: we also say that $\sum_{j=1}^{\infty} a_j$ converges.

Examples like this are called *series*, *infinite series* or *infinite sums*.

Series with different starting points are handled in the obvious way, e.g., for $(b_j)_{j\geq 0}$ we define

$$\sum_{j=0}^{\infty} b_j = \lim_{n\to\infty} \sum_{j=0}^{n} b_j$$

if the limit exists.

5.10 Example. (cf. Exercise 1.4). Consider the partial fraction decomposition

$$\frac{1}{j(j+1)} = \frac{1}{j} - \frac{1}{j+1}$$

We can write a partial sum

$$\sum_{j=1}^{n} \frac{1}{j(j+1)} = 1 - \frac{1}{2} + \frac{1}{2} - \frac{1}{3} + \cdots + \frac{1}{n} - \frac{1}{n+1}$$

All the terms except the initial 1 and the final $1/(n+1)$ cancel in pairs, leaving us with

$$s_n := \sum_{j=1}^{n} \frac{1}{j(j+1)} = 1 - \frac{1}{n+1}$$

We make the initial guess that the limit of this sequence as $n \to \infty$ is 1. To prove this, we have to consider the inequality $|s_n - 1| < \varepsilon$, which simplifies to

$$\frac{1}{n+1} < \varepsilon$$

which is in turn equivalent to

$$n > \frac{1}{\varepsilon} - 1 \tag{*}$$

Now, we can write down the argument: for any $\varepsilon > 0$, choose $N_\varepsilon \in \mathbb{N}$ such that $N_\varepsilon > 1/\varepsilon - 1$ (possible by Archimedes). Then, if $n > N_\varepsilon$, (*) is true and hence $|s_n - 1| < \varepsilon$. This shows that $s_n \to 1$ as $n \to \infty$. In summation notation,

$$\sum_{j=1}^{\infty} \frac{1}{j(j+1)} = 1$$

5.11 Theorem (Basic properties of convergent sequences)**.** *Suppose $(a_n)_{n\in\mathbb{N}}$ is a sequence and $a_n \to a$ as $n \to \infty$. Then:*

1. *$(a_n)_{n\in\mathbb{N}}$ is bounded; that is, there exists constants A and B such that $A \le a_n \le B$ for all $n \in \mathbb{N}$.*

2. *Changing finitely many terms does not affect the limit: if $(b_n)_{n\in\mathbb{N}}$ is another sequence such that $b_n = a_n$ except for finitely many n, then $b_n \to a$ as $n \to \infty$.*

3. *Starting at a later point (shifting) does not change the limit: if $k \in \mathbb{N}$ then $a_{n+k} \to a$ as $n \to \infty$.*

4. *Limits are unique: if $a_n \to b$ as $n \to \infty$ then $b = a$.*

Proof.

1. Put $\varepsilon = 1$ in the definition. There exists $N \in \mathbb{N}$ such that if $n > N$ then $|a_n - a| < 1$, or equivalently $a - 1 < a_n < a + 1$. There are only finitely many other terms, so we can let

$$A = \min\{a_1, a_2, \ldots, a_N, a-1\}; \qquad B = \max\{a_1, a_2, \ldots, a_N, a+1\}$$

   and it follows that $A \le a_n \le B$ for all $n \in \mathbb{N}$.

2. If the two sequences differ only in finitely many places, we can choose $N_0 \in \mathbb{N}$ such that $a_n = b_n$ for all $n > N_0$. Now, for any $\varepsilon > 0$ there exists $N_\varepsilon \in \mathbb{N}$ such that if $n > N$ then $|a_n - a| < \varepsilon$. Let $N'_\varepsilon = \max\{N_\varepsilon, N_0\}$. If $n > N'_\varepsilon$ then $n > N_0$, so $a_n = b_n$, and $n > N_\varepsilon$, so $|a_n - a| < \varepsilon$. It follows that $|b_n - a| < \varepsilon$, so $b_n \to a$.

3. Given $\varepsilon > 0$ there exists $N_\varepsilon \in \mathbb{N}$ such that if $n > N_\varepsilon$ then $|a_n - a| < \varepsilon$. If $n > N_\varepsilon$ then certainly $n + k > N_\varepsilon$, so $|a_{n+k} - a| < \varepsilon$. This shows that $a_{n+k} \to a$ as $n \to \infty$.

4. If $a_n \to a$ and $a_n \to b$ then for any $\varepsilon > 0$ there exist $N_\varepsilon^{(a)}$ such that if $n > N_\varepsilon^{(a)}$ then $|a_n - a| < \varepsilon$ and $N_\varepsilon^{(b)}$ such that if $n > N_\varepsilon^{(b)}$ then $|a_n - b| < \varepsilon$. If $n > \max\{N_\varepsilon^{(a)}, N_\varepsilon^{(b)}\}$ then we have both $|a_n - a| < \varepsilon$ and $|a_n - b| < \varepsilon$. We can write these as

$$-\varepsilon < a_n - a < \varepsilon; \qquad -\varepsilon < b - a_n < \varepsilon$$

   Add these together to give

$$-2\varepsilon < b - a < 2\varepsilon$$

   This is true for all $\varepsilon > 0$, so $a = b$ (Lemma 3.5).

□

5.12 Theorem (Limits and inequalities)**.** *Suppose $(a_n)_{n\in\mathbb{N}}$ is a real sequence such that $a_n \to a$ as $n \to \infty$ and that $b \in \mathbb{R}$. Then:*

1. *If there exists $N \in \mathbb{N}$ such that $a_n \ge b$ for all $n > N$ then $a \ge b$; if $a_n \le b$ for all $n > N$ then $a \le b$.*

2. *If $a < b$ then there exists $N \in \mathbb{N}$ such that $a_n < b$ for all $n > N$; if $a > b$ then there exists $N \in \mathbb{N}$ such that $a_n > b$ for all $n > N$.*

Proof. In each case, we prove the first statement only; the proof of the second statement is a minor modification.

1. For any $\varepsilon > 0$, there exists $N_\varepsilon \in \mathbb{N}$ such that if $n > N_\varepsilon$ then $|a_n - a| < \varepsilon$, or equivalently, $a - \varepsilon < a_n < a + \varepsilon$. For $n > \max(N, N_\varepsilon)$ we also have $a_n \geq b$; it follows that $b < a+\varepsilon$ and, this being true for all $\varepsilon > 0$, we can conclude (Lemma 3.5) that $b \leq a$.

2. Put $\varepsilon = b-a > 0$ in the definition of convergence of a_n to a to find $N \in \mathbb{N}$ such that if $n > N$ then $|a_n - a| < b - a$, which implies that $a_n - a < b - a$ and hence $a_n < b$.

□

5.13 Remark.

1. If $a_n > b$ for all n, it follows from the first part that $a \geq b$; however, it need not be that case that $a > b$. For example, $1/n > 0$ for all n but $1/n \to 0$ as $n \to \infty$. As a general rule, the limit of a strict inequality is a weak inequality.

2. In the second part, if $a > 0$ then $a > a/2$ so we have, for some $N \in \mathbb{N}$, $a_n > a/2 > 0$ for $n > N$; similarly, if $a < 0$ then, for some $N \in \mathbb{N}$, $a_n < a/2 < 0$ for $n > N$. Combining these, if $a \neq 0$ then $|a_n| > |a|/2$ for $n > N$. If, in addition, $a_n \neq 0$ for $n \in \mathbb{N}$ then we have $|a_n| \geq C > 0$ for all $n \in \mathbb{N}$ where

$$C = \min\{|a_1|, |a_2|, \ldots, |a_N|, |a|/2\}$$

This is referred to as being *bounded away from* 0: a non-zero sequence with a non-zero limit is bounded away from zero by the above quantity $C > 0$.

We can also assign a meaning to a sequence tending to $\pm\infty$. Here the idea is that, instead of the terms getting closer and closer to some value, they get larger and larger.

5.14 Definition. Suppose $(a_n)_{n\in\mathbb{N}}$ is a real sequence. We say that $a_n \to \infty$ as $n \to \infty$ if for any $K \in \mathbb{R}$ there exists $N_K \in \mathbb{N}$ such that if $n > N_K$ then $a_n > K$.

Similarly, we say that $a_n \to -\infty$ as $n \to \infty$ if for any $K \in \mathbb{R}$ there exists $N_K \in \mathbb{N}$ such that if $n > N_K$ then $a_n < K$.

It is sometimes simpler to consider only $K > 0$, or more generally $K > K_0$ for some stated constant K_0, in the definition of $a_n \to \infty$ — this is exactly equivalent. Similarly, if convenient, we may consider only $K < K_0$ in the definition of $a_n \to -\infty$.

It is easy to check that $a_n \to \infty$ if and only if $-a_n \to -\infty$.

5.15 Theorem. *Suppose $(a_n)_{n\in\mathbb{N}}$ is a sequence of non-zero numbers. Then the following are equivalent:*

1. $|a_n| \to \infty$ *as* $n \to \infty$

2. $1/a_n \to 0$ *as* $n \to \infty$.

Proof. The basic calculations are that

$$\left|\frac{1}{a_n} - 0\right| < \varepsilon \iff |a_n| > \frac{1}{\varepsilon}$$

and

$$|a_n| > K \iff \left|\frac{1}{a_n} - 0\right| < \frac{1}{K}$$

Suppose $|a_n| \to \infty$ and $\varepsilon > 0$. Put $K = 1/\varepsilon$ in the definition to see that there exists $N_\varepsilon \in \mathbb{N}$ such that if $n > N_\varepsilon$ then $|a_n| > 1/\varepsilon$. This leads to $|1/a_n| < \varepsilon$, so $1/a_n \to 0$.

Now suppose $1/a_n \to 0$ and $K > 0$. Put $\varepsilon = 1/K$ in the definition to see that there exists $N_K \in \mathbb{N}$ such that if $n > N_K$ then $|1/a_n| < 1/K$. This leads to $|a_n| > K$, so $|a_n| \to \infty$. Note that the case $K \leq 0$ can be ignored, as mentioned in the definition; or we can let N_K be anything at all, because $|a_n| > 0 > K$. □

5.16 Example. Suppose $q \in \mathbb{Q}$, $q > 0$, and consider the sequences

$$a_n = \frac{1}{n^q}; \qquad b_n = n^q$$

We can see directly that $a_n \to 0$ since, given $\varepsilon > 0$, we can let N_ε be a natural number greater than $1/\varepsilon^{1/q}$. If $n > N_\varepsilon$ then $n > 1/\varepsilon^{1/q}$ and hence $a_n = 1/n^q < \varepsilon$; since $a_n > 0$, $|a_n - 0| < \varepsilon$.

We can also see than $b_n \to \infty$ since, given $K > 0$, we can let N_K be a natural number greater than $K^{1/q}$. If $n > N_K$ then $n > K^{1/q}$ so $b_n = n^q > K$.

Alternatively, since $a_n = 1/b_n$, we could prove one of these facts and deduce the other using the previous theorem.

6 Algebra of Limits

The ε definition is powerful, but cumbersome. In this section we consider easier ways to calculate limits of sequences, by breaking them down into simpler parts. We need the following important result:

6.1 Theorem. *Suppose $x, y \in \mathbb{R}$. Then*

$$|x + y| \leq |x| + |y| \qquad \textit{(triangle inequality)}$$

$$\big||x| - |y|\big| \leq |x - y| \qquad \textit{(reverse triangle inequality)}$$

Proof. We have

$$(x + y)^2 = x^2 + 2xy + y^2 \leq |x|^2 + 2|x||y| + |y|^2 = (|x| + |y|)^2$$

If we take the non-negative square root of each side we see that

$$|x + y| \leq |x| + |y|$$

Similarly, we have

$$(x - y)^2 = x^2 - 2xy + y^2 \geq |x|^2 - 2|x||y| + |y|^2 = (|x| - |y|)^2$$

and taking non-negative square roots gives

$$|x-y| \geq \big||x|-|y|\big|$$

□

6.2 Lemma. *Suppose (a_n), (b_n) and (c_n) are real sequences such that, as $n \to \infty$, $a_n \to a$ and $b_n \to b$. If $c \in \mathbb{R}$ and there are constants $A, B \geq 0$ such that for all $n \in \mathbb{N}$*

$$|c_n - c| \leq A|a_n - a| + B|b_n - b|. \tag{*}$$

then $c_n \to c$ as $n \to \infty$.

Proof. Consider first the case $A > 0$ and $B > 0$. For any $\varepsilon > 0$, we choose $N_\varepsilon^{(a)}$ and $N_\varepsilon^{(b)}$ such that if $n > N_\varepsilon^{(a)}$ then $|a_n - a| < \varepsilon/(2A)$ and if $n > N_\varepsilon^{(b)}$ then $|b_n - b| < \varepsilon/(2B)$. Finally, it follows from (*) that if $n > N_\varepsilon := \max\{N_\varepsilon^{(a)}, N_\varepsilon^{(b)}\}$ then

$$|c_n - c| < A(\varepsilon/(2A)) + B(\varepsilon/(2B)) = \varepsilon.$$

The special cases where $A = 0$ or $B = 0$ are similar, but simpler (or, one can use the fact that if the inequality (*) holds for some particular values of A and B, then it also holds for any larger values of A and B). □

6.3 Theorem (Algebra of Limits / Combination Rules)**.** *Suppose $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ are sequences such that $a_n \to a$ as $n \to \infty$ and $b_n \to b$ as $n \to \infty$. Then, as $n \to \infty$,*

1. $a_n + b_n \to a + b$
2. $a_n b_n \to ab$
3. $a_n/b_n \to a/b$ *provided $b \neq 0$ and $b_n \neq 0$ for all n.*

Note that if $b \neq 0$ then there exists $N \in \mathbb{N}$ such that $b_n \neq 0$ for $n > N$; we can then apply (3) to see that $(a_n/b_n)_{n\geq N}$ has limit a/b.

Proof. In all cases we use the previous lemma. To do this, we need to find constants such that (*) is true.

1. Here the inequality is

$$|(a_n + b_n) - (a + b)| = |(a_n - a) + (b_n - b)| \leq |a_n - a| + |b_n - b|$$

2. Here the inequality is

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

where $C > 0$ is some constant such that $|a_n| \leq C$ for all $n \in \mathbb{N}$. (See Theorem 5.11(1)).

3. Here the first step towards the inequality is

$$\left|\frac{a_n}{b_n} - \frac{a}{b}\right| = \left|\frac{a_n b - a b_n}{b b_n}\right| = \left|\frac{(a_n - a)b - a(b_n - b)}{b_n b}\right| \le \frac{1}{|b_n|}|a_n - a| + \frac{|a|}{|b_n b|}|b_n - b|$$

To make this look like (*), we need a constant $C > 0$ such that $|1/b_n| \le C$ for all n. Such a constant exists by Remark 5.13(2): since $b_n \to b \neq 0$ and $b_n \neq 0$, we have $|b_n| \ge K > 0$ for some K and all n, so $|1/b_n| \le 1/K$ and we have

$$\left|\frac{a_n}{b_n} - \frac{a}{b}\right| \le \frac{1}{K}|a_n - a| + \frac{|a|}{K|b|}|b_n - b|$$

□

6.4 Remark. Two more rules follows immediately: $a_n - b_n = a_n + (-1) \times b_n$, so $a_n - b_n \to a - b$.

Also if $k \in \mathbb{Z}$, $k > 0$, then it follows by writing a_n^k as a product of k copies of a_n that $a_n^k \to a^k$. The same is true if $k < 0$, by writing $a_n^k = 1/a_n^{-k}$, provided $a_n \neq 0$ and $a \neq 0$.

This suddenly makes many algebraic examples much easier.

6.5 Example. Let

$$a_n = \frac{3n^2 - 2n + 1}{2n^2 + 7n - 2}$$

Divide through by the highest power of n present (the *dominant term*) to give

$$a_n = \frac{3 - 2/n + 1/n^2}{2 + 7/n - 2/n^2}$$

Everything except the constant tends to zero, so we can write down the answer:

$$a_n \to \frac{3 - 0 + 0}{2 + 0 - 0} = \frac{3}{2}$$

Every rational function (ratio of polynomials) can be handled in the same way. Let

$$b_n = \frac{n^2 + n}{3n^3 - 2n + 1}$$

Divide through by the highest power of n present, to give

$$b_n = \frac{1/n + 1/n^2}{3 - 2/n^2 + 1/n^3} \to \frac{0 + 0}{3 - 0 + 0} = 0$$

This technique leads to some useful general facts.

6.6 Theorem. *Suppose we have a sequence*

$$a_n = \frac{\alpha_1 n^{p_1} + \cdots + \alpha_j n^{p_j}}{\beta_1 n^{q_1} + \cdots + \beta_k n^{q_k}}$$

where all α, β coefficients are non-zero and $\max\{p_1, \ldots, p_j\} < \max\{q_1, \ldots, q_k\}$. *Then $a_n \to 0$ as $n \to \infty$.*

Proof. Dividing through by the highest power of n present, i.e. $\max\{q_1,\ldots,q_k\}$, gives only negative powers of n, all of which tend to zero, except for one or more constant terms in the denominator. The limit is therefore 0. □

Note that the powers of n do not have to be integers. If the largest power of n in the numerator is greater than that in the denominator, then the sequence tends to $\pm\infty$, depending on the signs of the coefficients of the largest power; easy proof omitted.

6.7 Theorem. *Suppose $\sum a_j$ is a convergent series. Then $a_j \to 0$ as $j \to \infty$. The converse is* not *true, as the harmonic series shows (Example 6.8).*

Proof. Let

$$s_n = \sum_{j=1}^{n} a_j$$

If the series converges, then by definition s_n converges to some finite limit, say a. Then s_{n-1} also converges to a (Theorem 5.11(3), Exercise 4.1), so (combination rules) $s_n - s_{n-1} \to 0$. But $s_n - s_{n-1} = a_n$, so $a_n \to 0$ as $n \to \infty$. □

6.8 Example (The Harmonic Series). The series

$$\sum_{j=1}^{\infty} \frac{1}{j} = 1 + \frac{1}{2} + \frac{1}{3} + \ldots$$

does not converge to a finite limit; in fact, even though the terms tend to zero, the partial sums tend to $+\infty$. To see this, we put the terms in blocks as follows:

$$1 + \frac{1}{2} + \underbrace{\frac{1}{3} + \frac{1}{4}}_{\text{2 terms}} + \underbrace{\frac{1}{5} + \cdots + \frac{1}{8}}_{\text{4 terms}} + \underbrace{\frac{1}{9} + \cdots + \frac{1}{16}}_{\text{8 terms}} + \ldots$$

Each block contains 2^k terms, of which the smallest is $1/2^{k+1}$. The sum of each block is therefore greater than $1/2$. The first two terms plus the first K blocks have a sum greater than $3/2 + K/2$; here K can be any natural number. This shows (Archimedes) that the partial sums are unbounded above; in particular, they cannot converge to a finite limit (Theorem 5.11(1)).

In fact, the partial sums tend to ∞: for any $C \in \mathbb{R}$, the blocking argument shows that there exists $N \in \mathbb{N}$ such that the Nth partial sum is larger than C; all subsequent partial sums are also larger than C.

6.9 Remark. Combination rules/algebra of limits for series are much more complicated than for sequences. One simple result is *linearity*: if we have two convergent series $\sum_{j=1}^{\infty} a_j$ and $\sum_{j=1}^{\infty} b_j$ and two real numbers α and β then

$$\sum_{j=1}^{\infty} (\alpha a_j + \beta b_j) = \alpha \sum_{j=1}^{\infty} a_j + \beta \sum_{j=1}^{\infty} b_j$$

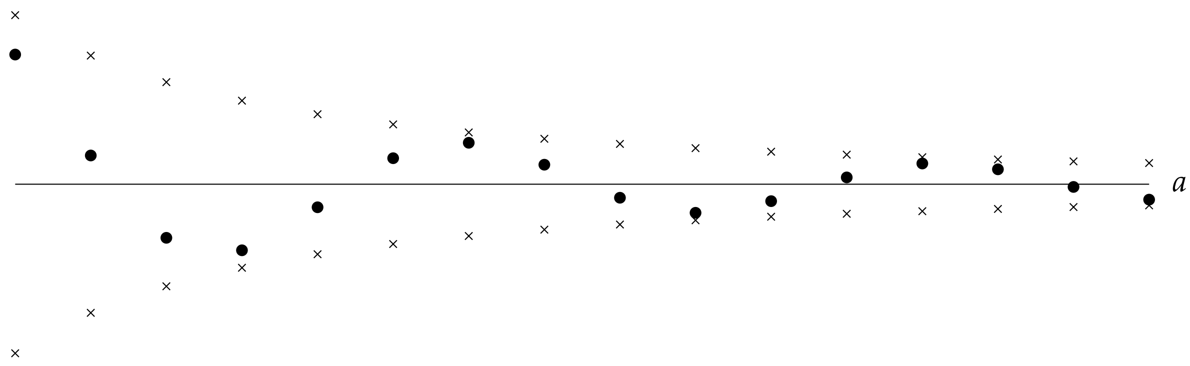


Figure 2: The sandwich theorem: the sequence drawn as solid dots is sandwiched between the two sequences drawn as crosses and forced to converge to their common limit, a.

We also have

$$\sum_{j=1}^{p} a_j + \sum_{j=p+1}^{\infty} a_j = \sum_{j=1}^{\infty} a_j$$

for any $p \in \mathbb{N}$, which is true if either one of the series is known to converge.

The proofs are easy exercises, based on the corresponding results for finite sums (Exercises 1).

There is a straightforward consequence worth mentioning: as $p \to \infty$, by definition,

$$\sum_{j=1}^{p} a_j \to \sum_{j=1}^{\infty} a_j$$

so by combination rules and the second identity above

$$\sum_{j=p+1}^{\infty} a_j \to 0$$

This sequence (of sums from $p+1$, or sometimes p, to infinity) is known as the *tail* of the series $\sum a_j$.

7 The Sandwich Theorem

7.1 Theorem (Sandwich Theorem). *Suppose we have three sequences $(a_n)_{n\in\mathbb{N}}$, $(b_n)_{n\in\mathbb{N}}$ and $(c_n)_{n\in\mathbb{N}}$ such that:*

1. *for all but finitely many values of n, $a_n \le b_n \le c_n$*
2. *as $n \to \infty$, $a_n \to a$ and $c_n \to a$*

Then $b_n \to a$ as $n \to \infty$. In particular (cf. Lemma 6.2), if $|x_n - x| \le y_n$ and $y_n \to 0$, then $x - y_n \le x_n \le x + y_n$ and $x \pm y_n \to x$, so $x_n \to x$.

Proof. Finitely many points do not affect the value of a limit (Theorem 5.11(2)), so we may proceed as if $a_n \le b_n \le c_n$ for all $n \in \mathbb{N}$.

Given $\varepsilon > 0$ there exist $N_\varepsilon^{(a)}, N_\varepsilon^{(c)} \in \mathbb{N}$ such that if $n > N_\varepsilon^{(a)}$ then $|a_n - a| < \varepsilon$ and if $n > N_\varepsilon^{(c)}$ then $|c_n - a| < \varepsilon$. If $n > N_\varepsilon := \max\{N_\varepsilon^{(a)}, N_\varepsilon^{(c)}\}$ then both of these inequalities are true. Write them in the form

$$-\varepsilon < a_n - a < \varepsilon; \qquad -\varepsilon < c_n - a < \varepsilon$$

and rearrange to give

$$a - \varepsilon < a_n < a + \varepsilon; \qquad a - \varepsilon < c_n < a + \varepsilon.$$

Now use the hypothesis $a_n \le b_n \le c_n$ to deduce that

$$a - \varepsilon < a_n \le b_n \le c_n < a + \varepsilon$$

so in particular

$$a - \varepsilon < b_n < a + \varepsilon$$

which is equivalent to $|b_n - a| < \varepsilon$. This shows that $b_n \to a$, as required. □

7.2 Example. Anticipating the definition of the sine function: consider the sequence

$$a_n = 1 + \frac{\sin(n)}{n}$$

All we actually need to know about sines is that $|\sin(n)| \le 1$ for all $n \in \mathbb{N}$, to show that this tends to 1 by sandwiching. The inequalities are

$$1 - \frac{1}{n} \le 1 + \frac{\sin(n)}{n} \le 1 + \frac{1}{n}$$

Since the left and right sequences both tend to 1, the sandwiched middle sequence also tends to 1.

7.3 Theorem. *Suppose $a_n \ge 0$ and $a_n \to a$ as $n \to \infty$. Then $\sqrt{a_n} \to \sqrt{a}$ as $n \to \infty$.*

Proof. First take the case $a > 0$ and write

$$|\sqrt{a_n} - \sqrt{a}| = \left| \frac{(\sqrt{a_n} - \sqrt{a})(\sqrt{a_n} + \sqrt{a})}{\sqrt{a_n} + \sqrt{a}} \right| = \frac{|a_n - a|}{\sqrt{a_n} + \sqrt{a}} \le \frac{|a_n - a|}{\sqrt{a}} \to 0$$

It now follows from the Sandwich Theorem that $\sqrt{a_n} \to \sqrt{a}$ as $n \to \infty$.

The case $a = 0$ is different (and easier): see Exercise 5.1.

Similar, but messier, arguments show that $a_n^p \to a^p$ for any $p \in \mathbb{Q}$ (or, in the case that $a = 0$, $p \in \mathbb{Q}$ with $p > 0$). □

There are also infinite analogues of the sandwich theorem. The proofs are left as an exercise for the reader.

7.4 Theorem. *If $a_n \to \infty$ as $n \to \infty$ and there exists $N \in \mathbb{N}$ such that $b_n \ge a_n$ for all $n > N$, then $b_n \to \infty$ as $n \to \infty$.*

Similarly, if $a_n \to -\infty$ as $n \to \infty$ and there exists $N \in \mathbb{N}$ such that $b_n \le a_n$ for all $n > N$, then $b_n \to -\infty$ as $n \to \infty$.

8 Standard Limits

In computing limits, the combination rules and the sandwich theorem are half the story. The other half is a library of standard limits. More complicated examples can then be broken down into pieces, whose limits are found in the library, and reassembled using combination rules. Many of the results can be thought of as part of a hierarchy of limits.

8.1 Definition. Suppose we have two sequences (a_n) and (b_n), both tending to $+\infty$ as $n \to \infty$. We say that a_n tends to infinity *faster* than b_n if $a_n/b_n \to \infty$ (equivalently, $b_n/a_n \to 0$) as $n \to \infty$. For example, n^4 and n^2 both tend to ∞, and $n^4/n^2 = n^2$ also tends to ∞, so n^4 tends to infinity faster than n^2.

If (c_n) and (d_n) are two sequences tending to zero, we say that c_n tends to 0 *faster* than d_n if $c_n/d_n \to 0$. For example, $1/n^2$ and $1/n$ both tend to zero, and $(1/n^2)/(1/n) = 1/n \to 0$, so $1/n^2$ tends to zero faster than $1/n$.

8.2 Theorem (Hierarchy of limits). *If (a_n) occurs higher up in the following table than (b_n) then, unless they are both bounded, $b_n/a_n \to 0$ as $n \to \infty$.*

$$
\begin{array}{ll}
\textit{tend to } \infty & \left\{\begin{array}{ll} n^n & \textit{FAST} \\ n! & \\ x^n\ (x > 1) & \\ n^q\ (q \in \mathbb{Q},\ q > 0) & \textit{SLOW} \end{array}\right. \\
\textit{remain bounded} & \left\{\begin{array}{ll} \textit{constants} & \\ (-1)^n \textit{ diverges} & \\ x^{1/n} \to 1\ (x > 0) & \\ n^{q/n} \to 1\ (q \in \mathbb{Q}) & \end{array}\right. \\
\textit{tend to } 0 & \left\{\begin{array}{ll} n^{-q}\ (q \in \mathbb{Q},\ q > 0) & \textit{SLOW} \\ x^n\ (|x| < 1) & \\ 1/n! & \\ n^{-n} & \textit{FAST} \end{array}\right.
\end{array}
$$

There are also sub-hierarchies, such as $n^p/n^q \to 0$ if $p < q$ and $x^n/y^n \to 0$ if $|x| < |y|$, but these would clutter up the table and can generally be deduced from the results already there.

Although this covers most of the common examples, such a table is always incomplete (e.g., how does $(n!)^2$ compare to n^n?).

The results restricted to rational powers also work for irrational powers; these are omitted because we haven't defined irrational powers yet!

Most of these results will be proved later in this section.

We can use the table to calculate the limits of many concrete examples.

8.3 Example.

$$\frac{n^{1/n} - 2^n}{n + 2^n} = \frac{n^{1/n}/2^n - 1}{n/2^n + 1} \to \frac{0 - 1}{0 + 1} = -1$$

as $n \to \infty$ (dominant term is 2^n)

$$\frac{n! + 2n^n + 1}{n^n + 2^n} = \frac{n!/n^n + 2 + 1/n^n}{1 + 2^n/n^n} \to \frac{0 + 2 + 0}{1 + 0} = 2$$

as $n \to \infty$ (dominant term is n^n).

The remainder of this section is devoted to proving the relationships in the table in Theorem 8.2. These are all examples of the use of the Sandwich Theorem and the Algebra of Limits.

8.4 Example. n^n tends to ∞ faster than $n!$. Precisely,

$$\frac{n!}{n^n} \to 0 \text{ as } n \to \infty.$$

This is because

$$0 < \frac{n!}{n^n} = \frac{1}{n}.\frac{2}{n}\ldots\frac{n}{n} \leq \frac{1}{n}$$

so $n!/n^n \to 0$ by the Sandwich Theorem.

8.5 Example. The next result from the Hierarchy of Limits is that if $x > 1$ then $n!$ tends to ∞ faster than x^n. The fact that $x^n \to \infty$ as $n \to \infty$ if $x > 1$ will actually be established in Example 8.6; here, we begin by establishing that for any $x > 0$,

$$\frac{x^n}{n!} \to 0 \text{ as } n \to \infty. \tag{*}$$

This is because

$$\frac{x^n}{n!} = \frac{x}{1}\frac{x}{2}\ldots\frac{x}{n}$$

Suppose $k \in \mathbb{N}$ with $k > x$. Then for $n > k$ we have

$$\frac{x^n}{n!} = \underbrace{\frac{x}{1}\frac{x}{2}\ldots\frac{x}{k-1}}_{=x^{k-1}/(k-1)!}\underbrace{\frac{x}{k}\ldots\frac{x}{n-1}}_{\leq 1}\frac{x}{n}$$

All of the terms from x/k to $x/n-1$ are less than or equal to 1. This gives us

$$0 < \frac{x^n}{n!} \leq \frac{x^{k-1}}{(k-1)!}\frac{x}{n} = \frac{x^k}{(k-1)!n} \to 0$$

By the Sandwich Theorem, $x^n/n! \to 0$ as $n \to \infty$.

In fact, (*) is true for any $x \in \mathbb{R}$. This is because

$$\underbrace{-\frac{|x|^n}{n!}}_{\to 0} \leq \frac{x^n}{n!} \leq \underbrace{\frac{|x|^n}{n!}}_{\to 0}$$

so $x^n/n! \to 0$ by the Sandwich Theorem.

8.6 Example. If $x > 1$ and $q \in \mathbb{Q}$, $q > 0$, then x^n tends to ∞ faster than n^q; precisely,

$$\frac{n^q}{x^n} \to 0 \text{ as } n \to \infty$$

(in fact, the proof will show that this is true for all $q \in \mathbb{Q}$, regardless of sign). This is a little more complicated.

Let $h = x - 1$ so $x = 1 + h$ and $h > 0$. Now expand x^n using the binomial theorem:

$$x^n = (1+h)^n = \sum_{k=0}^{n} \binom{n}{k} h^k$$

Choose some $k \in \mathbb{N}$ with $k > q$. If $n > k$ then we have

$$x^n > \binom{n}{k} h^k$$

so

$$0 < \frac{n^q}{x^n} < \frac{n^q}{\binom{n}{k} h^k}$$

The denominator is a polynomial of degree k and $k > q$, so the RHS tends to zero (Example 6.5). By the Sandwich Theorem, $n^q/x^n \to 0$ as $n \to \infty$.

In particular, if we put $q = 0$ then we see that $x^n \to \infty$ if $x > 1$. If $0 < x < 1$ then $1/x > 1$, so $(1/x)^n \to \infty$. But $(1/x)^n = 1/x^n$, so $1/x^n \to \infty$ and hence $x^n \to 0$ (Theorem 5.15). Next, if $-1 < x < 0$, we have

$$-\underbrace{|x|^n}_{\to 0} \le x^n \le \underbrace{|x|^n}_{\to 0}$$

so $x^n \to 0$ by the Sandwich Theorem. The cases $x = 0$ and $x = 1$ are trivial, so we can summarise:

$$x^n \to \begin{cases} \infty & \text{if } x > 1 \\ 1 & \text{if } x = 1 \\ 0 & \text{if } -1 < x < 1 \end{cases} \qquad \text{as } n \to \infty$$

8.7 Example (The Geometric Series). This is one of the most important examples of an infinite series. If $x \in \mathbb{R}$, $x \neq 1$, then we know that

$$\sum_{j=0}^{n} x^j = \frac{x^{n+1} - 1}{x - 1}$$

If (and only if) $|x| < 1$, this converges to a finite limit, because in this case $x^{n+1} \to 0$ as $n \to \infty$. (Example 8.6). We thus have

$$\sum_{j=0}^{\infty} x^j = \lim_{n\to\infty} \sum_{j=0}^{n} x^j = \lim_{n\to\infty} \frac{x^{n+1} - 1}{x - 1} = \frac{1}{1 - x}$$

provided $|x| < 1$.

8.8 Example. $n^{1/n} \to 1$ as $n \to \infty$. To see this, consider the inequality

$$|n^{1/n} - 1| < \varepsilon$$

Since $n \ge 1$, $n^{1/n} \ge 1$, so this is equivalent to $n^{1/n} - 1 < \varepsilon$ which is equivalent to $n < (1 + \varepsilon)^n$; provided $n \ge 2$, this in turn is equivalent to

$$n < 1 + n\varepsilon + \frac{1}{2} n(n-1)\varepsilon^2 + \cdots + \varepsilon^n$$

Now, we stop the "if and only if" reasoning and say instead that this is *implied by* (but does not imply)

$$n < \frac{1}{2}n(n-1)\varepsilon^2$$

This inequality can be solved; it is equivalent to $n > (2/\varepsilon^2) + 1$.

Formally, given $\varepsilon > 0$, choose $N_\varepsilon \in \mathbb{N}$ such that $N_\varepsilon \geq (2/\varepsilon^2) + 1$. If $n > N_\varepsilon$ then $n > (2/\varepsilon^2) + 1$ (and $n \geq 2$) and we can follow the above argument in reverse to give $|n^{1/n} - 1| < \varepsilon$, showing that $n^{1/n} \to 1$ as $n \to \infty$.

We can draw some extra conclusions from this. Firstly, $n^{q/n} = (n^{1/n})^q \to 1^q = 1$ for any $q \in \mathbb{Q}$ (Theorem 7.3).

Secondly, if $x \geq 1$ then, for $n > x$, $1 \leq x^{1/n} < n^{1/n} \to 1$. By the Sandwich Theorem, $x^{1/n} \to 1$.

Finally, if $0 < x < 1$ then

$$x^{1/n} = \frac{1}{(1/x)^{1/n}} \to \frac{1}{1} = 1$$

since $1/x > 1$, so in fact $x^{1/n} \to 1$ for any $x > 0$.

9 Part II: key points

The purpose of Part II is to develop the basic theory of convergent sequences and series. A *sequence* is an infinitely long list of real numbers, indexed by integers. Notation like $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\geq 0}$ is used to refer to sequences $(a_1, a_2, \dots)$, $(b_0, b_1, b_2, \dots)$, etc. (Definition 5.1).

A *series* is formed from a sequence by considering the *partial sums*: the nth partial sum is the sum of the first n terms of the sequence (Definition 5.9).

A sequence $(a_n)_{n\in\mathbb{N}}$ is said to:

- *converge* to a *limit* $a \in \mathbb{R}$ as $n \to \infty$ if for any $\varepsilon > 0$ there exists $N_\varepsilon \in \mathbb{N}$ such that if $n > N_\varepsilon$ then $|a_n - a| < \varepsilon$ (Definition 5.4);
- *diverge* if it does not converge to any limit $a \in \mathbb{R}$;
- *tend (or diverge) to* $+\infty$ *as* $n \to \infty$ if for any $K \in \mathbb{R}$ there exists $N_K \in \mathbb{N}$ such that if $n > N_K$ then $a_n > K$ (Definition 5.14);
- *tend (or diverge) to* $-\infty$ *as* $n \to \infty$ if for any $K \in \mathbb{R}$ there exists $N_K \in \mathbb{N}$ such that if $n > N_K$ then $a_n < K$ (*ibid.*).

These definitions are unambiguous, rigorous, testable, and mean the same thing to all mathematicians who use them. They are intended to capture the intuitive ideas we all have about convergence, but the definitions take precedence; in the case of a conflict between definition and intuition, the definition wins!

Some of the following properties probably seem "obvious" to you. We prove them because "this is obvious" is a statement about my or your intuitive idea of convergence, whereas the exact theorems are statements about the rigorous definition of convergence. You can, if you like, take the proof of an "obvious" fact as evidence that the rigorous definition is in tune with your intuition.

- The limit of a sequence, if it exists, is unique (Theorem 5.11).
- Changing finitely many terms of a sequence does not affect convergence (*ibid.*).
- Shifting a sequence (changing a_n to a_{n+k}) does not affect convergence (*ibid.*).
- Every convergent sequence is bounded (*ibid.*).
- If $a_n \to a \neq 0$ as $n \to \infty$ and $a_n \neq 0$ for all n then (a_n) is *bounded away from zero*; that is, there is a constant $C > 0$ such that $|a_n| \geq C$ for all n (Remark 5.13(2)).
- If $a_n \neq 0$ for all n then $|a_n| \to \infty$ as $n \to \infty$ if and only if $1/a_n \to 0$ as $n \to \infty$ (Theorem 5.15).
- If $q \in \mathbb{Q}$ and $q < 0$ then $n^q \to 0$ as $n \to \infty$; if $q > 0$ then $n^q \to \infty$ as $n \to \infty$ (Example 5.16).
- Limits interact well with the algebraic operations on the real numbers: if $a_n \to a$ and $b_n \to b$ as $n \to \infty$ then $a_n \pm b_n \to a \pm b$, $a_n b_n \to ab$ and $a_n/b_n \to a/b$ provided $b_n, b \neq 0$. These are known as the *combination rules* or the *algebra of limits* (Theorem 6.3 *et seq.*).

- Limits interact well with the order relation on the real numbers: if $a_n \le b$ for all n and $a_n \to a$ as $n \to \infty$ then $a \le b$; the same is true with "$\le$" replaced by "$\ge$", but not "$<$" or "$>$" (Theorem 5.12 *et seq.*). We also have the important *Sandwich Theorem*, Theorem 7.1: if $a_n \le b_n \le c_n$ for all n, $a_n \to a$ and $c_n \to a$ as $n \to \infty$, then $b_n \to a$ as $n \to \infty$. There are also analogues for divergence to $\pm\infty$: if $a_n \to \infty$ and $b_n \ge a_n$ then $b_n \to \infty$; if $a_n \to -\infty$ and $b_n \le a_n$ then $b_n \to -\infty$ (all as $n \to \infty$; Theorem 7.4).

- The Sandwich Theorem is the key to understanding many basic limits, which can be arranged as a hierarchy of speeds of divergence and convergence (Definition 8.1 *et seq.*). Together with the combination rules, this makes it quite straightforward to evaluate limits of many sequences defined by algebraic formulae (examples in Section 6).

The (infinite) sum of a series is defined as the limit of the sequence of partial sums, if this limit exists:

$$\sum_{j=1}^{\infty} a_j = \lim_{j\to\infty} \sum_{j=1}^{n} a_j$$

(Definition 5.9). If the series converges (i.e. if the sequence of partial sums converges), then we have $a_j \to 0$ as $j \to \infty$ (Theorem 6.7) but the converse is not true: the harmonic series

$$\sum_{j=1}^{n} \frac{1}{j} \to \infty \text{ as } n \to \infty$$

is the standard counterexample.

There are a few series whose sums we can evaluate explicitly, of which the most important is the *geometric series* or *geometric sum*:

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

provided $|x| < 1$ (Example 8.7).

Literature:

1. F. Mary Hart *Guide to Analysis* (2nd Edition, Palgrave Macmillan 2001, JBM classmark S7 Har) Chapters 2, 3.

2. Keith E. Hirst, *Numbers, Sequences and Series* (Edward Arnold, 1995, JBM classmark S2.81 Hir) Chapters 7, 8.

3. any other introduction to Real Analysis; browse classmark S7 in the library.

Part III
Convergence Tests

In all the previous examples of convergent sequences and series, we have given an explicit formula for the limit. In this part of the course, we consider techniques that prove that a sequence or series converges to a limit, without such a formula. This allows us to create new numbers as limits; for example, this is how we define the exponential and trigonometric functions.

10 Monotone Sequences

10.1 Definition. A sequence $(a_n)_{n\in\mathbb{N}}$ is called

$$\left.\begin{array}{lll} \text{increasing} & \text{if} & a_{n+1} \geq a_n \\ \text{strictly increasing} & \text{if} & a_{n+1} > a_n \\ \text{decreasing} & \text{if} & a_{n+1} \leq a_n \\ \text{strictly decreasing} & \text{if} & a_{n+1} < a_n \end{array}\right\} \text{for all } n$$

It is called *monotone* if it is satisfies any of the above. Note that some books use the terminology differently!

10.2 Theorem (Principle of Bounded Monotone Convergence). *If $(a_n)_{n\in\mathbb{N}}$ is an increasing sequence which is bounded above then, as $n \to \infty$, it converges to* $\sup\{a_n : n \in \mathbb{N}\}$.

If $(a_n)_{n\in\mathbb{N}}$ is a decreasing sequence which is bounded below then, as $n \to \infty$, it converges to $\inf\{a_n : n \in \mathbb{N}\}$.

Proof. Suppose first that $(a_n)_{n\in\mathbb{N}}$ is increasing and bounded above, and let

$$a = \sup\{a_n : n \in \mathbb{N}\}$$

Then for any $\varepsilon > 0$ there exists $N_\varepsilon \in \mathbb{N}$ such that $a_{N_\varepsilon} > a - \varepsilon$ (Remark 3.16(3)). If $n > N_\varepsilon$ then, because (a_n) is increasing, $a_n > a_{N_\varepsilon} > a - \varepsilon$. Also $a_n \leq a$ for any n. These two inequalities combine to give

$$-\varepsilon < a_n - a \leq 0$$

which implies that $|a_n - a| < \varepsilon$. Since this is true for all $n > N_\varepsilon$, $a_n \to a$ as $n \to \infty$.

The proof for decreasing sequences is similar, or we can apply the above result to $(-a_n)_{n\in\mathbb{N}}$. □

10.3 Example. We can use this to give an alternative derivation of a limit that we already know (cf. Example 8.6). Suppose $0 < x < 1$ and let $a_n = x^n$, so $(a_n)_{n\in\mathbb{N}}$ is a decreasing sequence, bounded below by 0. It follows that it converges to some limit a. But:

1. $a_{n+1} \to a$ as $n \to \infty$ (shift property, Theorem 5.11(3)) as $n \to \infty$ and

2. $a_{n+1} = xa_n \to xa$ as $n \to \infty$ (algebra of limits).

Now, we have two apparently different limits for the same sequence. But limits are unique (Theorem 5.11(4)), so they must in fact be the same: that is, $ax = a$, and hence $a = 0$. We conclude that $x^n \to 0$ as $n \to \infty$.

This argument required $x > 0$ to exploit monotonicity, but we can also note that the same conclusion is trivially true if $x = 0$ and that for $-1 < x < 0$ we have $|x^n - 0| = (-x)^n$ which tends to zero by the previous case; we thus have $x^n \to 0$ as $n \to \infty$ for any x with $|x| < 1$, exactly as in Example 8.6.

10.4 Example. Define a sequence recursively by

$$a_1 = \frac{5}{2}; \qquad a_{n+1} = \frac{1}{5}(a_n^2 + 6)$$

We can show that this converges to a limit by showing that it is bounded and decreasing.

First we show it is bounded: in fact, we claim that $2 < a_n < 3$ for all $n \in \mathbb{N}$. Certainly this is true for $n = 1$. If, for some particular n, we have $2 < a_n < 3$, then $4 < a_n^2 < 9$, so

$$\frac{1}{5}(4+6) < a_{n+1} < \frac{1}{5}(9+6)$$

so $2 < a_{n+1} < 3$. It now follows by induction that $2 < a_n < 3$ for all $n \in \mathbb{N}$; in particular, $(a_n)_{n\in\mathbb{N}}$ is bounded below.

Now we prove that $(a_n)_{n\in\mathbb{N}}$ is decreasing. For each $n \in \mathbb{N}$, we have

$$a_{n+1} - a_n = \frac{1}{5}(a_n^2 + 6) - a_n = \frac{1}{5}(a_n^2 - 5a_n + 6) = \frac{1}{5}(a_n - 2)(a_n - 3) < 0$$

since $2 < a_n < 3$.

We now know that (a_n) is decreasing and bounded below, so it must converge to some $a \in \mathbb{R}$ by the Principle of Bounded Monotone Convergence.

Now we know it converges, we can calculate the limit. Since $a_n \to a$ we have

$$a_{n+1} = \frac{1}{5}(a_n^2 + 6) \to \frac{1}{5}(a^2 + 6)$$

by the combination rules. But a_{n+1} also converges to a (shift property), so (uniqueness of limits) we are led to the equation

$$a = \frac{1}{5}(a^2 + 6) \iff a^2 - 5a + 6 = 0 \iff (a-3)(a-2) = 0$$

so either $a = 2$ or $a = 3$. Since (a_n) is a decreasing sequence starting at $5/2$, the limit must be 2.

For a series $\sum a_j$, the partial sums

$$s_n = \sum_{j=1}^{n} a_j$$

form an increasing sequence if and only if $a_j \geq 0$. The Principle of Bounded Monotone Convergence for sequences can be rewritten for series as

10.5 Corollary. *If $(a_j)_{j\in\mathbb{N}}$ is a sequence of non-negative terms and there is a bound b such that for all n*

$$\sum_{j=1}^{n} a_j \leq b$$

then $\sum_{j=1}^{\infty} a_j$ converges and

$$\sum_{j=1}^{\infty} a_j \leq b$$

Once context where this is useful is where the bound arises as a sum of another series. This is an important general result called the *comparison test*.

10.6 Corollary (Comparison Test)**.** *Suppose $0 \leq a_j \leq b_j$ for all j and that $\sum_{j=1}^{\infty} b_j$ converges. Then $\sum_{j=1}^{\infty} a_j$ also converges and*

$$\sum_{j=1}^{\infty} a_j \leq \sum_{j=1}^{\infty} b_j$$

Proof. Summing the inequality $a_j \leq b_j$ from $j = 1$ to n gives

$$\sum_{j=1}^{n} a_j \leq \sum_{j=1}^{n} b_j$$

The RHS is an increasing sequence with limit $\sum_{j=1}^{\infty} b_j$, so by the previous corollary the terms on the left converge to a limit no larger than this; that is,

$$\sum_{j=1}^{\infty} a_j \leq \sum_{j=1}^{\infty} b_j$$

□

10.7 Example. We know from Example 5.10 that

$$\sum_{j=1}^{\infty} \frac{1}{j(j+1)} = 1$$

Observe that if $j \in \mathbb{N}$ then $j \leq j^2$, so $j^2 + j \leq 2j^2$ and hence

$$\frac{1}{2j^2} \leq \frac{1}{j(j+1)}$$

By the Comparison Test,

$$\sum_{j=1}^{\infty} \frac{1}{2j^2} \leq 1$$

and hence

$$\sum_{j=1}^{\infty} \frac{1}{j^2} \le 2$$

This is Euler's [1] famous example; the exact answer turns out to be $\pi^2/6 \approx 1.6$. The importance of this technique is that we have established convergence without having to calculate the value of the sum.

Now suppose $q > 2$. Then $1/j^q < 1/j^2$, so

$$\sum_{j=1}^{\infty} \frac{1}{j^q}$$

converges by comparison with $\sum 1/j^2$. There is certainly no simple formula for the sum, even of

$$\sum_{j=1}^{\infty} \frac{1}{j^3},$$

about which very little is known.

One weakness in this example is that it needed the *ad hoc* inequality

$$\frac{1}{2j^2} \le \frac{1}{j(j+1)}$$

We can avoid having to drive such inequalities using a variant of the test:

10.8 Theorem (Limit Comparison Test). *Suppose $(a_j)_{j\in\mathbb{N}}$ and $(b_j)_{j\in\mathbb{N}}$ are strictly positive sequences and that a_j/b_j converges to a finite limit L as $j \to \infty$. Then:*

1. *If $\sum b_j$ converges, then $\sum a_j$ converges.*

2. *If $L > 0$ then $\sum a_j$ and $\sum b_j$ either both converge or both diverge.*

Proof.

1. Because a_j/b_j converges, it is bounded (Theorem 5.11); that is, there is a constant M such that $a_j/b_j \le M$, or equivalently $a_j \le Mb_j$. Now, $\sum Mb_j$ converges so $\sum a_j$ converges by the Comparison Test.

2. If $L > 0$ then $b_j/a_j \to 1/L$ as $j \to \infty$ and the previous part can also be applied with the roles of (a_j) and (b_j) reversed. This shows that, in this case, convergence of $\sum a_j$ is equivalent to convergence of $\sum b_j$.

□

10.9 Example. If $x > 0$ then the series

$$\sum_{j=0}^{\infty} \frac{x^j}{j!}$$

[1] Leonhard Euler, 1707–1783 (MacTutor, Wikipedia)

converges. We can see this by using the Limit Comparison Test: e.g.

$$(x^j/j!)/(1/2^j) = \frac{(2x)^j}{j!} \to 0$$

as $j \to \infty$. The use of the comparison series $\sum 1/2^j$ was somewhat arbitrary; many other examples would also work.

In fact, this series converges for $x < 0$ (and, trivially, for $x = 0$), to $\exp(x)$; but we are not yet in a position to prove that.

10.10 Example. Consider the series

$$\sum_{j=1}^{\infty} \frac{j^2 - 11}{7j^4 - 8j^2 - 1} \tag{1}$$

Note first that the numerator and denominator both tend to $+\infty$. It follows that they can have only finitely many negative terms, and hence that the series itself can have only finitely many negative terms; as finitely many terms do not affect the convergence of a series, we can proceed as if the series has only positive terms.

Since the terms are of the form quadratic/quartic, we expect them to behave something like $1/j^2$; precisely, we look at the ratio

$$\frac{j^2 - 11}{7j^4 - 8j^2 - 1}\Big/\frac{1}{j^2} = \frac{j^4 - 11j^2}{7j^4 - 8j^2 - 1} \to \frac{1}{7}$$

as $j \to \infty$. Since $\sum_{j=1}^{\infty} 1/j^2$ converges (Example 10.7), the limit comparison test tells us that (1) also converges.

Consider also

$$\sum_{j=1}^{\infty} \frac{3j + 5}{6j^2 - j - 2} \tag{2}$$

Here, the terms really are positive. The terms are linear/quadratic, so we expect behaviour something like $1/j$. We therefore look at the ratio

$$\frac{3j + 5}{6j^2 - j - 2}\Big/\frac{1}{j} = \frac{3j^2 + 5j}{6j^2 - j - 2} \to \frac{1}{2}$$

as $j \to \infty$. Since $\sum_{j=1}^{\infty} 1/j$ diverges (Example 6.8), the limit comparison test tells us that (2) diverges.

11 Subsequences and Cauchy Sequences

If $(a_n)_{n\in\mathbb{N}}$ is a sequence of real numbers, a subsequence is a sequence formed by choosing some of the numbers a_n, in the same order as they appear in the original sequence.

11.1 Definition. Suppose $(a_n)_{n\in\mathbb{N}}$ is a sequence of real numbers. A *subsequence* of (a_n) is a sequence which can be expressed in the form $\left(a_{n_k}\right)_{k\in\mathbb{N}}$ where $(n_k)_{k\in\mathbb{N}}$ is a strictly increasing sequence of natural numbers. (note $n_1 \geq 1$, $n_2 > n_1$, so $n_2 \geq n_1 + 1 \geq 2$; $n_3 > n_2$ so $n_3 \geq n_2 + 1 \geq 3$; by induction, $n_k \geq k$).

11.2 Example. If we let $n_k = 2k$ then

$$a_{n_k} = a_{2k}$$

the subsequence consisting of all terms in even places. Similarly if we let $n_k = 2k - 1$ then

$$a_{n_k} = a_{2k-1}$$

the subsequence consisting of all terms in odd places. If, for example, $a_n = (-1)^n$, this gives us the constant sequence (1) (from the even places) and the constant sequence (-1) from the odd places.

A trivial example of a subsequence is to take $n_k = k$, exhibiting the whole original sequence as a subsequence of itself.

A fundamental fact is that every subsequence of a convergent sequence is itself convergent, to the same limit.

11.3 Theorem. *Suppose $a_n \to a$ as $n \to \infty$ and $(a_{n_k})_{k\in\mathbb{N}}$ is a subsequence of $(a_n)_{n\in\mathbb{N}}$. Then $a_{n_k} \to a$ as $k \to \infty$.*

Proof. Since $a_n \to a$, given $\varepsilon > 0$ there exists $N_\varepsilon \in \mathbb{N}$ such that if $n > N_\varepsilon$ then $|a_n - a| < \varepsilon$. Since (n_k) is a strictly increasing sequence of natural numbers, $n_k \geq k$ for all k. It follows that if $k > N_\varepsilon$ then $n_k > N_\varepsilon$, so $|a - a_{n_k}| < \varepsilon$ and hence $a_{n_k} \to a$. □

Note that, as we saw in Example 11.2, divergent sequences sometimes have convergent subsequences. To investigate this idea further, we introduce the idea of a *peak point*:

11.4 Temporary Definition. Suppose $(a_n)_{n\in\mathbb{N}}$ is a real sequence. Say that n_0 is a *peak point* of the sequence if $a_{n_0} \geq a_n$ for all $n > n_0$ (see Figure 11).

11.5 Lemma. *Every real sequence has a monotone subsequence.*

Proof. We consider two cases:

1. **If there are infinitely many peak points**: Let $(n_k)_{k\in\mathbb{N}}$ be the sequence of all the peak points, arranged in increasing order. Then, for each k, $n_{k+1} > n_k$ and, since n_k is a peak point, $a_{n_{k+1}} \leq a_{n_k}$. This shows that $(a_{n_k})_{k\in\mathbb{N}}$ is a decreasing subsequence of $(a_n)_{n\in\mathbb{N}}$

2. **If there are only finitely many peak points**: choose n_1 such that every $n \geq n_1$ is not a peak point. Because n_1 is not a peak point, there exists $n_2 > n_1$ such that $a_{n_2} > a_{n_1}$. Because n_2 is not a peak point, there exists $n_3 > n_2$ such that $a_{n_3} > a_{n_2}$. Continuing in this way gives a strictly increasing sequence $(n_k)_{k\in\mathbb{N}}$ such that $a_{n_{k+1}} > a_{n_k}$ for all k; that is, an increasing subsequence of $(a_n)_{n\in\mathbb{N}}$.

□

11.6 Corollary (Bolzano[2] -Weierstrass[3] Theorem)**.** *Every bounded sequence has a convergent subsequence.*

[2]Bernard Bolzano, 1781–1848 (MacTutor, Wikipedia)
[3]Karl Weierstrass, 1815–1897 (MacTutor, Wikipedia)

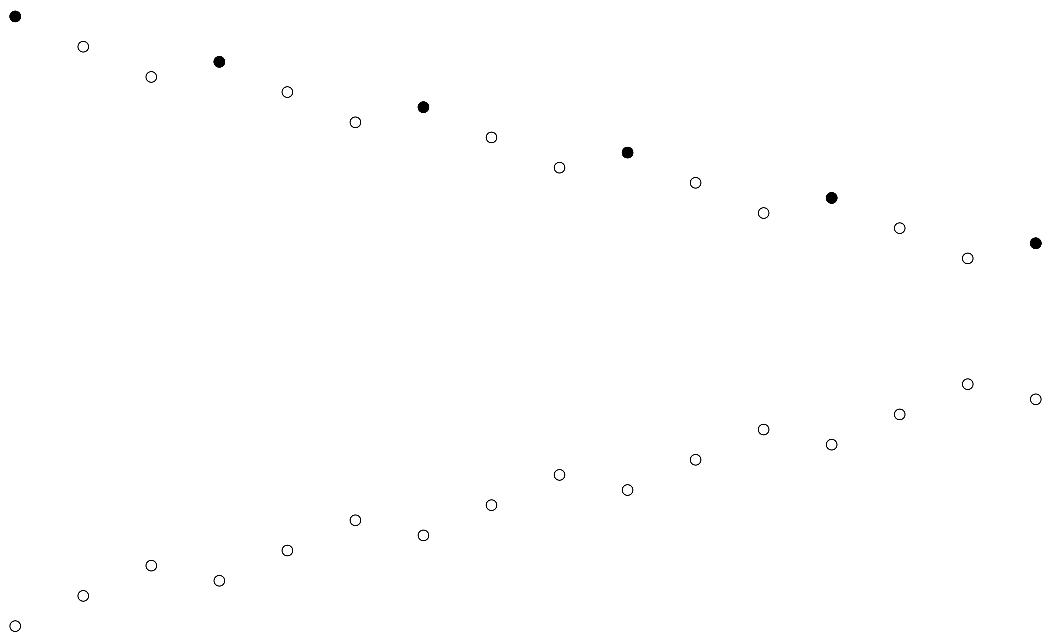


Figure 3: Top: a sequence with peak points shown as solid discs, non-peak points as open circles. Bottom: a sequence with no peak points.

Proof. Suppose $(a_n)_{n\in\mathbb{N}}$ is bounded. Then, by the previous lemma, it has a monotone subsequence. This subsequence is also bounded, so it converges by the principle of bounded monotone convergence. □

11.7 Definition. A real sequence $(a_n)_{n\in\mathbb{N}}$ is called a *Cauchy*[4] *sequence* if for any $\varepsilon > 0$ there exists $N_\varepsilon \in \mathbb{N}$ such that if $m, n > N_\varepsilon$ then $|a_m - a_n| < \varepsilon$.

Note that since $|a_m - a_n| = |a_n - a_m|$, it is equivalent to say that if $m > n > N_\varepsilon$, then $|a_m - a_n| < \varepsilon$; this is sometimes a more convenient form to use.

11.8 Theorem (Cauchy Criterion, Cauchy's general Principle of Convergence)**.** *Every convergent sequence is a Cauchy sequence. More importantly, every Cauchy sequences converges, so the Cauchy property is equivalent to convergence.*

Proof.

1. **Suppose $a_n \to a$ as $n \to \infty$.** Then for any $\varepsilon > 0$ there exists $N_\varepsilon \in \mathbb{N}$ such that if $n > N_\varepsilon$ then $|a_n - a| < \varepsilon/2$. If also $m > N_\varepsilon$ then $|a_m - a| < \varepsilon/2$ so

$$|a_m - a_n| = |(a_m - a) + (a - a_n)| < |a_m - a| + |a - a_n| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

   which shows that $(a_n)_{n\in\mathbb{N}}$ is Cauchy.

2. **Suppose $(a_n)_{n\in\mathbb{N}}$ is Cauchy**. If we put $\varepsilon = 1$ in the definition of the Cauchy property, we see that there exists $N \in \mathbb{N}$ such that if $n > N$ then $|a_n - a_m| < 1$. In particular, $|a_n - a_{N+1}| < 1$, so $a_{N+1} - 1 < a_n < a_{N+1} + 1$ for all $n > N$. This shows that $(a_n)_{n\in\mathbb{N}}$ is bounded.

[4]Augustin-Louis Cauchy, 1789–1857 (MacTutor, Wikipedia)

By the Bolzano-Weierstrass Theorem, $(a_n)_{n\in\mathbb{N}}$ has a convergent subsequence, say $a_{n_k} \to a$ as $k \to \infty$. We now claim that $a_n \to a$ as $n \to \infty$. Given $\varepsilon > 0$ there exists $N_\varepsilon \in \mathbb{N}$ such that if $n > N$ then $|a_m - a_n| < \varepsilon/2$ and K_ε such that if $k > K_\varepsilon$ then $|a_{n_k} - a| < \varepsilon/2$. Choose k such that $k > K_\varepsilon$ and $n_k > N_\varepsilon$. Then

$$|a_n - a| = |(a_n - a_{n_k}) + (a_{n_k} - a)| \le |a_n - a_{n_k}| + |a_{n_k} - a| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

This is true for all $n > N_\varepsilon$, so $a_n \to a$ as $n \to \infty$.

□

Remark. The Cauchy criterion is, of course, rather similar to the definition of convergence. They both start

For any $\varepsilon > 0$ there exists $N_\varepsilon \in \mathbb{N}$ such that ...

but they continue in different ways. The definition of convergence concludes

... if $n > N_\varepsilon$ then $|a_n - a| < \varepsilon$.

whereas the definition of the Cauchy criterion concludes

... if $m > N_\varepsilon$ and $n > N_\varepsilon$ then $|a_m - a_n| < \varepsilon$.

The reason that they are similar, but different, is that they answer similar, but different questions.

The definition of convergence addresses the question "Does this sequence converge to *this particular* limit a?" by comparing the terms of the sequence with a.

The Cauchy condition, on the other hand, addresses the question "Does this sequence converge to *some* limit?", by comparing the terms of the sequence with each other.

11.9 Corollary. *The series $\sum a_j$ converges if and only if it satisfies the Cauchy criterion: for any $\varepsilon > 0$ there exists $N_\varepsilon \in \mathbb{N}$ such that if $m, n \in \mathbb{N}$ with $n > m > N_\varepsilon$ then*

$$\left| \sum_{j=m+1}^{n} a_j \right| < \varepsilon$$

Proof. Let s_n be the partial sum

$$s_n = \sum_{j=1}^{n} a_j$$

The series $\sum a_j$ converges if and only if (s_n) converges. By the Cauchy criterion for sequences, this is equivalent to the statement:

For any $\varepsilon > 0$ there exists $N_\varepsilon \in \mathbb{N}$ such that if $m, n \in \mathbb{N}$ with $m > N_\varepsilon$ and $n > N_\varepsilon$ then $|s_m - s_n| < \varepsilon$.

If $n > m$, this is exactly what it written above, because then

$$s_n - s_m = \sum_{j=m+1}^{n} a_j$$

We should also consider the cases $m > n$ and $m = n$. The case $m > n$ can be ignored by symmetry: $|s_m - s_n| = |s_n - s_m|$. The case $m = n$ is trivial: $s_m - s_n = 0$. □

12 Two special convergence criteria for infinite series: Leibniz and Cauchy Condensation tests

The problem when dealing with infinite series is that in most cases, it's difficult or even impossible to determine the partial sums s_n for each $n \in \mathbb{N}$ and then check whether $(s_n)_{n\in\mathbb{N}}$ converges. Therefore, we want convergence criteria for which we don't have to determine all the s_n one by one. We have already seen one such test: the Comparison Test, Corollary 10.6. We now develop some more.

12.1 Theorem (Leibniz[5] alternating series test)**.** *Suppose $(a_j)_{j\in\mathbb{N}}$ is a decreasing sequence tending to zero. Then*

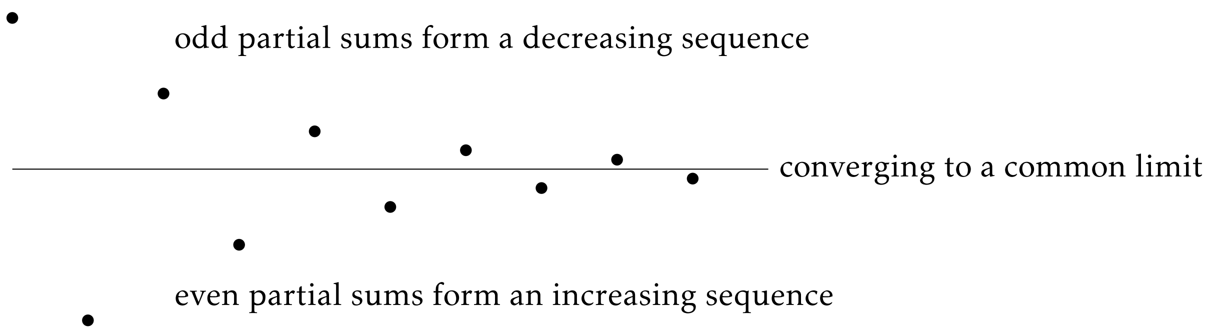
$$\sum_{j=1}^{\infty}(-1)^{j+1}a_j$$

converges.

Proof. Let

$$s_n = \sum_{j=1}^{n}(-1)^{j+1}a_j$$

The proof is contained in the following diagram:



Firstly, the even partial sums form an increasing sequence:

$$s_{2n+2} - s_{2n} = a_{2n+1} - a_{2n+2} \geq 0$$

since the (a_j) are decreasing.

Secondly, the odd partial sums form a decreasing sequence:

$$s_{2n+3} - s_{2n+1} = -a_{2n+2} + a_{2n+3} \leq 0$$

again since the (a_j) are decreasing.

[5]Gottfried Wilhelm Leibniz, 1646–1716 (MacTutor, Wikipedia)

Finally,

$$\begin{aligned} s_{2n+1} &= s_{2n} + a_{2n+1} \\ &\geq s_{2n} \\ &\geq s_2 \end{aligned} \tag{*}$$

because $a_{2n+1} \geq 0$ and (s_{2n}) is increasing. The (s_{2n+1}) are thus decreasing and bounded below; they therefore (Bounded Monotone Convergence) converge, say $s_{2n+1} \to s$ as $n \to \infty$. In (*), $a_{2n+1} \to 0$ as $n \to \infty$; it follows that $s_{2n} \to s$ as $n \to \infty$. Now, with s_{2n+1} and s_{2n} both converging to s, it follows that $s_n \to s$ as $n \to \infty$ (Exercise 4.3). □

12.2 Example. The series

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots$$
$$1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \dots$$

both converge by the Leibniz test. The limits are, in fact, $\log(2)$ and $\pi/4$, but the test can't tell us this: only that the series converge.

12.3 Theorem (Cauchy's Condensation Test)**.** *Suppose $(a_j)_{j\in\mathbb{N}}$ is a decreasing sequence of non-negative terms. Then the following are equivalent:*

1. *$\sum_{j=1}^{\infty} a_j$ converges;*
2. *$\sum_{k=0}^{\infty} 2^k a_{2^k}$ converges.*

Proof. (Compare with the argument for the harmonic series, Example 6.8). The idea is that both series are of non-negative terms, so their partial sums both form increasing sequences. The Principle of Bounded Monotone Convergence tells us that convergence is equivalent to boundedness; it is therefore enough to show that the partial sums in (1) are bounded if and only if the partial sums in (2) are bounded.

Boundedness in (1) implies boundeness in (2):

Let S be an upper bound for the partial sums in (1). Then, in particular,

$$a_1 + \underbrace{a_2}_{=1a_2} + \underbrace{a_3 + a_4}_{\geq 2a_4} + \underbrace{a_5 + \dots + a_8}_{\geq 4a_8} + \dots + \underbrace{a_{2^{n-1}+1} + \dots + a_{2^n}}_{\geq 2^{n-1}a_{2^n}} \leq S$$

Counting from $k = 1$, the kth block (indicated by a brace below the block; the first term, a_1, does not belong to any block) contains 2^{k-1} terms, of which the smallest is (because $(a_j)_{j\in\mathbb{N}}$ is decreasing) the last, namely a_{2^k}. The sum of terms in the kth block is thus greater than or equal to $2^{k-1}a_{2^k}$. We therefore have

$$a_1 + \sum_{k=1}^{n} 2^{k-1} a_{2^k} \leq S$$

and hence

$$\sum_{k=1}^{n} 2^k a_{2^k} \leq 2S - 2a_1$$

Adding a_1 to both sides gives

$$\sum_{k=0}^{n} 2^k a_{2^k} \leq 2S - a_1$$

which shows that the partial sums in (2) are bounded above by $2S - a_1$.

Boundedness in (2) implies boundeness in (1):

Suppose (2) converges and let

$$C = \sum_{k=0}^{\infty} 2^k a_{2^k}$$

Now consider the partial sum to $2^n - 1$ terms of (1):

$$\underbrace{a_1}_{=1a_1} + \underbrace{a_2 + a_3}_{\leq 2a_2} + \underbrace{a_4 + \cdots + a_7}_{\leq 4a_4} + \cdots + \underbrace{a_{2^{n-1}} + \cdots + a_{2^n-1}}_{\leq 2^{n-1}a_{2^{n-1}}}$$

Now counting from $k = 0$, the kth block (here every term belongs to a block) contains 2^k terms, of which the largest is (because $(a_j)_{j\in\mathbb{N}}$ is decreasing) the first, namely a_{2^k}. The sum of terms in the kth block is thus less than or equal to $2^k a_{2^k}$. We therefore have

$$\sum_{j=1}^{2^n-1} a_j \leq \sum_{k=0}^{n-1} 2^k a_{2^k} \leq C$$

(because the second sum is a partial sum from (2)). Finally, for any $n \in \mathbb{N}$, $n \leq 2^n - 1$, so

$$\sum_{j=1}^{n} a_j \leq \sum_{j=1}^{2^n-1} a_j \leq C$$

so all the partial sums in (1) are bounded above by C. □

12.4 Example. Not surprisingly, the condensation test establishes the divergence of the harmonic series (Example 6.8): the convergence of the two series

$$\sum_{j=1}^{\infty} \frac{1}{j} \tag{1}$$

$$\sum_{k=0}^{\infty} 2^k \frac{1}{2^k} = \sum_{k=0}^{\infty} 1 \tag{2}$$

are equivalent, by the condensation test; since (2) obviously diverges, (1) also diverges.

Now consider the more general series

$$\sum_{j=1}^{\infty} \frac{1}{j^p} \tag{3}$$

for some $p \in \mathbb{Q}$. This plainly diverges if $p \leq 0$, because the terms do not converge to zero. If $p > 0$ then it is a sum of non-negative, decreasing terms, so the condensation test applies, and convergence of (3) is equivalent to convergence of the condensed series

$$\sum_{k=0}^{\infty} 2^k \frac{1}{(2^k)^p} = \sum_{k=0}^{\infty} (2^{1-p})^k \tag{4}$$

This is a geometric sum, so converges if and only if $2^{1-p} < 1$; that is, if $1 - p < 0$ or equivalently $p > 1$. We thus have

$$\sum_{j=1}^{\infty} \frac{1}{j^p} \begin{cases} \text{converges} & p > 1 \\ \text{diverges} & p \leq 1 \end{cases}$$

Typically, the condensation test acts as an amplifier: it converts slow convergence or divergence into faster convergence or divergence.

13 Absolute Convergence and the Ratio Test

13.1 Definition. A series $\sum_{j=1}^{\infty} a_j$ is called *absolutely convergent* if $\sum_{j=1}^{\infty} |a_j|$ is convergent. Note that, since the partial sums $\sum_{j=1}^{n} |a_j|$ form an increasing sequence, this is equivalent to the partial sums being bounded above.

A series which is convergent, but not absolutely convergent, is called *conditionally convergent*.

13.2 Theorem. *Absolute convergence implies convergence; the converse is not true.*

Proof. Suppose $\sum_{j=1}^{\infty} |a_j|$ converges and $\varepsilon > 0$. Because $\sum |a_j|$ converges, it satisfies the Cauchy condition (Corollary 11.9): there exists $N_\varepsilon \in \mathbb{N}$ such that if $n > m > N_\varepsilon$ then

$$\sum_{j=m+1}^{n} |a_j| < \varepsilon$$

By the triangle inequality,

$$\left| \sum_{j=m+1}^{n} a_j \right| \leq \sum_{j=m+1}^{n} |a_j| < \varepsilon$$

so the partial sums $\sum_{j=1}^{n} a_j$ also form a Cauchy sequence; it follows that they converge.

This establishes that absolute convergence implies convergence. To show that the converse is not true, consider the series

$$\sum_{j=1}^{\infty} \frac{(-1)^{j+1}}{j}.$$

This converges by the Leibniz test (Example 12.2), but the sum of its absolute values is the harmonic series

$$\sum_{j=1}^{\infty} \frac{1}{j}$$

which diverges (Example 6.8, Example 12.4). This is thus an example of a conditionally convergent series. □

13.3 Example. Consider the series

$$\sum_{j=1}^{\infty} \frac{(-1)^{j+1}}{j^2} \tag{*}$$

We can see that this converges in various ways. One is to use the Leibniz test (Theorem 12.1). Another is to observe that $|(-1)^{j+1}/j^2| = 1/j^2$ and that

$$\sum_{j=1}^{\infty} \frac{1}{j^2}$$

converges (Example 10.7, Example 12.4).

As usual, we cannot determine the limit from this technique: it turns out that (*) evaluates to $\pi^2/12$.

More generally, suppose $\sigma_j = \pm 1$ for $j \in \mathbb{N}$. Then the series

$$\sum_{j=1}^{\infty} \frac{\sigma_j}{j^2}$$

is absolutely convergent, and hence convergent. The Lebniz test cannot handle this example, because we have no information about how the $\pm$ signs are distributed among the terms.

Absolute convergence is particularly powerful when used in combination with the Comparison Test (Corollary 10.6):

13.4 Theorem (Comparison and limit comparison test for signed terms)**.** *Comparison test: if $|a_j| \le b_j$ for all $j \in \mathbb{N}$ and $\sum_{j=1}^{\infty} b_j$ converges, then $\sum_{j=1}^{\infty} a_j$ converges absolutely and*

$$\left|\sum_{j=1}^{\infty} a_j\right| \le \sum_{j=1}^{\infty} |a_j| \le \sum_{j=1}^{\infty} b_j$$

Limit comparison test: if $b_j > 0$, $|a_j/b_j| \to L \in \mathbb{R}$ as $j \to \infty$ and and $\sum_{j=1}^{\infty} b_j$ converges, then $\sum_{j=1}^{\infty} a_j$ converges absolutely.

The proof is immmediate, given the fact that absolute convergence implies convergence. Note that there is no restriction on the sign of the a_j, but the comparison terms b_j are always non-negative: in the comparison test, $b_j \ge 0$ implicitly and in the limit comparison test $b_j > 0$ explicitly.

13.5 Example. Suppose $(a_j)_{j\in\mathbb{N}}$ is any sequence of integers. Then

$$\sum_{j=1}^{\infty} \frac{(-1)^{a_j}}{j^2+1}$$

converges, by the comparison inequality

$$\left|\frac{(-1)^{a_j}}{j^2+1}\right| \le \frac{1}{j^2}$$

13.6 Theorem (Ratio Test). *Suppose $(a_j)_{j\in\mathbb{N}}$ is a sequence of non-zero terms such that $|a_{j+1}/a_j| \to r$ as $j \to \infty$. Then*

$$\sum_{j=1}^{\infty} a_j \begin{cases} \text{converges absolutely} & r < 1 \\ \text{diverges} & r > 1 \\ ??? & r = 1 \end{cases}$$

The case that $|a_{j+1}/a_j| \to \infty$ as $j \to \infty$ should be interpreted as a limit greater than 1*; that is, the series diverges in this case.*

Proof. Consider first the convergence case: suppose $|a_{j+1}/a_j| \to r < 1$, and choose s such that $r < s < 1$. Then (Theorem 5.12) there exists $N \in \mathbb{N}$ such that if $j \geq N$ then $|a_{j+1}/a_j| < s$. Then $|a_{N+1}| \leq s|a_N|$, $|a_{N+2}| \leq s|a_{N+1}| \leq s^2|a_N|$, and by induction $|a_{N+k}| < s^k|a_N|$ ($k \in \mathbb{N}$). Since $s < 1$, the geometric series $\sum_{k=1}^{\infty} s^k$ converges; by the Comparison Test, $\sum_{k=1}^{\infty} a_{N+k}$ converges absolutely, and hence $\sum_{j=1}^{\infty} a_j$ converges absolutely.

Now consider the divergence case: suppose $|a_{j+1}/a_j| \to r > 1$, and choose s such that $r > s > 1$. Then (Theorem 5.12) there exists $N \in \mathbb{N}$ such that if $j \geq N$ then $|a_{j+1}/a_j| > s$. As above, by induction, we see that $|a_{N+k}| > s^k|a_N|$ ($k \in \mathbb{N}$). Since $s > 1$, $s^k \to \infty$ as $k \to \infty$; it follows that $|a_{N+k}| \to \infty$ as $k \to \infty$, so the series diverges. □

13.7 Example. Consider

$$\sum_{n=0}^{\infty} \frac{x^n}{n!}$$

for some $x \in \mathbb{R}$, $x \neq 0$ (if $x = 0$ the ratio test does not work but the series trivially converges). The absolute ratio of two successive terms is

$$\left| \frac{x^{n+1}}{(n+1)!} \frac{n!}{x^n} \right| = \frac{|x|}{n+1} \to 0$$

as $n \to \infty$. It follows that the series converges. Compare Example 10.9, which required a little more work and at the time only worked for $x \geq 0$.

Consider

$$\sum_{k=1}^{\infty} \frac{(2k)!}{(k!)^2}$$

The absolute ratio of two successive terms is

$$\frac{(2k+2)!}{((k+1)!)^2} \bigg/ \frac{(k!)^2}{(2k)!} = 2\frac{2k+1}{k+1} \to 4$$

as $k \to \infty$. It follows that the series diverges.

Finally, consider

$$\sum_{j=1}^{\infty} \frac{1}{j^p}$$

We already know (Example 12.4) that this converges if $p > 1$ and diverges if $p \leq 1$. Trying the ratio test:

$$\frac{j^p}{(j+1)^p} = \left(\frac{j}{j+1}\right)^p \to 1^p = 1$$

so the test is inconclusive.

14 Power Series

14.1 Definition. Fix $x_0 \in \mathbb{R}$ and a sequence $(a_n)_{n\geq 0}$, and consider for $x \in \mathbb{R}$ the series

$$\sum_{n=0}^{\infty} a_n(x-x_0)^n = a_0 + a_1(x-x_0) + a_2(x-x_0)^2 + \dots$$

This is called a *power series* or *Taylor*[6] *series* centred at x_0.

In this definition, think of the centre x_0 and the coefficients (a_n) as being fixed, and x as being variable, so the series defines a function of x wherever it converges. For this to be useful, we need to understand for which x the series converges. One immediate observation is that it trivially converges to a_0 when $x = x_0$, since all but the first term is zero (note the convention $0^0 = 1$ in this context).

14.2 Example. The simplest example is the geometric series:

$$\sum_{n=0}^{\infty} x^n \begin{cases} \text{converges to } 1/(1-x) & (|x| < 1) \\ \text{diverges} & (|x| \geq 1) \end{cases}$$

We thus have convergence on the interval $(-1, 1)$, centred at 0; this turns out to be typical of power series.

14.3 Lemma. *Consider the power series*

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

Let

$$S = \{r \geq 0 : a_n r^n \text{ is bounded as } n \to \infty\}$$

Then:

(a) If $r \in S$ and $|x-x_0| < r$ then (*) converges absolutely.

(b) If $|x-x_0| \notin S$ then (*) diverges.

Proof. (a) If $|x-x_0| < r \in S$ then $r > 0$ and there exists $C > 0$ such that $|a_n r^n| \leq C$ for all n. Write

$$|a_n(x-x_0)^n| = \left(\frac{|x-x_0|}{r}\right)^n |a_n r^n| \leq C\left(\frac{|x-x_0|}{r}\right)^n$$

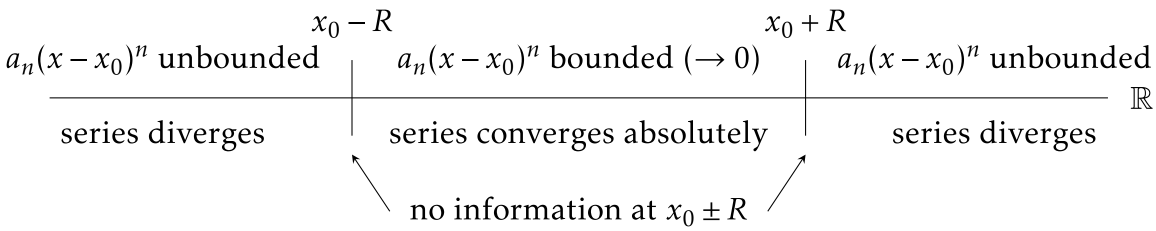
The sum of the terms on the RHS converges (geometric series) so, by comparison, the sum of the terms on the LHS converges; that is, (*) converges absolutely.

(b) If $|x-x_0| \notin S$, then $|a_n(x-x_0)^n|$ is unbounded as $n \to \infty$; in (*), the terms do not tend to zero so the series diverges.

□

[6] Brook Taylor (1685–1731) (MacTutor, Wikipedia)

Figure 4: Convergence of a power series with finite radius of convergence



14.4 Theorem. *Of the set S and the series (*) in the previous Lemma, exactly one of these three statements is true:*

1. *S is unbounded and (*) converges absolutely for all $x \in \mathbb{R}$.*
2. *$S = \{0\}$ and (*) converges only for $x = x_0$.*
3. *S is bounded and not equal to $\{0\}$; with $R = \sup(S) > 0$, (*) converges absolutely for $|x - x_0| < R$ and diverges for $|x - x_0| > R$.*

Proof.

1. Suppose S is unbounded. Then for any $x \in \mathbb{R}$ there exists $r \in S$ such that $r > |x - x_0|$; by part (a) of the Lemma, (*) converges absolutely.
2. Suppose $S = \{0\}$. Then, for any $x \neq x_0$, $|x - x_0| \notin S$; by part (b) of the Lemma, (*) diverges.
3. The remaining case is that S is bounded, with $S \neq \{0\}$. We have $0 \in S$, so S is non-empty; we can therefore let $R = \sup(S)$. Since $S \subseteq [0, \infty)$ and $S \neq \{0\}$, S must contain a strictly positive number, and hence $R > 0$. If $|x - x_0| < R$ then we must have $|x - x_0| < r$ for some $r \in S$ (otherwise, $|x - x_0|$ would be an upper bound for S with $|x - x_0| < R = \sup(S)$). It follows from part (a) of the Lemma that (*) converges. If $|x - x_0| > R = \sup(S)$ then $|x - x_0| \notin S$ so, by part (b) of the Lemma, (*) diverges.

□

14.5 Definition. The quantity R in case (3) of the corollary is called the *radius of convergence* of the power series. We extend the notion by writing $R = \infty$ in case (1) and $R = 0$ in case (2).

The radius of convergence is often calculated by the ratio test, because

$$\frac{a_{n+1}(x - x_0)^{n+1}}{a_n(x - x_0)^n} = \frac{a_{n+1}}{a_n}(x - x_0)$$

14.6 Example. To find the radius of convergence of

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

we look at the absolute ratio of two terms (for $x \neq 1$; then $x = 1$ case is trivial):

$$\left| \frac{n+1}{n+2}(x-1)^{n+2} \bigg/ \frac{n}{n+1}(x-1)^n \right| = \frac{(n+1)^2}{n(n+2)}|x-1| \to |x-1|$$

as $n \to \infty$, the ratio test tells us that we have convergence if $|x-1| < 1$ and divergence if $|x-1| > 1$; that is, $R = 1$.

We have already seen (Example 13.7) that

$$\sum_{n=0}^{\infty} \frac{x^n}{n!}$$

converges for all x, so we have radius of convergence $R = \infty$.

If we look at

$$\sum_{n=1}^{\infty} n^n (x+2)^n$$

then, provided $x \neq -2$, the terms are non-zero and the ratio is

$$\left| \frac{(n+1)^{n+1}(x+2)^{n+1}}{n^n (x+2)^n} \right| = (n+1)\left(1 + \frac{1}{n}\right)^n |x+2|$$

Now, $(1 + 1/n)^n > 1$ (in fact, as $n \to \infty$, $(1 + 1/n)^n \to \mathrm{e}$ but we don't need to know this) so, the case $x = -2$ not being under consideration, the ratio tends to $+\infty$. We thus have divergence for $x \neq -2$, or radius of convergence $R = 0$. Alternatively, we could use the hierarchy of limits to observe that $n^n(x+2)^n \to \infty$ as $n \to \infty$ for any $x \neq -2$.

14.7 Example. Some power series have only odd or only even terms. The ratio test still works for these. For example,

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

The absolute ratio of two successive terms is

$$\left| \frac{((n+1)!)^2 x^{2n+1}}{(2n+1)!} \bigg/ \frac{(n!)^2}{(2n)!} x^{2n+1} \right| = \frac{(n+1)^2}{(2n+2)(2n+1)}|x|^2 \to \frac{|x|^2}{4}$$

as $n \to \infty$. By the ratio test, we have convergence if $|x|^2 < 4$ and divergence if $|x|^2 > 4$, so the radius of convergence is 2.

The series

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

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

can both be seen to have infinite radius of convergence by the ratio test:

$$\left|\frac{x^{2n+3}}{(2n+3)!}\frac{(2n+1)!}{x^{2n+1}}\right| = \frac{x^2}{(2n+2)(2n+3)}$$
$$\left|\frac{x^{2n}}{(2n)!}\frac{x^{2n+2}}{(2n+2)!}\right| = \frac{x^2}{(2n+2)(2n+1)}$$

both of which tend to $0 < 1$ as $n \to \infty$.

14.8 Example. Suppose we have a power series

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

with radius of convergence R, where $0 < R < \infty$, so we have convergence when $|x-x_0| < R$ and divergence when $|x-x_0| > R$. What happens when $|x-x_0| = R$, i.e. $x-x_0 = \pm R$? The answer is that all possible combinations of behaviours can occur.

The simplest example is the geometric series

$$\sum_{n=0}^{\infty} x^n$$

which has radius of convergence 1, and converges to $1/(1-x)$. This diverges at both endpoints ± 1.

If we look at

$$\sum_{n=1}^{\infty} \frac{x^n}{n},$$

which also has radius of convergence 1, then this converges when $x = -1$ (Leibniz test) and diverges if $x = 1$ (harmonic series).

Finally, the series

$$\sum_{n=2}^{\infty} \frac{x^n}{n(n-1)}$$

also has radius of convergence 1, but converges at both endpoints ± 1 because it is absolutely convergent: $|(\pm 1)^n[n/n(n-1)]| = 1/[n(n-1)]$ and we know that $\sum_{n=2}^{\infty} 1/[n(n-1)]$ converges by e.g. limit comparison with $\sum 1/n^2$.

Incidentally, these examples were built by starting with the geometric series and integrating twice; the sums of the second and third examples, when they converge, are respectively $-\log(1-x)$, and $(1-x)\log(1-x)+x$.

15 Elementary Functions Defined by Power Series

If we have a power series

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

with finite radius of convergence R, we can think of this as defining a function $f : (x_0-R, x_0+R) \to \mathbb{R}$ in the obvious way:

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

The series might or might not naturally define $f(x_0-R)$ and $f(x_0+R)$; we shall not investigate these cases.

Similarly, if the power series has infinite radius of convergence, it defines a function on the whole of $\mathbb{R}$.

We can now make the crucial move to understand the elementary (exponential and trigonometric) functions: we break in by *defining* them to be the values of convergent power series (with infinite radius of convergence).

15.1 Definition. For $x \in \mathbb{R}$, define:

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

At this stage, we cannot go much further than the definitions and the obvious facts that $\sin(0) = 0$, $\cos(0) = \exp(0) = 1$ and that sin is odd and cos is even. The properties of these functions (sketched in Section 1) will be established as the course progresses.

16 Part III: key points

The main idea in Part III is that we can prove that a sequence or series converges, without knowing in advance what the limit is, and without ending up with any kind of concrete formula for the limit. The purpose of this is to create useful new numbers and functions. We saw the genesis of this in Part I, where we constructed $\sqrt{2}$ directly from the completeness axiom (Example 3.6), but Part III offers much more powerful tools: in particular, the idea of defining exponential and trigonometric functions as the sums of convergent power series (Section 15). In later sections, we shall see that all of the properties of these elementary functions can be deduced from the power series representations.

A sequence is called *monotone (or monotonic)* if it is either *increasing* (for all n, $a_{n+1} \geq a_n$) or *decreasing* (for all n, $a_{n+1} \leq a_n$). The key fact about monotone sequences is the *Principle of Bounded Monotone Convergence*: every bounded, monotonic sequence converges. For increasing sequences, the limit is the supremum of the terms of the sequence; for decreasing sequences, the limit is the infimum. This is where the theory of bounds and completeness from Part I meets the theory of convergence.

The Principle of Bounded Monotone Convergence has two major theoretical consequences, known as the *Bolzano-Weierstrass Theorem* (Corollary 11.6) and *Cauchy's General Principle of Convergence* (Theorem 11.8).

The Bolzano-Weierstrass Theorem states that every bounded sequence has a convergent subsequence; that is, if there is a constant K such that $|a_n| \leq K$ for all n then there is a strictly increasing sequence of natural numbers $(n_k)_{k\in\mathbb{N}}$ and a real number a such that $a_{n_k} \to a$ as $k \to \infty$.

Cauchy's General Principle of Convergence states that a sequence $(a_n)_{n\in\mathbb{N}}$ converges if and only if it is a *Cauchy sequence*, which is to say that for any $\varepsilon > 0$ there exists $N_\varepsilon \in \mathbb{N}$ such that if $m, n > N_\varepsilon$ then $|a_m - a_n| < \varepsilon$ (Definition 11.7). There is also a variant for series (Corollary 11.9) where the relevant inequality is

$$\left|\sum_{j=m+1}^{n} a_j\right| < \varepsilon$$

for $n > m > N_\varepsilon$. The Principle of Bounded Monotone Convergence, the Bolzano-Weierstrass Theorem and Cauchy's General Principle of Convergence have one important point in common: they assert the existence of the limit of a sequence, without the limit being known in advance (cf. the definition of convergence, in which the value of the limit is part of the definition). This means that they can be used to construct numbers with certain properties (e.g., solutions to equations), just as the axiom of completeness was used in Part I to construct square roots.

Power series have the form

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

We think of the Taylor coefficients $(a_n)_{n\geq 0}$ as fixed and the series as definining a function of x. There are three basic possibilities: the series could converge only for $x = x_0$, could converge for all x, or could converge for $|x - x_0| < R$ and diverge for $|x - x_0| > R$.

This R, where it exists, is called the radius of convergence of the power series; we extend this idea by saying that, in the first two cases, $R = 0$ or $R = \infty$, respectively (Theorem 14.4, Definition 14.5). In concrete examples, tha ratio test is often a good way to compute the radius of convergence.

The remaining topic in this section is the use of convergence tests for infinite series; these are summarised in a separate section, below.

Literature:

1. F. Mary Hart *Guide to Analysis* (2nd Edition, Palgrave Macmillan 2001, JBM classmark S7 Har) Chapters 2, 3.

2. Keith E. Hirst, *Numbers, Sequences and Series* (Edward Arnold, 1995, JBM classmark S2.81 Hir) Chapters 7, 8.

3. any other introduction to Real Analysis; browse classmark S7 in the library.

17 Convergence tests: summary

Convergence tests are techniques which we can apply to tell us whether an infinite series converges or diverges. Different examples typically need different tests (though sometimes more than one test will work); the following notes describe the tests and give some ideas about the type of examples that the tests apply to.

> If $a_j \not\to 0$ as $j \to \infty$ then $\sum_{j=1}^{\infty} a_j$ diverges.

This is most primitive test of all. Take care not to apply it the wrong way round: if $a_j \to 0$ then the series might converge or diverge (remember the harmonic series).

> Leibniz alternating series test: if $(a_j)_{j\in\mathbb{N}}$ is decreasing and $a_j \to 0$ as $j \to \infty$ then the series $\sum_{j=1}^{\infty}(-1)^{j+1} a_j$ converges.

This is potentially useful for any series where the terms alternate in sign, but take care to check the other hypotheses as well. The Leibniz test is the only test in the course which can detect conditional convergence.

> Cauchy condensation test: if $(a_j)_{j\in\mathbb{N}}$ is a non-negative, decreasing sequence then the series
> $$\sum_{j=1}^{\infty} a_j; \qquad \sum_{k=0}^{\infty} 2^k a_{2^k}$$
> either both converge or both diverge.

This is good for slowly-converging or slowly-diverging series, because the condensed series on the right typically converges or diverges much more quickly than the original series on the left, making it easier to see how it behaves. Examples like $\sum_{j=1}^{\infty} 1/j^p$ and $\sum_{j=1}^{\infty} 1/(j(\log(j)^p))$ (in both cases, converge if $p > 1$, diverge if $p \leq 1$) can be handled by condensation. For more algebraically complicated examples, consider using it in combination with the limit comparison test.

> Absolute convergence implies convergence: if $\sum_{j=1}^{\infty} |a_j|$ converges, then so does $\sum_{j=1}^{\infty} a_j$.

This is most useful in combination with the comparison test.

> Comparison test: if $|a_j| \leq b_j$ and $\sum_{j=1}^{\infty} b_j$ converges, then $\sum_{j=1}^{\infty} a_j$ converges absolutely.

This is very powerful (most of the earlier examples in the notes using bounded monotone convergence can be recast into comparison test examples), but finding the right comparison sequence (b_j) can be tricky. As a minimum, you need to know that $\sum_{j=1}^{\infty} 1/j^p$ converges if $p > 1$ and diverges if $p \leq 1$ (this follows from the condensation test), and that $\sum_{j=1}^{\infty} x^j$ converges if $|x| < 1$ and diverges if $|x| \geq 1$. The limit comparison and ratio tests are based on the comparison test, and are often easier to use for specific examples.

Limit comparison test: if $a_j/b_j \to c \neq 0$ then either the series

$$\sum_{j=1}^{\infty} a_j; \qquad \sum_{j=1}^{\infty} b_j$$

are both absolutely convergent, or neither is absolutely convergent.

This is often easier to use than the comparison test in specific examples. e.g., if $a_j = (12j^2+7)/(9j^4-6j^3+1)$ then $b_j = 1/j^2$ works in the limit comparison test and proves absolute convergence; finding an appropriate sequence for the comparison test is much harder! In this example, the convergence of $\sum_{j=1}^{\infty} 1/j^2$ follows from the Condensation Test; applying the Condensation Test directly to $\sum_{j=1}^{\infty} a_j$ would be impractical, but the two tests together work well. Any sum of a rational function (ratio of polynomials) can be handled like this.

Ratio test: if $|a_{j+1}/a_j| \to r$ as $j \to \infty$ then $\sum_{j=1}^{\infty} a_j$ converges absolutely if $r < 1$, and diverges if $r > 1$.

This is only useful for examples which converge or diverge quite quickly; other series tend to produce the inconclusive result $r = 1$. If a_j has terms like $j!$, x^j or j^j, then there is typically a good deal of algebraic simplification in a_{j+1}/a_j, making the test easy to use; e.g., if $a_j = j!/(2j)!$ then $a_{j+1}/a_j = 1/(2(2j+1)) \to 0$, so $\sum_{j=1}^{\infty} j!/(2j)!$ converges. The ratio test is the prime tool in calculating the radius of convergence of a power series $\sum_{n=0}^{\infty} a_n(x-x_0)^n$, because of the simplification

$$\left|\frac{a_{n+1}(x-x_0)^{n+1}}{a_n(x-x_0)^n}\right| = \left|\frac{a_{n+1}}{a_n}\right||x-x_0|$$

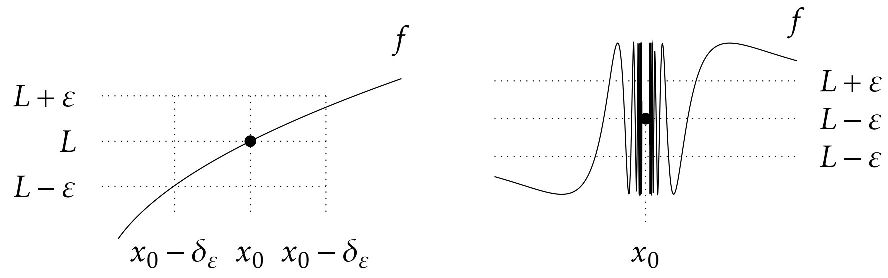


Figure 5: (Left) A valid ε-δ relationship for the limit $f(x) \to L$ as $x \to x_0$ (Definition 18.1). When x is closer to x_0 than δ_ε, $f(x)$ is closer to L than ε; the curve between $x_0 \pm \delta$ is contained in the rectangle with vertical sides at $x_0 \pm \delta$ and horizontal sides at $L \pm \varepsilon$. Right: a function without a limit at x_0. For small enough ε (e.g. as shown), no rectangle with horizontal sides at $L \pm \varepsilon$ contains the whole curve between $x_0 \pm \delta$, for any L or any δ.

Part IV

Limits and Continuity

18 Limits at a point

The previous parts of the course have been about limits as an integer n approaches infinity. We now turn to what happens as a real number x approaches a finite number x_0. The idea is to leave the ε part of the definition alone but replace $n > N_\varepsilon$, meaning n is very large, with $0 < |x - x_0| < \delta_\varepsilon$, meaning that x is closer to x_0 than δ_ε, but not actually equal to x_0.

18.1 Definition. Suppose $S \subseteq \mathbb{R}$, $f : S \to \mathbb{R}$ and $x_0 \in \mathbb{R}$. We say that $f(x)$ converges to a limit L as $x \to x_0$ (through S) if for any $\varepsilon > 0$ there exists $\delta_\varepsilon > 0$ such that if $x \in S$ and $0 < |x - x_0| < \delta_\varepsilon$ then $|f(x) - L| < \varepsilon$. When this happens, we also write $\lim_{x \to x_0} f(x) = L$.

Here we should think of both ε and δ as being very small. The idea is that, however small a tolerance (ε) we specify, if we get close enough (δ_ε) to x_0, but not equal to x_0, $f(x)$ will agree with L, up to the stated tolerance. See Figure 5 for an illustration. The qualifier "through S" is often omitted, if it is clear what the domain should be, but sometimes it is useful to specify it.

18.2 Remark (Localisation). Suppose $f(x) \to L$ as $x \to x_0$ through S.

1. The point x_0 does not have to lie in the set S. Equivalently, $f(x_0)$ need not be defined (just as for sequences, where there is never a sequence term for "$n = \infty$"). But, even if $f(x_0)$ is defined, its value has no effect on the limit.

2. Given some $\varepsilon > 0$, if the definition holds for some particular δ_ε, then it also holds for any smaller δ_ε. This allows us to fix in advance some $r > 0$ and replace δ_ε with $\min\{\delta_\varepsilon, r\}$. In this case, every x considered has $|x - x_0| < r$; values of $f(x)$ for

x with $|x-x_0| \geq r$ are never used. In consequence, values of $f(x)$ where $|x-x_0| \geq r$ have no effect on the limit.

Another way to think of this is to say that the statements "$f(x) \to L$ as $x \to x_0$ through S" and "$f(x) \to L$ as $x \to x_0$ through $S \cap (x_0 - r, x_0 + r)$" are equivalent — as indeed is "$f(x) \to L$ as $x \to x_0$ through $S \cap \big((x_0 - r, x_0) \cup (x_0, x_0 + r)\big)$".

This is analogous to the property that the limit of a sequence is unaffected by any finite number of terms.

3. Another related property: if $L \neq 0$ then we can put $\varepsilon = |L|/2$ in the definition of limit to find $\delta > 0$ such that if $x \in S$ and $0 < |x - x_0| < \delta$ then $|f(x) - L| < |L|/2$. If follows that $|f(x)| > |L|/2$; in particular, $f(x) \neq 0$. This corresponds to Remark 5.13(2) for sequences. In consequence we can, for example, safely divide by $f(x)$, provided $0 < |x_0 - x| < \delta$: see the division property in Theorem 18.6. Taking slightly more care with the sign of L, we can in fact see that if $L > 0$ then $f(x) > 0$ if $x \in S$ and $0 < |x - x_0| < \delta$ and that if $L < 0$ then $f(x) < 0$ if $x \in S$ and $0 < |x - x_0| < \delta$.

18.3 Example. To show that $1/x \to 1/2$ as $x \to 2$ through $\mathbb{R} \setminus \{0\}$, we need to end up with the conclusion $|1/x - 1/2| < \varepsilon$, which is equivalent to

$$\left|\frac{x-2}{2x}\right| < \varepsilon$$

Dealing with this for general x is awkward: it has an absolute value and is singular at 0. In view of the localisation principle, we can decide only to work with x for which $|x-2| < 1$, i.e. $1 < x < 3$. For such x, we can use the estimate $x > 1$ on the bottom line to write

$$\left|\frac{x-2}{2x}\right| < \varepsilon \impliedby \frac{|x-2|}{2} < \varepsilon \iff |x-2| < 2\varepsilon$$

Our carefully presented proof now reads: given $\varepsilon > 0$, let $\delta_\varepsilon = \min\{2\varepsilon, 1\}$. If $0 < |x-2| < \delta_\varepsilon$ then $|x-2| < 1$, so $x > 1$ and $0 < 1/x < 1$. It follows that

$$\left|\frac{1}{x} - \frac{1}{2}\right| = \left|\frac{x-2}{2x}\right| < \frac{|x-2|}{2} < \frac{2\varepsilon}{2} = \varepsilon$$

showing that $1/x \to 1/2$ as $x \to 2$.

18.4 Example (Heaviside function). Define $H : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ by

$$H(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x > 0 \end{cases}$$

Suppose $\varepsilon > 0$. If $x_0 < 0$ then we can choose $\delta = -x_0$; if $|x - x_0| < \delta$ then $2x_0 < x < 0$ so $H(x) = 0$; in particular, $|H(x) - 0| < \varepsilon$. This shows that $H(x) \to 0$ as $x \to x_0$.

Similarly, if $x_0 > 0$ then $\delta = x_0$ works in the definition to show that $H(x) \to 1$ as $x \to x_0$.

If $x_0 = 0$ then for any $\delta > 0$ both $x = \pm\delta/2$ satisfy $|x - 0| < \delta$; but $f(\delta/2) = 1$ and $f(-\delta/2) = 0$. If H had a a limit L at 0, it would have to satisfy both $|L - 0| < \varepsilon$ and $|L - 1| < \varepsilon$, for all $\varepsilon > 0$. This is impossible, so no limit exists at 0.

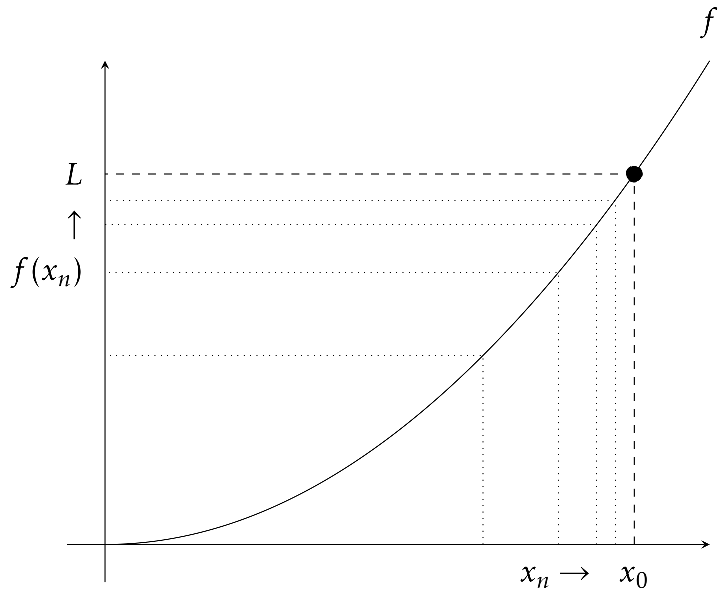


Figure 6: Sequential characterisation of limits. As $x \to x_0$, $f(x) \to L$; as $n \to \infty$, $x_n \to x_0$ and $f(x_n) \to L$.

18.5 Theorem (Sequential characterisation of limits). *Suppose $S \subseteq \mathbb{R}$, $f : S \to \mathbb{R}$ and $x_0, L \in \mathbb{R}$. Then the following are equivalent:*

1. *$f(x) \to L$ as $x \to x_0$ through S;*
2. *for every sequence (x_n) in $S \setminus \{x_0\}$ such that $x_n \to x_0$ as $n \to \infty$, we have $f(x_n) \to L$.*

(see Figure 6 for an illustration).

Proof. (1) $\Longrightarrow$ (2): suppose that $f(x) \to L$ as $x \to x_0$ and that $x_n \to x_0$ as $n \to \infty$, with $x_n \in S \setminus \{x_0\}$. For any $\varepsilon > 0$ there exists δ_ε such that if $x \in S$ and $0 < |x - x_0| < \delta_\varepsilon$ then $|f(x) - L| < \varepsilon$. Also there exists N_ε such that if $n > N_\varepsilon$ then $|x_n - x_0| < \delta_\varepsilon$; we now have, for $n > N_\varepsilon$, $|f(x_n) - L| < \varepsilon$, showing that $f(x_n) \to L$ as $\to \infty$.

(2) $\Longrightarrow$ (1): we use the contrapositive. Suppose $f(x) \nrightarrow L$ as $x \to x_0$. Then for some $\varepsilon > 0$ we cannot find a suitable δ_ε; that is, for any $\delta > 0$ there exists $x \in S$ with $0 < |x - x_0| < \delta$ but $|f(x) - L| \geq \varepsilon$. In particular, for $n \in \mathbb{N}$ we can put $\delta = 1/n$ and find $x_n \in S \setminus \{x_0\}$ with $|x_n - x_0| < 1/n$ and $|f(x_n) - L| \geq \varepsilon$. We now have $x_n \to x_0$ in $S \setminus \{x_0\}$ but $f(x_n) \nrightarrow L$. □

18.6 Theorem (Algebra of Limits). *Suppose $S \subseteq \mathbb{R}$, $f, g : S \to \mathbb{R}$, $x_0 \in \mathbb{R}$, $f(x) \to L$ as $x \to x_0$ and $g(x) \to M$ as $x \to x_0$ (through S)*

1. *$f(x) + g(x) \to L + M$ as $x \to x_0$ (through S)*
2. *$f(x)g(x) \to LM$ as $x \to x_0$ (through S)*
3. *If $M \neq 0$ then $f(x)/g(x) \to L/M$ (through $\{x \in S : g(x) \neq 0\}$)*

Proof. To prove (1): If $(x_n)_{n\in\mathbb{N}}$ is any sequence in $S \setminus \{x_0\}$ such that $x_n \to x_0$ as $n \to \infty$ then $f(x_n) \to L$ and $g(x_n) \to M$ (sequential characterisation, (1) $\Longrightarrow$ (2)) so $f(x_n) + g(x_n) \to L+M$ (algebra of limits for sequences) and hence $f(x)+g(x) \to L+M$ (sequential characterisation, (2) $\Longrightarrow$ (1)).

To prove (2) and (3), we just change addition to multiplication and division. Note that, in (3), because $M \neq 0$, there exists $\delta > 0$ such that $g(x) \neq 0$ provided $0 < |x - x_0| < \delta$; see Remark 18.2. □

18.7 Example. If P is any polynomial, then the algebra of limits for sequences tells us that for any sequence $(x_n)_{n\in\mathbb{N}}$ such that $x_n \to x_0$ as $n \to \infty$, we have $P(x_n) \to P(x_0)$. The sequential characterisation of limits now tells us that $P(x) \to P(x_0)$ as $x \to x_0$.

Alternatively, we could start with the constant functions and the identity function $f(x) = x$, check that their limiting properties work as expected (they do!) and draw the same conclusion about P from Theorem 18.6.

18.8 Remark (Uniqueness of Limits). Suppose $S \subseteq \mathbb{R}$, $f : S \to \mathbb{R}$, $x_0 \in \mathbb{R}$ and, as $x \to x_0$, we have $f(x) \to L$ and $f(x) \to M$. We would expect to find that $L = M$: that limits are unique, as they are for sequences. Indeed, we can offer a very plausible proof of this: if $(x_n)_{n\in\mathbb{N}}$ is some sequence in S converging to x_0, then the sequential characterisation of limits shows that, as $n \to \infty$, $f(x_n) \to L$ and $f(x_n) \to M$; by uniqueness of limits for sequences, $L = M$.

There is, however, a small hole in this argument: it is possible that no such sequence $(x_n)_{n\in\mathbb{N}}$ exists. For example, if $f : [0,1] \to \mathbb{R}$, what is $\lim_{x\to 2} f(x)$? The problem here is that it is not possible for x to approach 2 while remaining in $[0,1]$ and the definition of limit starts to break down. However, as long as such pathological examples are excluded, limits, if they exist, are uniquely defined.

We shall soon restrict attention to sets where this will not be a problem. If we were to continue with general sets, we would have to require, for uniqueness of limits at x_0, that there exists a sequence in $S \setminus \{x_0\}$ converging to x_0; equivalently,that for any $\delta > 0$ there exists $x \in S$ with $0 < |x - x_0| < \delta$. Such a point x_0 is called a *limit point* (or *cluster point* or *point of accumulation*) of S; we shall not need to work explicitly with these.

One-sided limits (from the left or right) are a straightforward modification: in the definition of limit, we can replace "$0 < |x - x_0| < \delta_\varepsilon$" with "$0 < x_0 - x < \delta_\varepsilon$" for a left limit or "$0 < x - x_0 < \delta_\varepsilon$" for a right limit. Equivalently, we can change the domain S:

18.9 Definition (One-sided Limits)**.** Suppose $S \subseteq \mathbb{R}$, $f : S \to \mathbb{R}$ and $x_0 \in \mathbb{R}$. Say that:

1. $f(x)$ converges to a limit L as $x \to x_0$ through S from the left (written $x \to x_0-$) if $f(x) \to L$ as $x \to x_0$ through $S \cap (-\infty, x_0)$. We also write $\lim_{x\to x_0-} f(x) = L$ and $f(x-) = L$.

2. $f(x)$ converges to a limit L as $x \to x_0$ through S from the right (written $x \to x_0+$) if $f(x) \to L$ as $x \to x_0$ through $S \cap (x_0, \infty)$. We also write $\lim_{x\to x_0+} f(x) = L$ and $f(x+) = L$.

To contrast with these, we sometimes call the limit defined in Definition 18.1 a *two-sided* limit.

18.10 Remark. The virtue of this way of expressing one-sided limits is that the sequential characterisation of one-sided limits follows immediately from Theorem 18.5: just restrict to sequences $x_n \in S$ with $x_n \to x_0$ and either $x_n < x_0$ for a left limit or $x_n > x_0$ for a right limit. The Algebra of Limits then works in exactly the same way for one-sided limits.

18.11 Theorem. *Suppose $S \subseteq \mathbb{R}$, $f : S \to \mathbb{R}$, $x_0 \in \mathbb{R}$, $L \in \mathbb{R}$. Then the following are equivalent:*

1. $f(x) \to L$ *as* $x \to x_0$
2. $f(x) \to L$ *as* $x \to x_0-$ *and* $f(x) \to L$ *as* $x \to x_0+$

Proof.

(1) $\Longrightarrow$ (2) Given $\varepsilon > 0$ there exists $\delta_\varepsilon > 0$ such that if $x \in S$ and $0 < |x_0 - x| < \varepsilon$ then $|f(x) - L| < \varepsilon$. Equivalently, if $0 < x_0 - x < \delta_\varepsilon$ or $0 < x - x_0 < \delta_\varepsilon$ then $|f(x) - L| < \varepsilon$; the first of these shows that $f(x) \to L$ as $x \to x_0-$ and the second that $f(x) \to L$ as $x \to x_0+$

(2) $\Longrightarrow$ (1) Given $\varepsilon > 0$ there exist δ_ε^- and δ_ε^+ such that, for $x \in S$, if $0 < x_0 - x < \delta_\varepsilon^-$ then $|f(x) - L| < \varepsilon$ and if $0 < x - x_0 < \delta_\varepsilon^+$ then $|f(x) - L| < \varepsilon$. Let $\delta_\varepsilon = \min\{\delta_\varepsilon^-, \delta_\varepsilon^+\}$. Now, if $x \in S$ and $|x - x_0| < \delta_\varepsilon$ then $|f(x) - L| < \varepsilon$, showing that $f(x) \to L$ as $x \to x_0$.

□

18.12 Definition (Infinite limits, limits at $\pm\infty$)**.** There are many other variations on the basic definition, most of which we shall not need in this course. Suppose $S \subseteq \mathbb{R}$ and $f : S \to \mathbb{R}$.

For infinite limits, say that e.g. $f(x) \to \infty$ as $x \to x_0$ if for any $K \in \mathbb{R}$ there exists $\delta_K > 0$ such that if $x \in S$ and $0 < |x - x_0| < \delta_K$ then $f(x) > K$. Here we should think of K as large and positive and of δ_K as small. This can be modified in the obvious way for $f(x) \to -\infty$, and for one-sided limits.

For limits at $\pm\infty$, we can adapt the definition for sequences. Say $f(x) \to L$ as $x \to \infty$ if for any $\varepsilon > 0$ there exists $X_\varepsilon \in \mathbb{R}$ such that if $x \in S$ and $x > X_\varepsilon$ then $|f(x) - L| < \varepsilon$. Replace "$x > X_\varepsilon$" with"$x < X_\varepsilon$" for a limit at $-\infty$. This can also be modified in an obvious way for infinite limits at $\pm\infty$.

We can enumerate all the possible limit definitions, based on the idea that the definition of a limit comes in four pieces. In the definition of $f(x) \to L$ as $x \to x_0$ ($x_0, L \in \mathbb{R}$), these are:

1. For any $\varepsilon > 0$
2. there exists $\delta > 0$
3. such that if $x \in S$ and $0 < |x - x_0| < \delta$
4. then $|f(x) - L| < \varepsilon$

In all the variations of the definition, the same four pieces are visible but they change slightly. Here are all the possibilities. It is important to be clear on terminology: a limit "at x_0" means that $x \to x_0$, whereas a limit "is L" means that $f(x) \to L$.

1. (a) Limit is $L \in \mathbb{R}$ (finite limit): $\underline{\text{For any } \varepsilon > 0}$
   (b) Limit is $\pm\infty$: $\underline{\text{For any } K \in \mathbb{R}}$
2. (a) Limit at $x_0 \in \mathbb{R}$: $\underline{\text{there exists } \delta > 0}$
   (b) Limit at $\pm\infty$: $\underline{\text{there exists } X \in \mathbb{R}}$
3. $\underline{\text{such that if } x \in S \text{ and}}$
   (a) (Two-sided) limit at $x_0 \in \mathbb{R}$: $\underline{0 < |x_0 - x| < \delta}$
   (b) Left limit at $x_0 \in \mathbb{R}$: $\underline{0 < x_0 - x < \delta}$
   (c) Right limit at $x_0 \in \mathbb{R}$: $\underline{0 < x - x_0 < \delta}$
   (d) Limit at $-\infty$: $\underline{x < X}$
   (e) Limit at $+\infty$: $\underline{x > X}$
4. (a) Limit is $L \in \mathbb{R}$ (finite limit): $\underline{\text{then } |f(x) - L| < \varepsilon}$
   (b) Limit is $-\infty$: $\underline{\text{then } f(x) < K}$
   (c) Limit is $+\infty$: $\underline{\text{then } f(x) > K}$

Notation like δ_ε, emphasising the fact that δ depends on ε, has been suppressed because δ and X can depend on either ε or K.

19 Continuous functions on intervals

We now abandon the generality of functions defined on arbitrary sets and concentrate on the most important example: intervals.

19.1 Definition. Suppose $I \subseteq \mathbb{R}$ is an interval, $f : I \to \mathbb{R}$ and $x_0 \in I$. We say that f is *continuous at* x_0 if $f(x) \to f(x_0)$ as $x \to x_0$. We say that f is *continuous on* I if f is continuous at every point $x_0 \in I$.

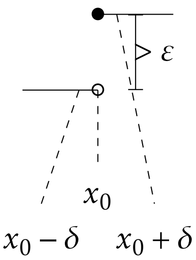
In the definition of limit, the exclusion $0 < |x - x_0|$ is no longer needed (because it excludes only the trivially true statement $|f(x_0) - f(x_0)| < \varepsilon$). We thus have:

f is continuous at x_0 if for any $\varepsilon > 0$ there exists $\delta_\varepsilon > 0$ such that if $x \in I$ and $|x - x_0| < \delta_\varepsilon$ then $|f(x) - f(x_0)| < \varepsilon$.

In sequential terms, f is continuous at x_0 if and only if for every sequence $(x_n)_{n\in\mathbb{N}}$ in I such that $x_n \to x_0$ as $n \to \infty$, we have $f(x_n) \to f(x_0)$.

It is occasionally useful to use one-sided limits: we say that f is *right-continuous* or *continuous from the right* at x_0 if $f(x) \to f(x_0)$ as $x \to x_0+$. Continuity from the left and one-sided continuity on an interval are then defined in the obvious way.

The idea is that if there is any jump or gap in the graph of the function then, for sufficiently small ε, the definition will fail (i.e. no such δ will exist). In the case of a simple jump discontinuity, this will happen when ε is smaller than the height of the jump.



19.2 Example. We can now look back at Example 18.7 which says that, for any polynomial P and any $x_0 \in \mathbb{R}$, $P(x) \to P(x_0)$ as $x \to x_0$. In the new terminology, this says that every polynomial is a continuous function on $\mathbb{R}$ (and any smaller interval).

We can also use the division property from the algebra of limits: if P and Q are two polynomials and I is an interval on which Q is never zero, then P/Q defines a continuous function on I.

We saw in Theorem 7.3 and Exercise 5.1 that if $x_0 \geq 0$, $x_n \geq 0$ and $x_n \to x_0$ as $n \to \infty$, then $\sqrt{x_n} \to \sqrt{x_0}$ as $n \to \infty$. In the new terminology, $\sqrt{\cdot}$ is continuous on $[0, \infty)$.

We saw in Example 18.4 that if we define

$$H(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x > 0 \end{cases}$$

then H has no limit at 0. This means that we cannot extend H to be a continuous function from $\mathbb{R} \to \mathbb{R}$: however we define $H(0)$, 0 will always be a point of discontinuity. However, if we were to define $H(0) = 0$ then H would become continuous from the left (but not the right) at 0 and if we were to define $H(0) = 1$ then H would become continuous from the right (but not the left) at 0.

19.3 Notation. If f and g are two real-valued functions with the same domain, we define functions $f + g$, $f - g$, fg and f/g in the natural way:

$$(f + g)(x) = f(x) + g(x)$$

etc. If g is zero at any point, the domain of f/g will have to be reduced.

The following is an immediate consequence of Theorem 18.6:

19.4 Theorem (Algebra of Continuous Functions)**.** *Suppose $I \subseteq \mathbb{R}$ is an interval and $f, g : I \to \mathbb{R}$ is continuous. Then $f + g$, $f - g$, fg and, provided $g(x) \neq 0$ for $x \in I$, f/g are all continuous.*

We also have another way to combine continuous functions: composition.

19.5 Theorem (Composition of Continuous Functions)**.** *Suppose $I, J \subseteq \mathbb{R}$ are intervals and that $f : J \to \mathbb{R}$ and $g : I \to J$ are continuous functions. Then their composition $f \circ g : I \to \mathbb{R}$ defined by $(f \circ g)(x) = f(g(x))$ is continuous.*

Proof. Fix x_0 in I. If $(x_n)_{n \in \mathbb{N}}$ is any sequence in I with limit x_0 then, as $n \to \infty$, $g(x_n)$ converges to $g(x_0)$ in J and $f(g(x_n)) \to f(g(x_0))$, showing that $f \circ g$ is continuous at x_0. Since x_0 was an arbitrary element of I, $f \circ g$ is continuous on I.

□

19.6 Theorem. *Suppose f is defined by a power series*

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

which converges on $(x_0 - R, x_0 + R)$. Then f is continuous on $(x_0 - R, x_0 + R)$

⚠ The proof of this result is more demanding than most in the course; you might prefer to skip the details at first reading.

Proof. We shall work with $x_0 = 0$ to simplify notation; for the general case, replace x by $x - x_0$ throughout.

Suppose $x^* \in (-R, R)$; we shall prove continuity at x^*. For $N \in \mathbb{N}_0$ and $x \in (-R, R)$, let

$$f_N(x) = \sum_{n=0}^{N} a_n x^n$$

We need an estimate on how quickly the series converges. Convergence can fail at the endpoints, when $x = \pm R$, so we need to stay away from them; to this end, fix r such that $|x^*| < r < R$ and suppose $|x| < r$. Then

$$\left.\begin{aligned} |f(x) - f_N(x)| &= \left| \sum_{n=N+1}^{\infty} a_n x^n \right| \\ &\leq \sum_{n=N+1}^{\infty} |a_n||x|^n \\ &< \sum_{n=N+1}^{\infty} |a_n| r^n \\ &=: c_N \end{aligned}\right\} \qquad (|x| < r) \tag{*}$$

and we know that $c_N \to 0$ as $N \to \infty$ because the series $\sum_{n=0}^{N} a_n r^n$ is absolutely convergent (Theorem 14.4).

Now we prove continuity at x^*. Given $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $c_N < \varepsilon/3$; since $|x^*| < r$, we have $|f(x^*) - f_N(x^*)| < c_N < \varepsilon/3$ by (*). Because f_N is a polynomial, it is continuous, so there exists $\delta_\varepsilon \in (0, r - |x^*|)$ such that if $|x^* - x| < \delta_\varepsilon$ then $|f_N(x^*) - f_N(x)| < \varepsilon/3$. The choice of δ_ε smaller than $r - |x^*|$ implies that if $|x^* - x| < \delta_\varepsilon$ then $|x| < r$, so $|f(x) - f_N(x)| < c_N < \varepsilon/3$ by (*). Finally, we estimate that if $|x - x^*| < \delta_\varepsilon$ then

$$\begin{aligned} |f(x^*) - f(x)| &= |f(x^*) - f_N(x^*) + f_N(x^*) - f_N(x) + f_N(x) - f(x)| \\ &\leq |f(x^*) - f_N(x^*)| + |f_N(x^*) - f_N(x)| + |f_N(x) - f(x)| \\ &< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon \end{aligned}$$

showing that f is continuous at x^*. □

19.7 Example. The functions sin, cos and exp were defined in Definition 15.1 by power series with infinite radius of convergence; they are therefore continuous on $\mathbb{R}$.

We now reach one of the most important properties of continuous functions.

19.8 Theorem (Bolzano's Theorem)**.** *Suppose $a < b$, $g : [a,b] \to \mathbb{R}$ is continuous and $g(a)$ and $g(b)$ have opposite signs (i.e. either $g(a) < 0$ and $g(b) > 0$ or $g(a) > 0$ and $g(b) < 0$). Then there exists $x_0 \in (a,b)$ such that $g(x_0) = 0$.*

Proof. We shall prove this in the case $g(a) < 0$ and $g(b) > 0$. The other case follows by applying this result to $-g$.

Because $g(b) > 0$ and g is continuous, there is an interval $[b', b]$ on which g is positive (Remark 18.2). Let

$$S = \{x \in [a,b] : g(x) \leq 0\}$$

Then $a \in S$ and S is bounded above by b'. Let $x_0 = \sup(S)$, so $a \leq x_0 \leq b' < b$.

Now, we construct two sequences. Firstly, for any $n \in \mathbb{N}$, there exists $v_n \in [a,b]$ such that $x_0 < v_n < x_0 + 1/n$. Then $v_n \to x_0$ and $v_n \notin S$ so $g(v_n) > 0$; letting $n \to \infty$, we have $g(x_0) \geq 0$.

Secondly, for any $n \in \mathbb{N}$, there exists $u_n \in S$ with $u_n > x_0 - 1/n$; now we have $u_n \to x_0$ and $g(u_n) \leq 0$; letting $n \to \infty$, $g(x_0) \leq 0$. Combine these to give $g(x_0) = 0$. □

19.9 Corollary (Intermediate Value Theorem (IVT))**.** *Suppose I is an interval, $f : I \to \mathbb{R}$ is continuous and $a, b \in I$ with $a < b$. If y is any number such that $f(a) \leq y \leq f(b)$ or $f(b) \leq y \leq f(a)$ then there exists $x \in [a,b]$ such that $f(x) = y$. In other words (Theorem 3.17), the range (sometimes called image) of f, $f(I) = \{f(x) : x \in I\}$, is an interval.*

Proof. If $f(a) = y$ or $f(b) = y$, the result is true with $x = a$ or $x = b$. If not, then we can apply the previous result on $[a,b]$ with $g(x) = f(x) - y$ to find x such that $f(x) - y = 0$, i.e. $f(x) = y$. □

The IVT can be thought of as solving the equation $f(x) = y$ (which is why it is important). However, there can be many solutions to the equation. We can impose extra conditions on f to make the solution unique, in which case we end up constructing an inverse function.

19.10 Definition. Suppose $I \subseteq \mathbb{R}$ is an interval and $f : I \to \mathbb{R}$. We say that f is *increasing* on I if (for $x, y \in I$) $x \leq y$ implies $f(x) \leq f(y)$. We say that f is *strictly increasing* on I if (for $x, y \in I$) $x < y$ implies $f(x) < f(y)$.

The *decreasing* and *strictly decreasing* properties are defined in the obvious way. The terms *monotonic* and *strictly monotonic* mean "either increasing or decreasing" and "either strictly increasing or strictly decreasing".

19.11 Theorem (Inverse Function Theorem for Continuous Functions)**.** *Suppose I is an interval and $f : I \to \mathbb{R}$ is continuous and strictly increasing. Let $J = f(I)$ (so J is also an interval by the IVT, Corollary 19.9). Then $f : I \to J$ is invertible and $f^{-1} : J \to I$ is continuous. The same is true if f is strictly decreasing.*

Proof. We shall prove the result for increasing functions only; for a decreasing f, e.g. apply the increasing result to $-f$.

By definition of J, f is surjective onto J. We can also see that f is injective, because if $x_1, x_2 \in I$ with $x_1 \neq x_2$ then either $x_1 < x_2$ and $f(x_1) < f(x_2)$ or $x_1 > x_2$ and $f(x_1) > f(x_2)$;

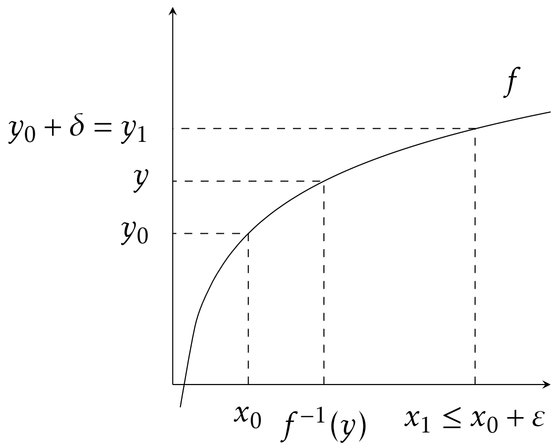


Figure 7: Inequalities in the proof of the Inverse Function Theorem

in either case, $f(x_1) \neq f(x_2)$. We now know that f is bijective, so it has an inverse $f^{-1} : J \to I$.

We shall show that f^{-1} is continuous from the right; continuity from the left follows from a similar argument, and the two combine to show that f^{-1} is continuous.

Suppose $y_0 \in J$ and $\varepsilon > 0$. If y_0 is the maximum value in J, there is nothing to do. Otherwise, let $x_0 = f^{-1}(y_0)$ and note that x_0 cannot be the maximum value of I (otherwise $y_0 = f(x_0)$ would be the maximum value of J). We can therefore choose $x_1 \in I$ with $x_1 > x_0$ and $x_1 \le x_0 + \varepsilon$. Let $y_1 = f(x_1)$ and $\delta = y_1 - y_0$.

Now, if $0 < y - y_0 < \delta$ then $y_0 < y < y_0 + \delta = y_1$; applying the strictly increasing function f^{-1}, we obtain $f^{-1}(y_0) < f^{-1}(y) < f^{-1}(y_1)$; but $f^{-1}(y_1) = x_1 \le x_0 + \varepsilon = f^{-1}(y_0) + \varepsilon$, so we have $f^{-1}(y_0) < f^{-1}(y) < f^{-1}(y_0) + \varepsilon$ and in particular $|f^{-1}(y_0) - f^{-1}(y)| < \varepsilon$. This shows that $f^{-1}(y_0)$ is the right limit of f^{-1} at y_0, i.e. f^{-1} is continuous from the right at y_0. $\square$

19.12 Example. All qth root functions (defined as inverses of qth power functions) are continuous on $[0, \infty)$; compare Theorem 7.3). We know that exp, defined by a power series, is continuous; it is tempting to use the Inverse Function Theorem to define log as its inverse, but we cannot do this yet: we do not have enough information to show that exp is strictly increasing, or that the range of exp is the interval $(0, \infty)$.

19.13 Theorem (Extreme Value Theorem)**.** *Suppose $a \le b$ and $f : [a,b] \to \mathbb{R}$. Then $f([a,b])$ is a closed, bounded interval; equivalently, there exist $x_{\min}$ and $x_{\max}$ such that for all $x \in [a,b]$ we have*

$$f(x_{\min}) \le f(x) \le f(x_{\max})$$

Proof. If $f([a,b])$ were unbounded above, then for each $n \in \mathbb{N}$ we could find x_n such that $f(x_n) \ge n$; this $(x_n)_{n\in\mathbb{N}}$ would have a convergent subsequence $(x_{n_k})_{k\in\mathbb{N}}$ with limit $x \in [a,b]$ and we would have $f(x_{n_k}) \to f(x)$ as $k \to \infty$, contradicting $f(x_{n_k}) \ge n_k \to \infty$.

$f([a,b])$ must therefore be bounded above. As such, it has a supremum

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

Now, for any $n \in \mathbb{N}$ there must exist $x_n \in [a,b]$ such that $f(x_n) > M - 1/n$ (otherwise $M - 1/n$ would be an upper bound for $f([a,b])$, smaller than the least upper bound M). This $(x_n)_{n\in\mathbb{N}}$ has a convergent subsequence $(x_{n_k})_{k\in\mathbb{N}}$ with limit $x \in [a,b]$. Because f is continuous, $f(x_{n_k}) \to f(x)$ as $k \to \infty$; but $M - 1/n_k < f(x_{n_k}) \leq M$ so, by the Sandwich Theorem, $f(x_{n_k}) \to M$. We now have $f(x) = M$ and can let $x_{\max} = x$.

We can construct $x_{\min}$ in an analogous way, or apply the above to $-f$. □

19.14 Remark. This does not work for other kinds of interval! Consider e.g. $f(x) = 1/x$ on $(0,1)$. This is unbounded above; it is bounded below with infimum 1 but there is no $x \in (0,1)$ with $f(x) = 1$.

20 Part IV: Key Points

The idea of a limit of a sequence at infinity can be adapted to a limit at a point x_0 by replacing "$n > N_\varepsilon$", meaning that n is very large, with "$0 < |x - x_0| < \delta_\varepsilon$", meaning that x is very close to x_0 (Definition 18.1).

The value of a limit at x_0 depends only on values near but not equal to x_0: the values of $f(x_0)$ and of $f(x)$ where $|x - x_0| \geq r > 0$ have no effect on the limit (this is analogous to the fact that the limit of a sequence is not affected by any finite collection of terms) (Remark 18.2).

Limits at a point can be expressed in terms of limits of sequences: $f(x) \to L$ as $x \to x_0$ is equivalent to $f(x_n) \to L$ for every sequence (x_n) converging to x_0 but never equalling x_0 (Theorem 18.5). The Algebra of Limits then carries over almost unchanged from the sequential case (Theorem 18.6).

A function f on an interval I is said to be continuous at x_0 if $f(x_0) = \lim_{x \to x_0} f(x)$, and continuous on I if it is continuous at all $x_0 \in I$ (Definition 19.1). Continuity, unlike the idea of a limit, is a global property.

This can be recast in sequential form: f is continuous on I if whenever $x_n \to x$ as $n \to \infty$ in I, we also have $f(x_n) \to f(x)$ as $n \to \infty$. In this form, we can interpret some earlier results about sequences as expressions of continuity: e.g. polynomials and the square root are continuous functions (Example 18.7, Example 19.2). We can also see that any function described by a convergent power series is continuous, in particular the elementary functions (sin, cos, exp) defined earlier.

We can also interpret the Algebra of Limits in terms of continuity: sums, differences, products and (if we avoid division by zero), ratios of continuous functions are continuous (Theorem 19.4). We also have a new combination rule: compositions of continuous functions are continuous (Theorem 19.5).

Any continuous function on an interval has the Intermediate Value property: if it attains any two values, it also attains every value between them (Corollary 19.9). This can be used to infer the existence of solutions to equations, but is most useful when combined with strict monotonicity: a strictly monotonic continuous function on an interval has a continuous, strictly monotonic inverse (Theorem 19.11). This can be used to see that all the qth root functions are continuous.

Finally, any continuous function on a closed, bounded interval $[a, b]$ is bounded and takes on both maximal and minimal values on that interval (Theorem 19.13). We haven't seen any applications of this yet, but it's somewhat analogous to the fact that every convergent sequence is bounded, as a source of cheap upper bounds.

Literature:

1. F. Mary Hart *Guide to Analysis* (2nd Edition, Palgrave Macmillan 2001, JBM classmark S7 Har) Chapter 4.

2. any other introduction to Real Analysis; browse classmark S7 in the library.

Part V
Differentiation

Much of this section has already been covered in Calculus and will be just briefly reprised here. The main differences are:

- We now have a rigorous definition of a limit, so the definition of derivative is on more secure footings.
- Rolle's Theorem, the foundation stone of Taylor's Theorem and Taylor series, can be properly proved (based on the Extreme Value Theoerem, Theorem 19.13).
- The elementary functions have been defined in a more constructive way, so their properties can be more rigorously established.

21 The Derivative

21.1 Definition. Suppose $I \subseteq \mathbb{R}$ is an interval, $f : I \to \mathbb{R}$ and $x_0 \in I$. We say that f is *differentiable* at $x_0 \in I$ if the limit

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

exists. In that case, we say that the *derivative* of f at x_0, denoted $f'(x_0)$, is that limit. We say that f is *differentiable on I* if f is differentiable at each point $x_0 \in I$. We say that f is *continuously differentiable* on I if, in addition, $f' : I \to \mathbb{R}$ is continuous. Note that, implicitly, $x \to x_0$ through I; at an endpoint of I contained in I, the limit is effectively one-sided.

In this definition, it is often convenient to change notation from x to $x_0 + h$ and write

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

Note that h can be positive or negative.

21.2 Example. If f is constant on an interval I, say $f(x) = c$, then the difference quotients are just

$$\frac{c - c}{x - x_0} = 0$$

so f is plainly differentiable and $f'(x_0) = 0$.

Now fix $n \in \mathbb{N}$ and define $f : I \to \mathbb{R}$ by $f(x) = x^n$. We can differentiate f using the difference quotient in the form

$$\frac{f(x_0 + h) - f(x_0)}{h}$$

and the binomial expansion

$$(x_0+h)^n = \sum_{k=0}^{n} \binom{n}{k} x_0^{n-k} h^k$$

This gives

$$\frac{f(x_0+h)-f(x_0)}{h} = n x_0^{n-1} + \sum_{k=2}^{n} \binom{n}{k} x_0^{n-k} h^{k-1}$$

All the terms in the sum are multiples of a positive power of h and therefore tend to zero as $h \to 0$; the RHS therefore tends to $n x_0^{n-1}$ as $h \to 0$. We conclude that

$$f'(x_0) = n x_0^{n-1}$$

Especially for more general results, the following characterisation is often useful (see note after theorem for a geometrical interpretation of p and Figure 8 for an illustration).

21.3 Theorem. *Suppose $I \subseteq \mathbb{R}$ is an interval, $f : I \to \mathbb{R}$ and $x_0 \in I$. Then the following are equivalent:*

1. *f is differentiable at x_0, with $f'(x_0) = d$*
2. *We can write*

$$f(x_0+h) = f(x_0) + hp(h) \tag{*}$$

for h such that $x_0 + h \in I$, where the function p is continuous at 0 and satisfies $p(0) = d$ (equivalently, $p(0) = d$ and $p(h) \to d$ as $h \to 0$).

Proof. (1) $\Longrightarrow$ (2): let

$$p(h) = \begin{cases} \dfrac{f(x_0+h)-f(x_0)}{h} & \text{if } x_0+h \in I \text{ and } h \neq 0 \\ f'(x_0) \quad [= d] & \text{if } h = 0 \end{cases}$$

Then (*) is true by a simple rearrangement. Also, $p(h) \to f'(x_0) = d = p(0)$ as $h \to 0$ by definition of the derivative, so p is continuous at 0. (2) $\Longrightarrow$ (1): we have

$$\frac{f(x_0+h)-f(x_0)}{h} = p(h) \to p(0) = d$$

as $h \to 0$, using continuity of p at 0 and $p(0) = d$; by definition of the derivative, f is differentiable at x_0 with $f'(x_0) = d$. □

The geometrical point of this (Figure 8) is that $p(h)$ is the gradient of the chord from $(x_0, f(x_0))$ to $(x_0+h, f(x_0+h))$ for non-zero h, and the gradient of the tangent at x_0 for $h = 0$.

21.4 Theorem. *Suppose $I \subseteq \mathbb{R}$ is an interval and $f : I \to \mathbb{R}$ is differentiable. Then f is continuous. The converse is not true: there exist continuous functions that are not differentiable.*

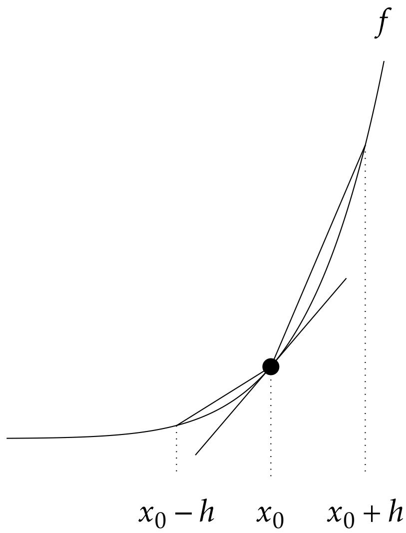


Figure 8: The function p in Theorem 21.3: $p(-h)$ and $p(h)$ are the gradients of the chords shown; $p(0)$ is the gradient of the tangent to f at x_0.

Proof. If $x_0 \in I$ then f is differentiable at x_0, so

$$f(x_0+h) = f(x_0) + hp(h) \tag{†}$$

where $p(h) \to f'(x_0)$ as $h \to 0$; we can now let $h \to 0$ in (†) to give $f(x_0+h) \to f(x_0)$ as $h \to 0$, showing that f is continuous at x_0. Since x_0 was an arbitrary element of I, f is continuous on I.

For an example of a continuous function that is not differentiable, we can use $f(x) = |x|$ on $\mathbb{R}$. The reverse triangle inequality

$$||x| - |x_0|| \leq |x - x_0|$$

shows that f is continuous ($\delta_\varepsilon = \varepsilon$).

But if we write down the difference quotient at $x_0 = 0$, we get

$$\frac{f(x)-f(0)}{x-0} = \frac{|x|}{x} = \begin{cases} 1 & \text{if } x > 0 \\ -1 & \text{if } x < 0 \end{cases}$$

which has no limit at the origin, because its left and right limits (-1 and 1, respectively) are not equal; f is thus not differentiable at 0.

Note that, because of localisation, we can also say that

$$f(x) = x \text{ on } (0,\infty) \implies f'(x) = 1 \text{ on } (0,\infty)$$

$$f(x) = -x \text{ on } (0,\infty) \implies f'(x) = -1 \text{ on } (-\infty,0)$$

It is only at 0, where the two definitions collide, that we have to check the difference quotient.

□

22 Rules of Differentiation

22.1 Theorem. *Suppose $I \subseteq \mathbb{R}$ is an interval and $f, g : I \to \mathbb{R}$ are differentiable. Then:*

1. *$f + g$ is differentiable on I and $(f+g)' = f' + g'$*
2. *fg is differentiable on I and $(fg)' = fg' + f'g$ (product rule)*
3. *if $g(x) \neq 0$ for $x \in I$ then f/g is differentiable on I and*

$$\left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2} \qquad \textit{(quotient rule)}$$

Proof. For some fixed, but arbitrary, $x \in I$, and h such that $x + h \in I$, write

$$\begin{aligned} f(x+h) &= f(x) + hp(h) \qquad & p(0) &= f'(x) \\ g(x+h) &= g(x) + hq(h) \qquad & q(0) &= g'(x) \end{aligned}$$

so p, q are continuous at 0.

1. Calculate:

$$\begin{aligned} (f+g)(x+h) &= f(x+h) + g(x+h) \\ &= f(x) + hp(h) + g(x) + hq(h) \\ &= (f+g)(x) + h[\underline{p(h) + q(h)}] \end{aligned}$$

   Now, because p and q are continuous at 0, so is the underlined term $p+q$; we can find $(f+g)'(x)$ by substituting $h = 0$, giving

$$(f+g)'(x) = p(0) + q(0) = f'(x) + g'(x)$$

2. Calculate:

$$\begin{aligned} (fg)(x+h) &= f(x+h)g(x+h) \\ &= [f(x) + hp(h)][g(x) + hq(h)] \\ &= (fg)(x) + h[\underline{f(x)q(h) + g(x)p(h) + hp(h)q(h)}] \end{aligned}$$

   Again because p and q are continuous at 0, the underlined term is continuous at $h = 0$ and we can find $(fg)'(x)$ by substituting $h = 0$:

$$(fg)'(x) = f(x)q(0) + g(x)p(0) = f(x)g'(x) + g(x)f'(x).$$

3. Calculate initially:

$$\frac{f(x+h)}{g(x+h)} - \frac{f(x)}{g(x)} = \frac{f(x) + hp(h)}{g(x) + hq(h)} - \frac{f(x)}{g(x)} = \frac{h[g(x)p(h) - f(x)q(h)]}{(g(x) + hq(x))g(x)}$$

   Now rearrange this to give

$$\frac{f(x+h)}{g(x+h)} = \frac{f(x)}{g(x)} + h\underline{\frac{g(x)p(h) - f(x)q(h)}{(g(x) + hq(x))g(x)}}$$

Just as in the previous two cases, we can substitute $h = 0$ in the underlined (continuous at $h = 0$) term to give

$$\left(\frac{f}{g}\right)'(x) = \frac{g(x)p(0) - f(x)q(0)}{g(x)^2} = \frac{g(x)f'(x) - f(x)g'(x)}{g(x)^2}$$

Alternatively, one can show that $(1/g)' = -g'/g^2$, which is a slightly simpler calculation on the same lines, then use the product rule. Even more alternatively, one can derive $(1/g)' = -g'/g^2$ from the fact that $\mathrm{d}/\mathrm{d}x 1/x = -1/x^2$ (see next example) and the Chain Rule (Theorem 22.3 below).

□

22.2 Example. We can now see, starting with constant functions and the identity function $f(x) = x$, that every polynomial is differentiable, and every rational function with non-zero denominator is differentiable, and use familiar techniques to find these derivatives. In particular, we can start with the fact that if $n \in \mathbb{N}$ and $f(x) = x^n$ then $f'(x) = nx^{n-1}$ (Example 21.2) and extend this to negative powers: in Leibniz notation,

$$\frac{\mathrm{d}}{\mathrm{d}x}x^{-n} = \frac{\mathrm{d}}{\mathrm{d}x}\frac{1}{x^n} = \frac{x^n \cdot 0 - nx^{n-1} \cdot 1}{x^{2n}} = -nx^{-n-1}$$

provided, of course, $x \neq 0$.

One more important result remains:

22.3 Theorem (Chain Rule)**.** *Suppose I and J are intervals and $f : J \to \mathbb{R}$ and $g : I \to J$ are differentiable. Then $f \circ g : I \to \mathbb{R}$ is differentiable and*

$$(f \circ g)'(x) = f'(g(x))g'(x)$$

for each $x \in I$.

Proof. Write, for h and k such that $x + h \in I$ and $g(x) + k \in J$,

$$\begin{aligned} g(x+h) &= g(x) + hq(h) & \qquad q(0) &= g'(x) \\ f(g(x)+k) &= f(g(x)) + kp(k) & \qquad p(0) &= f'(g(x)) \end{aligned}$$

so p, q are continuous at 0 and hence everywhere on their domains (because e.g. for $h \neq 0$ we have $q(h) = (g(x+h) - g(x))/h$; this describes a continuous function by the algebra of limits, Theorem 19.4, using the fact that g is continuous, because it is differentiable, Theorem 21.4).

$$\begin{aligned} (f \circ g)(x+h) &= f(g(x+h)) \\ &= f(g(x) + hq(h)) \\ &= f(g(x)) + h\underline{q(h)p(hq(h))} \end{aligned}$$

Now, as in the arguments above, we note that the underlined term is continuous (using the algebra of limits, Theorem 19.4, and the composition property, Theorem 19.5, for continuous functions) and substitute $h = 0$ to find the derivative, in this case of $f \circ g$:

$$(f \circ g)'(x) = q(0)p(0) = f'(g(x))g'(x)$$

□

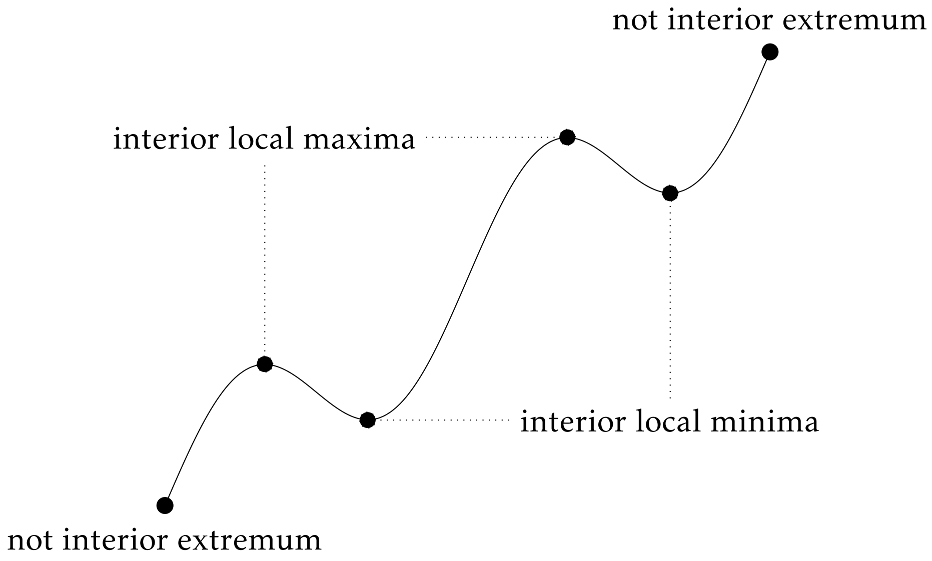


Figure 9: Interior local extrema

23 Monotonicity and extrema

23.1 Definition. Suppose $I \subseteq \mathbb{R}$ is an interval and $f : I \to \mathbb{R}$. A point $x_0 \in I$ is called an *interior local maximum* of f if there exists $\delta > 0$ such that if $|x - x_0| < \delta$ then $x \in I$ and $f(x) \le f(x_0)$. Replace "$f(x) \le f(x_0)$" with "$f(x) \ge f(x_0)$" for the definition of a *interior local minimum*. An *interior local extremum* means either one of the above two cases. The interior condition here, that $x \in I$ if $|x_0 - x| < \delta$, rules out the endpoints of the interval (if it has any). See Figure 9 for an illustration.

23.2 Theorem. *Suppose $I \subseteq \mathbb{R}$ is an interval and $f : I \to \mathbb{R}$ is differentiable. If f is increasing on I then $f' \ge 0$ on I; if f is decreasing on I then $f' \le 0$ on I; if x is an interior local extremum then $f'(x) = 0$.*

Proof. If f is increasing then, for $x, x+h \in I$, we have

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

and we can let $h \to 0$ to give $f'(x) \ge 0$. The decreasing case is similar. If x is an interior local maximum then there exists $\delta > 0$ such that if $|h| < \delta$ then $x + h \in I$ and we have

$$\frac{f(x+h) - f(x)}{h} \begin{cases} \le 0 & \text{if } h > 0 \\ \ge 0 & \text{if } h < 0 \end{cases}$$

Now, if we let $h \to 0$ from the right, we get $f'(x) \le 0$; if we let $h \to 0$ from the left, we get $f'(x) \ge 0$; we conclude that $f'(x) = 0$. A similar argument works for an interior local minimum, with the inequalities reversed.

Note that the interior property is important here, as it ensures that we have $x+h \in I$ with both positive and negative h. An extreme point at an endpoint need not have zero

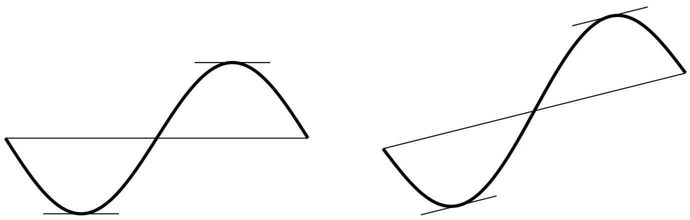


Figure 10: Rolle's Theorem (left) and the Mean Value Theorem (right): tangents are parallel to the chord joining the endpoints.

derivative; e.g. $f(x) = x$ on $[a,b]$ has a minimum at a and a maximum at b but never has derivative zero. □

23.3 Example. If we define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^3$, then f is strictly increasing and its derivative, $f'(x) = 3x^2$ is non-negative, as expected. Note though that $f'(0) = 0$. This shows that:

- Strictly monotonic functions need not have strictly single-signed derivatives;
- the derivative can be zero at an interior point that is not an extremum.

Going in the opposite direction, inferring information about monotonicity and extreme values from information about derivatives, requires more sophisticated machinery, which we shall now develop. The following two results are illustrated in Figure 10

23.4 Theorem (Rolle's Theorem)**.** *Suppose $a < b$ and $g : [a,b] \to \mathbb{R}$ is a continuous function, differentiable on (a,b), and such that $g(a) = g(b)$. Then there exists $x_0 \in (a,b)$ such that $g'(x_0) = 0$.*

Proof. By the Extreme Value Theorem (Theorem 19.13), we can find $x_{\min}$ and $x_{\max} \in [a,b]$ such that for all $x \in [a,b]$,

$$g(x_{\min}) \le g(x) \le g(x_{\max})$$

If both $x_{\min}$ and $x_{\max}$ are endpoints (a or b) then, because $g(a) = g(b)$, g is constant and we have $g'(x) = 0$ for all $x \in (a,b)$.

If not, then one of $x_{\min}$ and $x_{\max}$ is an interior local extremum, at which the derivative of f is zero. □

We can generalise this to functions with non-equal endpoint values by first adding a linear function, chosen so the endpoint values of the sum are equal. This leads to

23.5 Theorem (Mean Value Theorem (MVT))**.** *Suppose $a < b$ and $f : [a,b] \to \mathbb{R}$ is a continuous function, differentiable on (a,b). Then there exists $x_0 \in (a,b)$ such that*

$$f'(x_0) = \frac{f(b) - f(a)}{b - a}.$$

Proof. We seek $m \in \mathbb{R}$ such that, with $g(x) = f(x) + mx$, we have $g(a) = g(b)$; that is,

$$f(a) + ma = f(b) + mb$$

This is easily solved:

$$m = -\frac{f(b) - f(a)}{b - a}$$

We can now apply Rolle's Theorem to

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

to conclude that there is an $x_0 \in (a, b)$ such that $g'(x_0) = 0$; that is,

$$f'(x_0) - \frac{f(b) - f(a)}{b - a} = 0$$

as claimed. □

Geometrically, the quantity $m = (f(b) - f(a))/(b - a)$ is the gradient of the chord joining the points $(a, f(a))$ and $(b, f(b))$. The MVT states that there is a point in (a, b) where the tangent to f is parallel to the chord.

Now, we can finish the connection between monotonicity and the sign of the derivative.

23.6 Theorem. *Suppose $a < b$ and $f : [a, b] \to \mathbb{R}$ is a continuous function, differentiable on (a, b). Then:*

if $f'(x)$ is	≥ 0	*on (a, b) then f is*	*increasing*	*on $[a, b]$*
	> 0		*strictly increasing*	
	≤ 0		*decreasing*	
	< 0		*strictly decreasing*	
	$= 0$		*constant*	

Proof. If $x, y \in [a, b]$ with $x < y$ then

$$f(y) - f(x) = \underbrace{(y - x)}_{>0} f'(z)$$

for some $z \in (x, y) \subseteq (a, b)$. The results now follow immediately, e.g. if $f' > 0$ then $f(y) - f(x) > 0$ giving $f(x) < f(y)$ for $x < y$, i.e. f is strictly increasing. The other cases are similar. □

This was framed using a function continuous on $[a, b]$ and differentiable on (a, b), in order to use the MVT. We can modify it for other situations:

23.7 Corollary. *Suppose I is an interval and $f : I \to \mathbb{R}$ is differentiable on I. Then:*

if $f'(x)$ is	≥ 0	*on I then f is*	*increasing*	*on I*
	> 0		*strictly increasing*	
	≤ 0		*decreasing*	
	< 0		*strictly decreasing*	
	$= 0$		*constant*	

Proof. Suppose $f' > 0$ on I and $x, y \in I$ with $x < y$. The previous result applied to the interval $[x, y]$ shows that f is strictly increasing on $[x, y]$; in particular, $f(x) < f(y)$, showing that f is strictly increasing on I. The other parts are similar. □

24 Inverse Functions

We now know that a differentiable function with strictly positive derivative is strictly increasing and continuous, and therefore (Theorem 19.11) has a strictly increasing, continuous inverse function. What can we say about the differentiability of the inverse function?

24.1 Theorem (Inverse Function Theorem for Differentiable Functions)**.** *Suppose I is an interval in $\mathbb{R}$ and $f : I \to \mathbb{R}$ is differentiable, with either $f' > 0$ on I or $f' < 0$ on I. Let $J = f(I)$, so by the IVT (Corollary 19.9) J is an interval and (Theorem 19.11) $f^{-1} : J \to I$ exists and is continuous. Then f^{-1} is differentiable on J and for any $y \in J$ we have*

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

Proof. Fix $y \in J$ and let $x = f^{-1}(y)$. We can now write, for h such that $x + h \in I$,

$$f(x+h) = f(x) + hp(h) \tag{*}$$

where p is continuous at 0 and $p(0) = f'(x)$. We can also define for k such that $y + k \in J$ and $k \neq 0$

$$q(k) = \frac{f^{-1}(y+k) - f^{-1}(y)}{k}$$

Our goal is to show that $q(k)$ has a limit as $k \to 0$ and to calculate this limit; this is the derivative of f^{-1} at y. Start with the rearrangement

$$f^{-1}(y+k) = f^{-1}(y) + kq(k) = x + kq(k) \tag{†}$$

(where $y + k \in J$ and $k \neq 0$) and apply f to give

$$y + k = f(x + kq(k))$$

Now, $kq(k)$ can play the role of h in (*) to give

$$y + k = f(x) + kq(k)p(kq(k))$$

and we have some cancellations: $y = f(x)$, so the first terms on each side cancel, and we can divide by $k \neq 0$ to give

$$1 = q(k)p(kq(k)) \implies q(k) = \frac{1}{p(kq(k)} \tag{‡}$$

Going back to (†), because f^{-1} is continuous (Theorem 19.11), we see that $kq(k) \to 0$ as $k \to 0$. We can therefore let $k \to 0$ in (‡) to give

$$q(k) \to \frac{1}{p(0)} = \frac{1}{f'(x)} = \frac{1}{f'(f^{-1}(y))}$$

as $k \to 0$, which finishes the proof. □

24.2 Example. If $q \in \mathbb{Z} \setminus \{0\}$ then we can define $f : (0,\infty) \to (0,\infty)$ by $f(x) = x^q$. We can also consider the inverse function $f^{-1} : (0,\infty) \to (0,\infty)$, $f^{-1}(y) = y^{1/q}$ and we know that

$$f'(x) = qx^{q-1} \begin{cases} > 0 & \text{if } q > 0 \\ < 0 & \text{if } q < 0 \end{cases}$$

so f satisfies the hypotheses of the Inverse Function Theorem. We therefore have

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

$$(f^{-1})'(y) = \frac{1}{q(y^{1/q})^{q-1}} = \frac{1}{q} y^{1/q-1}$$

A straightforward calculation with the chain rule now gives, in Leibniz notation,

$$\frac{\mathrm{d}}{\mathrm{d}x} x^{p/q} = \frac{p}{q} x^{p/q-1}$$

for $x \in (0,\infty)$ and p/q any non-zero rational number.

25 Taylor's Theorem

25.1 Notation. For $n \in \mathbb{N}_0$, we use the notation $f^{(n)}$ to mean the nth derivative of a function f, so $f^{(0)} = f$, $f^{(1)} = f'$, $f^{(2)} = f''$, etc. More formally, if $f : I \to \mathbb{R}$ then

$$f^{(0)} = f; \qquad f^{(n+1)} = \left(f^{(n)}\right)'$$

for those $n \in \mathbb{N}_0$ such that $f^{(n)}$ is differentiable on I. A function f for which $f^{(n)}$ exists on I is said to be *n times differentiable* on I.

25.2 Remark. Some observations on polynomials and their derivatives:

1. Suppose P is a polynomial, say

   $$P(x) = \sum_{n=0}^{N} a_n x^n$$

   Then we can easily express the derivatives of P at the origin in terms of the coefficients: $P(0) = a_0$, $P'(0) = a_1$, $P''(0) = 2a_2$ and in general

   $$P^{(n)}(0) = \begin{cases} n!a_n & \text{if } n \le N \\ 0 & \text{if } n > N \end{cases}$$

2. We can also turn this around and, from the derivatives at the origin, we can find the coefficients: given $c_0, c_1, \ldots, c_N \in \mathbb{R}$, if we define

   $$P(x) = \sum_{n=0}^{N} \frac{c_n}{n!} x^n$$

then we have

$$P^{(n)}(0) = \begin{cases} c_n & \text{if } n \le N \\ 0 & \text{if } n > N \end{cases}$$

Moreover, P is the unique polynomial with this property: if Q is another polynomial with the same derivatives at 0 then the degree of Q is no more than N, because the derivatives beyond the Nth are all zero, then applying part (1) to P and to $Q(x) = \sum_{n=1}^{N} b_n x^n$ gives $n!a_n = n!b_n$, hence $a_n = b_n$, for all n, showing that P and Q are exactly the same polynomial.

3. We can also do this at other points: if $x_0 \in \mathbb{R}$ and

$$P(x) = \sum_{n=0}^{N} \frac{c_n}{n!}(x - x_0)^n$$

then

$$P^{(n)}(x_0) = \begin{cases} c_n & \text{if } n \le N \\ 0 & \text{if } n > N \end{cases}$$

and again P is the unique polynomial with this property; or, equivalently, the unique polynomial of degree N or less with $P^{(n)}(x_0) = c_n$ for $0 \le n \le N$.

25.3 Definition. Suppose I is an interval, $x_0 \in I$ and $f : I \to \mathbb{R}$ is N times differentiable. The *Nth-order Taylor polynomial* of f at x_0 is defined by

$$P_N(x) = \sum_{n=0}^{N} \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n$$

In view of the previous remark, this is the unique polynomial of degree no more than N such that for $0 \le n \le N$ we have

$$P_N^{(n)}(x_0) = f^{(n)}(x_0)$$

Equivalently, $f - P_N$ and its first N derivatives are zero at x_0.

Taylor's Theorem describes how, under some circumstances, the Taylor polynomials for f at x_0 are good approximations for f near x_0. The advantage of having a good polynomial approximation to f is that many calculations can be done explicitly for polynomials, after which the results can be interpreted as approximations for the corresponding calculations for the function f. Much of Numerical Analysis, e.g. numerical integration and the numerical solution of differential equations, is based on ideas of polynomial approximation.

25.4 Theorem (Taylor's Theorem). *Suppose $x_0, h \in \mathbb{R}$ with $h \neq 0$. Let I be the closed interval with endpoints x_0 and $x_0 + h$, i.e. $I = [\min\{x_0, x_0 + h\}, \max\{x_0, x_0 + h\}]$. Suppose f is $N + 1$ times differentiable on I. Then there exists c between x_0 and $x_0 + h$ such that*

$$f(x_0 + h) = P_N(x_0 + h) + \frac{f^{(N+1)}(c)}{(N+1)!}h^{N+1}$$

where P_N is the Nth order Taylor polynomial of f at x_0:

$$P_N(x_0+h) = \sum_{n=0}^{N} \frac{f^{(n)}(x_0)}{n!} h^n$$

Note that h can be positive or negative; c being "between" x_0 and x_0+h means either $x_0 < c < x_0 + h$ for $h > 0$ or $x_0 + h < c < x_0$ for $h < 0$. The difference $f(x_0+h) - P_N(x_0+h)$ is called the *remainder* at $x_0 + h$. There are several different formulae known for the remainder; the one given here is called the *Lagrange form*.

⚠ The proof of this result is more demanding than most in the course; you might prefer to skip the details at first reading.

Proof of Taylor's Theorem. As remarked in Definition 25.3

$$f(x) - P_N(x) \text{ and its first } N \text{ derivatives are zero when } x = x_0.$$

Direct calculation also shows that

$$(x - x_0)^{N+1} \text{ and its first } N \text{ derivatives are zero when } x = x_0.$$

It follows that for any $m \in \mathbb{R}$

$$g(x) := f(x) - P_N(x) + m(x-x_0)^{N+1} \text{ and its first } N \text{ derivatives are zero when } x = x_0.$$

Now fix m so $g(x_0+h) = 0$: explicitly,

$$m = -\frac{1}{h^{N+1}} \left(f(x_0+h) - P_N(x_0+h) \right)$$

Let $c_0 = x_0 + h$. Now,

$$\begin{array}{rcrl}
 & \text{(Rolle)} & & \\
g(x_0) = g(c_0) = 0 & \Longrightarrow & g'(c_1) = 0 & \text{(some } c_1 \text{ between } x_0 \text{ and } c_0) \\
g'(x_0) = g'(c_1) = 0 & \Longrightarrow & g''(c_2) = 0 & \text{(some } c_2 \text{ between } x_0 \text{ and } c_1) \\
g''(x_0) = g''(c_2) = 0 & \Longrightarrow & g'''(c_3) = 0 & \text{(some } c_3 \text{ between } x_0 \text{ and } c_2) \\
 & \vdots & & \\
g^{(N)}(x_0) = g^{(N)}(c_N) = 0 & \Longrightarrow & g^{(N+1)}(c_{N+1}) = 0 & \text{(some } c_{N+1} \text{ between } x_0 \text{ and } c_N)
\end{array}$$

Now let $c = c_{N+1}$ and, referring to the definition of $g(x)$, note that the $(N+1)$st derivatives of $P_N(x)$ and $(x-x_0)^{N+1}$ are respectively 0 (because P_N has degree no more than N) and $(N+1)!$ (by direct calculation). The last line above now reduces to

$$f^{(N+1)}(c) + m(N+1)! = 0$$

Now substitute the definition of m and rearrange to give

$$f(x_0+h) = P_N(x_0+h) + \frac{f^{(N+1)}(c)}{(N+1)!} h^{N+1}$$

□

26 Differentiation of Power Series

In this section, we consider when it is possible to differentiate power series term by term, i.e. when it is true that

$$\frac{\mathrm{d}}{\mathrm{d}x}\sum_{n=0}^{\infty} a_n(x-x_0)^n = \sum_{n=1}^{\infty} na_n(x-x_0)^{n-1}$$

This is by no means a trivial question; for example (see Calculus) the Fourier series

$$\sum_{n=1}^{\infty} \frac{2(1-(-1)^n)}{n\pi}\sin(nt)$$

converges for all $t \in \mathbb{R}$ to the square wave function

$$\begin{cases} -1 & \text{if } (2k-1)\pi < t < 2k\pi \\ 1 & \text{if } 2k\pi < t < (2k+1)\pi \qquad (k \in \mathbb{Z}) \\ 0 & \text{if } t = k\pi \end{cases}$$

but its term-by-term derivative

$$\sum_{n=1}^{\infty} \frac{2(1-(-1)^n)}{\pi}\cos(nt)$$

has terms that do not tend to zero, and therefore cannot converge.

The most illuminating answers to this question require complex numbers (i.e. Complex Analysis; see Functions of a Complex Variable in the second year), so in this Real Analysis course for now we just sketch two basic results.

26.1 Theorem. *The power series*

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

and its formal derivative,

$$\sum_{n=1}^{\infty} na_n(x-x_0)^{n-1}$$

have the same radius of convergence.

Proof. Let (cf. Lemma 14.3)

$$S = \{r \geq 0 : a_n r^n \text{ is bounded as } n \to \infty\}$$
$$S' = \{r \geq 0 : na_n r^{n-1} \text{ is bounded as } n \to \infty\}$$

According to Theorem 14.4, the radii of convergence of the two series are respectively $\sup(S)$ and $\sup(S')$, taking the radius of convergence to be ∞ if the relevant set is unbounded. It is therefore enough to show that if either of these sets is bounded then so is the other, and that in that case they have the same supremum.

Firstly, suppose S is bounded and that $r \in S'$. Then for some C we have $|na_n r^{n-1}| \le C$ for all $n \in \mathbb{N}$; we can now estimate

$$|a_n r^n| = \left|\frac{r}{n} na_n r^{n-1}\right| \le \frac{rC}{n} \le rC$$

for all $n \in \mathbb{N}$, showing that $r \in S$. This shows that $S' \subseteq S$; in consequence, if S is bounded then so is S' and $\sup(S') \le \sup(S)$.

Secondly, suppose S' is bounded and $r \in S$. We need to show that $r \le \sup(S')$. If $r = 0$ then this is trivially true. If $r > 0$ then for some C we have $|a_n r^n| \le C$ for all $n \in \mathbb{N}$; then, for any s with $0 < s < r$ we have

$$|na_n s^{n-1}| = n\left(\frac{s}{r}\right)^{n-1} \frac{1}{r}|a_n r^n| \le \frac{Cn}{r}\left(\frac{s}{r}\right)^{n-1}$$

which tends to zero as $n \to \infty$ by the hierarchy of limits; in particular, it is bounded, so $s \in S'$ and hence $s \le \sup(S')$. Since this is true for every $s < r$, we have $r \le \sup(S')$, showing that S is bounded and that $\sup(S) \le \sup(S')$.

Combining these two results, if either S or S' is bounded then so is the other and the two inequalities $\sup(S') \le \sup(S)$ and $\sup(S) \le \sup(S')$ both hold, showing that $\sup(S') = \sup(S)$. □

Note that, by induction, all the higher-order derivative series also have the same radius of convergence.

26.2 Theorem. *Suppose a function is defined by a power series*

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

which converges on the interval $(x_0 - R, x_0 + R)$. Then f is differentiable on $(x_0 - R, x_0 + R)$ and for $x \in (x_0 - R, x_0 + R)$ we have

$$f'(x) = \sum_{n=1}^{\infty} na_n (x - x_0)^{n-1}$$

⚠ The proof of this result is more demanding than most in the course; you might prefer to skip the details at first reading.

Proof. As we have often done before, we work with $x_0 = 0$ to simplify notation. To recover the general result, replace x with $x - x_0$ throughout.

Fix $x \in (-R, R)$. To avoid problems at $\pm R$, where the series might diverge, choose some r with $|x| < r < R$; we will work in the interval $[-r, r]$, so suppose $0 < |h| < r - |x|$, which guarantees that $|x + h| < r$. Now consider the difference quotient

$$\begin{aligned}
\frac{f(x+h) - f(x)}{h} &= \frac{1}{h}\left[\sum_{n=0}^{\infty} a_n (x+h)^n - \sum_{n=0}^{\infty} a_n x^n\right] \\
&= \sum_{n=1}^{\infty} a_n \frac{(x+h)^n - x^n}{h} \\
&= a_1 + \sum_{n=2}^{\infty} a_n \frac{(x+h)^n - x^n}{h}
\end{aligned}$$

Here, the $n = 0$ terms cancel (both are a_0) and the $n = 1$ terms combine to give $(a_1(x+h) - a_1x)/h = a_1$. We break off this term because it slightly simplifies the next step, which is to apply Taylor's Theorem to $(x+h)^n$: we have

$$(x+h)^n = x^n + nx^{n-1}h + \frac{1}{2}n(n-1)c_n^{n-2}h^2$$

for some c_n between x and $x+h$, and hence

$$\frac{(x+h)^n - x^n}{h} = nx^{n-1} + \frac{1}{2}n(n-1)c_n^{n-2}h$$

Substituting this into the formula for the difference quotient, we obtain

$$\begin{aligned}\frac{f(x+h)-f(x)}{h} &= a_1 + \sum_{n=2}^{\infty} a_n\left[nx^{n-1} + \frac{1}{2}n(n-1)c_n^{n-2}h\right] \\ &= a_1 + \sum_{n=2}^{\infty} na_nx^{n-1} + h\sum_{n=2}^{\infty}\frac{1}{2}n(n-1)a_nc_n^{n-2} \\ &= \sum_{n=1}^{\infty} na_nx^{n-1} + h\sum_{n=2}^{\infty}\frac{1}{2}n(n-1)a_nc_n^{n-2}\end{aligned}$$

Here, the first series converges because it is (apart from, initially, one term), the formal derivative of the series we're working with, so it has the same radius of convergence. The second series then converges because the first one does. We now want to let $h \to 0$. To understand what happens, we need an estimate on how large the second series could be: since c_n lies between x and $x+h$, we have $|c_n| \le r$ so

$$\left|\sum_{n=2}^{\infty}\frac{1}{2}n(n-1)a_nc_n^{n-2}\right| \le \sum_{n=2}^{\infty}\frac{1}{2}n(n-1)|a_n|r^{n-2}$$

which converges because the second term-by-term derivative of the series describing f converges absolutely within its radius of convergence, which is R by the previous theorem; and we chose r with $0 < r < R$.

We can therefore let $h \to 0$ in the last formula for the difference quotient and conclude that

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

as claimed. □

26.3 Example. In Definition 15.1, we defined

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

It is an immediate consequence of Theorem 26.2 that $\exp' = \exp$, that $\sin' = \cos$ and that $\cos' = -\sin$.

27 Elementary Functions

We can now finally establish properties E1–E4 in the original stated objectives of the course (Section 1).

We know that exp satisfies the Initial Value Problem

$$\exp' = \exp; \qquad \exp(0) = 1 \tag{IVP}$$

Suppose u is any solution of (IVP), i.e. that $u : \mathbb{R} \to \mathbb{R}$, $u' = u$ and $u(0) = 1$. We can differentiate $u(t)\exp(-t)$

$$\frac{\mathrm{d}}{\mathrm{d}t} u(t)\exp(-t) = -u(t)\exp(-t) + u(t)\exp(-t) = 0$$

using the fact that exp and u are both solutions to (IVP), in particular that $\exp' = \exp$ and $u' = u$. It follows that $u(t)\exp(-t)$ is constant. To find its constant value, we use the other part of the fact that exp and u both satisfy (IVP), i.e. that $u(0) = \exp(0) = 1$; since $u(0)\exp(-0) = 1$, we have $u(t)\exp(-t) = 1$ for all t. This is true for *any* solution u of (IVP); in particular, it is true for $u(t) = \exp(t)$, giving $\exp(t)\exp(-t) = 1$. This gives $\exp(-t) = 1/\exp(t)$ and we can substitute this into $u(t)\exp(-t) = 1$ to give $u(t) = \exp(t)$, showing that exp is the only solution to (IVP).

This looks rather unmotivated, but there is a good way to think about it: the idea is that we started with the initial value 1 at time 0, follow the solution u forward in time by t to give $u(t)$, then follow the solution exp backwards in time by t to give $u(t)\exp(-t)$. For the two solutions to be the same, we expect this to bring us back where we started, i.e. we expect to find that $u(t)\exp(-t) = 1$, which is exactly what the calculation above established.

The other fact that we should establish is that $\exp(s+t) = \exp(s)\exp(t)$. We prove this in much the same way as before: fix s and differentiate $\exp(s+t)\exp(-s)\exp(-t)$ with respect to t, giving

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}\exp(s+t)\exp(-s)\exp(-t) = \\ &\qquad -\exp(s+t)\exp(-s)\exp(-t) + \exp(s+t)\exp(-s)\exp(-t) = 0\end{aligned}$$

so $\exp(s+t)\exp(-s)\exp(-t)$ remains constant as t varies. To find its constant value, we put $t = 0$, giving $\exp(s)\exp(-s) = 1$ (as seen above). We thus have for all s and t

$$\exp(s+t)\exp(-s)\exp(-t) = 1$$

which we can multiply by $\exp(s)\exp(t)$ to give $\exp(s+t) = \exp(s)\exp(t)$ as expected.

Again, this can be thought of in a dynamical way: starting at time 0 with the initial value 1, we evolve forward in time by $s+t$ to give $\exp(s+t)$, then use this as a new initial value and evolve backwards in time by s, giving $\exp(s+t)\exp(-s)$, then backwards again by t, giving $\exp(s+t)\exp(-s)\exp(-t)$. We expect to get back where we started, i.e. to find that $\exp(s+t)\exp(-s)\exp(-t) = 1$, which is exactly what the calculation above established.

In summary:

27.1 Theorem. *The exponential function is the unique solution to the IVP* $\exp' = \exp$; $\exp(0) = 1$. *It satisfies the functional equation* $\exp(s+t) = \exp(s)\exp(t)$ *for all* $s, t \in \mathbb{R}$.

It is immediate from the definition

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

that, for $t \geq 0$, we have $\exp(t) \geq 1$ (the first term of the series). Since $\exp(t)\exp(-t) = 1$, we deduce that if $t \leq 0$ then $0 < \exp(t) \leq 1$; in particular, $\exp(t) > 0$ for all t. Since $\exp' = \exp$, this shows that exp is strictly increasing. We also have, for $t \geq 0$, $\exp(t) \geq 1+t$ (the first two terms of the series). It follows that $\exp(t) \to \infty$ as $t \to \infty$ and, because $\exp(t)\exp(-t) = 1$, that $\exp(t) \to 0$ as $t \to -\infty$. The range of exp is therefore the open halfline $(0, \infty)$. We can use the Inverse Function Theorem (Theorem 24.1) to see that there is a strictly increasing inverse function, which we call $\log : (0, \infty) \to \mathbb{R}$, which satisfies

$$\log'(x) = \frac{1}{\exp'(\log(x))} = \frac{1}{x}$$

because $\exp' = \exp$ and exp is the inverse function to log. We also have for $x, y > 0$

$$\exp(\log(x) + \log(y)) = \exp(\log(x))\exp(\log(y)) = xy$$

Applying log to the left-hand and right-hand expressions:

$$\log(x) + \log(y) = \log(xy)$$

In summary:

27.2 Theorem. *The exponential function is strictly increasing on* $\mathbb{R}$ *with range* $(0, \infty)$. *The inverse function is denoted* log *(or* ln*), and satisfies* $\log(1) = 0$ *and* $\log'(x) = 1/x$. *It satisfies the functional equation* $\log(xy) = \log(x) + \log(y)$.

Finally, we can define arbitrary powers of positive numbers:

27.3 Definition. We now define for any $x > 0$ and $y \in \mathbb{R}$,

$$x^y = \exp(y \log(x)).$$

If we define $\mathrm{e} = \exp(1)$, then this gives $\mathrm{e}^x = \exp(x)$.

All the expected properties can be easily verified, $x^{y+z} = x^y x^z$ follows from the functional equation for exp and we can use the chain rule to differentiate:

$$\frac{\mathrm{d}}{\mathrm{d}x} x^y = \exp(y \log(x)) \frac{y}{x} = x^y \frac{y}{x} = y x^{y-1}$$

Similarly, we can establish properties T1–T4 from Section 1.

27.4 Theorem. *The trigonometric functions are the unique solution to the IVP* $\sin' = \cos$, $\cos' = -\sin$, $\sin(0) = 0$, $\cos(0) = 1$. *They satisfy the Pythagoras identity* $\cos^2 + \sin^2 = 1$ *and the addition formulae*

$$\sin(x+y) = \sin(x)\cos(y) + \sin(y)\cos(x)$$
$$\cos(x+y) = \cos(x)\cos(y) - \sin(x)\sin(y)$$

Both sin *and* cos *are periodic. There is a number* $\pi \approx 3.14$ *such that* 2π *is their minimal period.*

The proof has been relegated to an appendix; much of it is very similar to the arguments used for the exponential function, but somewhat more complicated because it uses (small!) vectors and matrices to work with the two coupled equations at the same time. Also, the construction of π is quite fiddly.

28 Part V: Key Points

The theory of differentiation has few surprises for anyone familiar with calculus.

If I is an interval, $f : I \to \mathbb{R}$ and $x_0 \in I$ then we define

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

provided this limit exists. This can also be expressed as

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

(Definition 21.1) or

$$f(x_0 + h) = f(x_0) + hp(h)$$

where p is continuous at 0 and $p(0) = f'(0)$ (Theorem 21.3).

The familiar rules for calculating derivatives (linearity, product rule, quotient rule, chain rule) work as expected (Theorem 22.1, Theorem 22.3).

The precision of the definition of the derivative as a limit helps to clarify the exact relationship between monotonicity and the sign of the derivative (Theorem 23.2, Theorem 23.6, Corollary 23.7). In summary:

- A differentiable function on an interval is increasing if and only if its derivative is non-negative.
- If a differentiable function on an interval has strictly positive derivative then it is strictly increasing.
- A strictly increasing function on an interval need not have strictly positive derivative.
- Replace "increasing", "non-negative" and "strictly positive" with "decreasing", "non-positive" and "strictly negative" in the above.
- A differentiable function on an interval is constant if and only if its derivative is identically zero.
- At an interior local extremum, the derivative is zero; endpoints of an interval can be local extrema with non-zero derivative; points with zero derivative need not be local extrema.

The fundamental result that makes this relationship work is the Mean Value Theorem (Theorem 23.5): if f is continuous on $[a, b]$ and differentiable on (a, b) then for some $x_0 \in (a, b)$ we have

$$f'(x_0) = \frac{f(b) - f(a)}{b - a}$$

This understanding of monotonicity helps us formulate the Inverse Function Theorem (Theorem 24.1, cf. 19.11): if f is differentiable on an interval I and $f' > 0$ on I or $f' < 0$ on I then f^{-1} is differentiable on $J = f(I)$ and for $y \in J$ we have

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

This allows us to differentiate fractional powers, verifying the usual formula.

Power series can, within their radius of convergence, be differentiated term by term. The resulting series has the same radius of convergence as the original series and its sum is the derivative of the sum of the original series (Theorem 26.2). It follows that the exponential and trigonometric functions, defined earlier (Definition 15.1) as power series, obey the expected differential equations: $\exp' = \exp$, $\sin' = \cos$ and $\cos' = -\sin$ (Example 26.3). Moreover, if we specify the initial values $\exp(0) = 1$, $\sin(0) = 0$ and $\cos(0) = 1$, then these are the only solutions to these initial value problems. This can be exploited to establish the well-known relationships $\exp(s + t) = \exp(s)\exp(t)$, $\cos^2 + \sin^2 = 1$ and the addition formulae for trigonometric functions, and to show that sin and cos are periodic and thereby define π (Theorem 27.1, Theorem 27.4).

We can also observe that exp is a strictly increasing bijection from $\mathbb{R}$ to $(0, \infty)$ and define log to be its inverse function; the properties established for exp then lead to the expected properties of log, that $\log(xy) = \log(x) + \log(y)$ and that $\mathrm{d}/\mathrm{d}x \log(x) = 1/x$ (Theorem 27.2).

With exp and log in place, we can define

$$x^y = \exp(y \log(x))$$

for any $x > 0$ and $y \in \mathbb{R}$, and verify that

$$\frac{\mathrm{d}}{\mathrm{d}x} x^y = y x^{y-1}$$

(Definition 27.3).

Literature:

1. F. Mary Hart *Guide to Analysis* (2nd Edition, Palgrave Macmillan 2001, JBM classmark S7 Har) Chapters 5.

2. any other introduction to Real Analysis; browse classmark S7 in the library.

B The trigonometric functions

This is a technical appendix, for anyone who would like a deeper understanding of the way the the trigonometric functions work; in particular, none of this is examinable!

We know that sin and cos satisfy $\sin' = \cos$, $\cos' = -\sin$, $\sin(0) = 0$, $\cos(0) = 1$. Consider the Initial Value Problem

$$u'(t) = -v(t), v'(t) = u(t) \qquad (t \in \mathbb{R}; u(0) = u_0, v(0) = v_0) \tag{IVP}$$

We can express a solution of this using $\sin(t)$ and $\cos(t)$. In matrix form, this looks like

$$\begin{bmatrix} u(t) \\ v(t) \end{bmatrix} = \begin{bmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{bmatrix} \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} \tag{SOL}$$

(you are invited to check that this really does satisfy the IVP). In this solution, we start with a *state vector* $[u_0, v_0]^T$ representing $u(t)$ and $v(t)$ at time 0; at time 0; multiplying by the matrix gives us the corresponding values at time t. In fact, because t does not appear in the ODE (it is *autonomous*), $[u_0, v_0]^T$ could represent the state at time t_0, in which case $[u(t), v(t)]^T$ would represented the state at time $t_0 + t$.

In fact, this is the only solution. To see this, take any functions u, v satisfying (IVP) and consider

$$\begin{bmatrix} \cos(t) & \sin(t) \\ -\sin(t) & \cos(t) \end{bmatrix} \begin{bmatrix} u(t) \\ v(t) \end{bmatrix} = \begin{bmatrix} \cos(t)u(t) + \sin(t)v(t) \\ -\sin(t)u(t) + \cos(t)v(t) \end{bmatrix} \tag{A}$$

The matrix here is created by replacing t with $-t$ in the solution above; the idea is that it should take the state vector $[u(t), v(t)]$ at time t and return the corresponding state vector at time 0, i.e. $[u_0, v_0]^T$; that is, the vector on the RHS should be constant in time. To verify this, we differentiate the RHS of(A) wrt t giving (because $u' = -v$ and $v' = u$)

$$\begin{bmatrix} -\cos(t)v(t) - \sin(t)u(t) + \sin(t)u(t) + \cos(t)v(t) \\ \sin(t)v(t) - \cos(t)u(t) + \cos(t)u(t) - \sin(t)v(t) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

This shows that the quantities in (A) are constant as t varies; in particular, they are always equal to their values at $t = 0$, i.e. for all t we have

$$\begin{bmatrix} \cos(t) & \sin(t) \\ -\sin(t) & \cos(t) \end{bmatrix} \begin{bmatrix} u(t) \\ v(t) \end{bmatrix} = \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} \tag{B}$$

Since this is true for any solution to the IVP, we can substitute the particular solution (SOL) to give

$$\begin{bmatrix} \cos(t) & \sin(t) \\ -\sin(t) & \cos(t) \end{bmatrix} \begin{bmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{bmatrix} \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} = \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} \tag{3}$$

Because this is true for all vectors $[u_0, v_0]$, the two matrices on the LHS must be inverses to each other. We can therefore multiply equation (B) by the matrix

$$\begin{bmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{bmatrix}$$

to give

$$\begin{bmatrix} u(t) \\ v(t) \end{bmatrix} = \begin{bmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{bmatrix} \begin{bmatrix} u_0 \\ v_0 \end{bmatrix}$$

As a formula, this is identical to (SOL), but the context is different: we already knew that (SOL) describes a solution to (IVP), and we have now shown that every solution to (IVP) is represented by SOL). That is (IVP) has one and only one solution, namely (SOL).

We can also now note that if u and v between them satisfy the ODE $u'(t) = -v(t)$, $v'(t) = u(t)$ ($t \in \mathbb{R}$), with no initial values or other conditions specified, then we can manufacture an IVP that u and v satisfy by specifying $u_0 = u(0)$ and $v_0 = v(0)$ in (IVP). Since we know that (SOL) is the only solution of (IVP), we now have

$$\begin{aligned} \begin{bmatrix} u(t) \\ v(t) \end{bmatrix} = \begin{bmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{bmatrix} \begin{bmatrix} u(0) \\ v(0) \end{bmatrix} &= \begin{bmatrix} u(0)\cos(t) - v(0)\sin(t) \\ v(0)\cos(t) + u(0)\sin(t) \end{bmatrix} \\ &= \begin{bmatrix} A\cos(t) - B\sin(t) \\ B\cos(t) + A\sin(t) \end{bmatrix} \qquad \text{(C)} \end{aligned}$$

where we can think of A and B as unknown constants; this describes all possible solutions of $u'(t) = -v(t)$, $v'(t) = u(t)$ ($t \in \mathbb{R}$).

The addition formulae and the Pythagoras identity $\cos^2(t) + \sin^2(t) = 1$ are now easily established. Here are two ways to think about them:

1. In terms of general solutions: if we fix s and define $u(t) = \cos(t+s)$ and $v(t) = \sin(t+s)$ then we have a solution to the ODE $u'(t) = -v(t)$ and $v'(t) = u(t)$. According to (C), we must have

   $$\begin{bmatrix} u(t) \\ v(t) \end{bmatrix} = \begin{bmatrix} u(0)\cos(t) - v(0)\sin(t) \\ v(0)\cos(t) + u(0)\sin(t) \end{bmatrix} = \begin{bmatrix} \cos(s)\cos(t) - \sin(s)\sin(t) \\ \sin(s)\cos(t) + \cos(s)\sin(t) \end{bmatrix}$$

   which are the addition formulae.

2. In terms of initial value problems: the matrices

   $$\begin{bmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{bmatrix} \qquad \begin{bmatrix} \cos(s) & -\sin(s) \\ \sin(s) & \cos(s) \end{bmatrix} \qquad \begin{bmatrix} \cos(t+s) & -\sin(t+s) \\ \sin(t+s) & \cos(t+s) \end{bmatrix}$$

   act on a state vector to find the state vector at times t, s and $t+s$ in the future. The product of the first two will act on a state vector to give the state vector at time s in the future then at time t in the future from s; that is, at time $t+s$ in the future from the original beginning. It follows that

   $$\begin{bmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{bmatrix} \begin{bmatrix} \cos(s) & -\sin(s) \\ \sin(s) & \cos(s) \end{bmatrix} = \begin{bmatrix} \cos(t+s) & -\sin(t+s) \\ \sin(t+s) & \cos(t+s) \end{bmatrix}$$

   and multiplying out gives the addition formulae for $\cos(t+s)$ and $\sin(t+s)$.

In either case,

$$\cos(t-t) = \cos(t)\cos(-t) - \sin(t)\sin(-t)$$

which, given that cos is even and sin is odd (which are clear from the Taylor series) gives $\cos^2(t)+\sin^2(t) = 1$. Or, avoiding the use of the odd and even properties, we can differentiate $\cos^2(t)+\sin^2(t)$ to give $-2\cos(t)\sin(t)+2\sin(t)\cos(t) = 0$, directly from the ODE; we now have that $\cos^2(t)+\sin^2(t)$ is constant and determine that constant value by substituting $t = 0$.

The last challenge is to define π. For any $x \in (0,1]$, the Taylor series for both $\sin(x)$ and $\cos(x)$ consist of terms of alternating sign and strictly decreasing magnitude. Referring back to the proof of the Leibniz Alternating Series test (Theorem 12.1), the odd partial sums are all greater than the infinite sum and the even partial partial sums are all less than the infinite sum. In particular, we have

$$\begin{aligned}
\sin(x) &> x - \frac{x^3}{3!} > 0 \\
\cos(x) &> 1 - \frac{x^2}{2!} > 0 \\
\cos(x) &< 1 - \frac{x^2}{2!} + \frac{x^4}{4!}
\end{aligned}$$

The first and third of these inequalities give, for $x = 1$,

$$\begin{aligned}
\sin(1) &> 1 - \frac{1}{6} = \frac{5}{6} \\
\cos(1) &< 1 - \frac{1}{2} + \frac{1}{24} = \frac{13}{24}
\end{aligned}$$

This gives us

$$\cos(1) - \sin(1) < \frac{13}{24} - \frac{5}{6} = -\frac{7}{24}$$

in comparison with

$$\cos(0) - \sin(0) = 1$$

By the Intermediate Value Theorem, there exists $x_0 \in (0,1)$ such that $\sin(x_0)-\cos(x_0) = 0$. Now suppose there also exists a different $x_1 \in (0,1)$ with $\sin(x_1)-\cos(x_1) = 0$. By Rolle's Theorem, there exists x_2 between x_0 and x_1 such that $\cos(x_2)+\sin(x_2) = 0$. But we have already seen that, since $x_2 \in (0,1)$, $\sin(x_2) > 0$ and $\cos(x_2) > 0$, so this cannot happen. We therefore have one and only one $x_0 \in (0,1)$ such that $\sin(x_0) = \cos(x_0)$. Given this value of x_0, we define $\pi = 4x_0$. Note that $0 < x_0 < 1$, so $0 < \pi < 4$.

Now, we have some tedious and routine steps towards establishing periodicity. We know that $\cos(\pi/4)$ and $\sin(\pi/4)$ are both positive and equal; since $\cos^2+\sin^2 = 1$, we have $\cos(\pi/4) = \sin(\pi/4) = \sqrt{2}/2$. We can apply the addition formulae:

$$\begin{aligned}
\cos(x+\pi/4) &= \frac{\sqrt{2}}{2}(\cos(x)-\sin(x)) \\
\sin(x+\pi/4) &= \frac{\sqrt{2}}{2}(\cos(x)+\sin(x))
\end{aligned}$$

and then

$$\begin{aligned}\cos(x+\pi/2) &= \cos((x+\pi/4)+\pi/4)\\ &= \frac{\sqrt{2}}{2}[\cos(x+\pi/4)-\sin(x+\pi/4)]\\ &= \frac{1}{2}[\cos(x)-\sin(x)-\cos(x)-\sin(x)]\\ &= -\sin(x)\end{aligned}$$

and

$$\begin{aligned}\sin(x+\pi/2) &= \sin((x+\pi/4)+\pi/4)\\ &= \frac{\sqrt{2}}{2}[\cos(x+\pi/4)+\sin(x+\pi/4)]\\ &= \frac{1}{2}[\cos(x)-\sin(x)+\cos(x)+\sin(x)]\\ &= \cos(x)\end{aligned}$$

This quickly yields

$$\cos(x+\pi) = \cos((x+\pi/2)+\pi/2) = -\sin(x+\pi/2) = -\cos(x)$$

and

$$\sin(x+\pi) = \sin((x+\pi/2)+\pi/2) = \cos(x+\pi/2) = -\sin(x)$$

which finally gives

$$\cos(x+2\pi) = \cos(x); \qquad \sin(x+2\pi) = \sin(x)$$

so cos and sin are 2π-periodic.

We should also check that 2π is the minimal period. We already have $\sin(x+\pi/2) = \cos(x)$, so sin and cos have the same minimal period and we need only consider one of them; we focus on sin.

We have already seen that sin and cos are positive on $(0,1]$; since cos is even and $\cos(0) = 1$, it is in fact positive on $[-1,1]$. Now consider $\sin(x+\pi/2) = \cos(x)$, which is positive on $[-1,1]$; it follows that sin is positive on $[\pi/2-1, \pi/2+1]$. We have seen that $\pi < 4$, so $2 > \pi/2$ and $1 > \pi/2-1$; the intervals $(0,1]$ and $[\pi/2-1, \pi/2+1]$ overlap, and we have sin positive on $(0, \pi/2+1]$. Because $\sin(\pi/2+x) = \cos(x)$ is even, we can stretch this a little more to see than sin is positive on $(0,\pi)$. We then have $\sin(\pi) = 0$ and, because $\sin(x+\pi) = -\sin(x)$, sin is negative on $(\pi, 2\pi)$.

Now, for any $0 < L < 2\pi$, we have $\sin(\pi - L/2) > 0$ and $\sin(L+\pi/2) < 0$, so L cannot be a period for sin; 2π is thus the minimal period.

Part VI
The Riemann Integral

Throughout this section, we consider bounded functions on closed, bounded intervals $f : [a,b] \to \mathbb{R}$; that is, functions for which the set

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

is a bounded set. The ideas of supremum and infimum from the beginning of the course are heavily used.

29 The Integral

29.1 Definition. A *partition* of an interval $[a,b]$ is a finite list of numbers

$$P = (x_0, x_1, \ldots, x_N)$$

where

$$a = x_0 < x_1 < \cdots < x_N = b.$$

A partition P divides $[a,b]$ into N subintervals, from x_{n-1} to x_n for $1 \le n \le N$. Now suppose $f : [a,b] \to \mathbb{R}$ is bounded. We define the *lower sum* and *upper sum* of f over P by

$$L(f,P) = \sum_{n=1}^{N} (x_n - x_{n-1}) \inf_{x \in [x_{n-1}, x_n]} f(x)$$
$$U(f,P) = \sum_{n=1}^{N} (x_n - x_{n-1}) \sup_{x \in [x_{n-1}, x_n]} f(x)$$

The *lower integral* and *upper integral* of f from a to b are defined by

$$L_a^b f = \sup\{L(f,P) : \ P \text{ is a partition of } [a,b]\}$$
$$U_a^b f = \inf\{U(f,P) : \ P \text{ is a partition of } [a,b]\}$$

(see remark below for why we are taking the supremum and infimum of bounded sets here). Finally, we say that f is *integrable* over $[a,b]$ if $L_a^b f = U_a^b f$. In this case, we define the integral over $[a,b]$ of f by

$$\int_a^b f = L_a^b f = U_a^b f$$

The idea here (see Figure 11) is that, based on an intuitive idea of the "area under the curve", every lower sum is a lower bound for the area, and therefore the lower integral is a lower bound for the area (in fact, the greatest lower bound, obtainable with these techniques). Equally, every upper sum is an upper bound for the area and

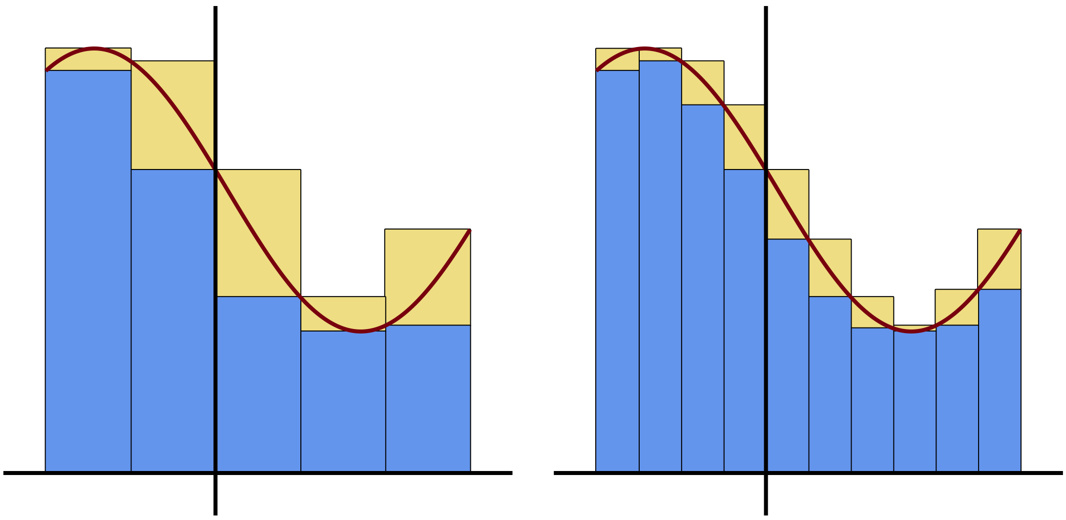


Figure 11: Illustrating many things! Two partitions and their lower and upper sums; the second is a refinement (Definition 29.5) of the first. The lower sum (Definition 29.1) is the total area of the lower, blue, rectangles, reaching up from the axis to touch the curve. The upper sum is the total area of all the rectangles, containing the region between the curve and the axis. The area of the upper, yellow, rectangles is the difference between the upper and lower sums, which appears in the Cauchy criterion for integration (Theorem 29.11). The vertical line represents the point c in Theorem 30.3; the sum of the areas of the rectangles (of any type) in the whole graph is the sum of the areas to the left of the line plus the sum of the areas to the right of the line.

therefore the upper integral is an upper bound for the area (in fact, the least upper bound, obtainable with these techniques). In symbols:

$$L_a^b f \leq (\text{area under the curve}) \leq U_a^b f$$

If the upper and lower integrals coincide then there is only one possible value for the area: the common value of the two bounds.

The definition of the integral was proposed by Gaston Darboux[7] as a simplification of an earlier definition first given by Bernhard Riemann[8]; it is equivalent to Riemann's so this theory of integration is often called the "Riemann integral." It stands in contrast with the "Lebesgue integral", a much more powerful but technically more demanding theory of integration developed by Henri Lebesgue[9]. See Appendix C for a presentation of Riemann's approach and its equivalence wth Darboux's.

29.2 Remark. If $P = (x_0, x_1, \dots, x_N)$ is a partition of $[a,b]$ and $f : [a,b] \to \mathbb{R}$ is bounded then

$$\begin{aligned} L(f,P) &= \sum_{n=1}^{N} (x_n - x_{n-1}) \inf_{x \in [x_{n-1}, x_n]} f(x) \\ &\leq \sum_{n=1}^{N} (x_n - x_{n-1}) \sup_{x \in [a,b]} f(x) \\ &= \left(\sup_{x \in [a,b]} f(x) \right) \sum_{n=1}^{N} (x_n - x_{n-1}) \\ &= (b-a) \sup_{x \in [a,b]} f(x) \end{aligned}$$

(telescopic sum; terms cancel in pairs leaving only $-x_0 = -a$ and $x_N = b$). This shows that the lower sums are bounded above, so their supremum $L_a^b f$ is well-defined and

$$L_a^b f \leq (b-a) \sup_{x \in [a,b]} f(x)$$

A similar argument shows that $U_a^b f$ is well defined and that

$$U_a^b f \geq (b-a) \inf_{x \in [a,b]} f(x)$$

29.3 Corollary (A Really Rough Estimate)**.** *If $f : [a,b] \to \mathbb{R}$ is integrable, then*

$$(b-a) \inf_{x \in [a,b]} f(x) \leq \int_a^b f \leq (b-a) \sup_{x \in [a,b]} f(x)$$

Proof. If f is integrable, we can replace both $L_a^b f$ and $U_a^b f$ by $\int_a^b f$ in the inequalities at the end of the Remark 29.2. $\square$

[7]Gaston Darboux, 1826–1866 (MacTutor, Wikipedia)
[8]Bernhard Riemann, 1826–1866 (MacTutor, Wikipedia)
[9]Henri Lebesgue, 1875–1941 (MacTutor, Wikipedia)

29.4 Example (constant functions). Define $f : [a,b] \to \mathbb{R}$ by $f(x) = c$. Then, in the definitions of lower and upper sums, all the sup and inf terms work out simply to c. All the upper and lower sums have the form

$$\sum_{n=1}^{N}(x_n - x_{n-1})c$$

Exactly as in Remark 29.2, this is a telescoping series, with the terms cancelling in pairs to give

$$L(f,P) = U(f,P) = (x_N - x_0)c = (b-a)c$$

for all P. The lower and upper integrals are thus the supremum and infimum of the singleton $\{(b-a)c\}$, so both are equal to $(b-a)c$; we conclude that constant functions are integrable and that

$$\int_a^b f = \int_a^b c = (b-a)c$$

Simple as this is, it is one of very few examples that we can work out without developing some more theory. In particular, we need to establish the intuitively obvious fact that $L_a^b f \le U_a^b f$, for which we introduce the idea of *refinements*.

29.5 Definition. A partition P of an interval $[a,b]$ is called a *refinement* of a partition Q of $[a,b]$ if every point of Q is also a point of P; that is, if P is either equal to Q or can be obtained from Q by adding extra points.

29.6 Theorem. *Any two partitions P and Q of an interval $[a,b]$ have a common refinement.*

Proof. If

$$P = (x_0, x_1, \ldots, x_M)$$
$$Q = (y_0, y_1, \ldots, y_N)$$

then we can construct a partition R by listing all the points x_m and y_n in a single list, in increasing order, discarding duplicates. This new partition contains every point of P and every point of Q, so is a refinement of both P and Q. □

29.7 Theorem. *If P and Q are partitions of $[a,b]$, $f : [a,b] \to \mathbb{R}$ is bounded and P is a refinement of Q then*

$$L(f,P) \ge L(f,Q)$$
$$U(f,P) \le U(f,Q)$$

So, refining a partition increases every associated lower sum and decreases every associated upper sum.

Proof. It is enough to show that if we add one point to a partition, then the lower sum increases and the upper sum decreases; the general result then follows by induction.

Suppose then that

$$Q = (x_0, x_1, \ldots, x_N)$$

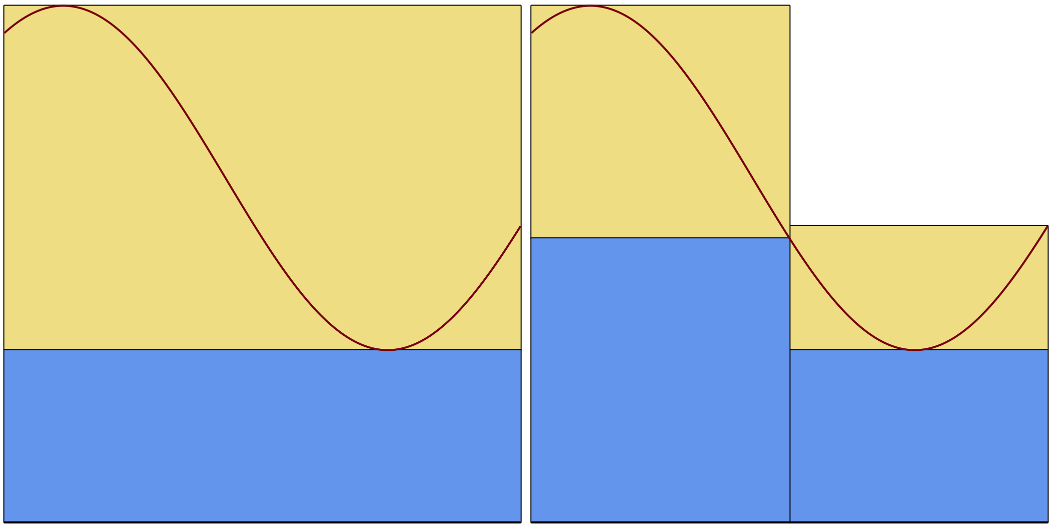


Figure 12: Refinement in one interval: lower sum rises, upper sum falls (Theorem 29.7)

and that we create P by adding a single point x^*, not equal to any of the existing points. Then for some n we have $x_{n-1} < x^* < x_n$; the subinterval $[x_{n-1}, x_n]$ in Q is replaced by two subintervals $[x_{n-1}, x^*]$ and $[x^*, x_n]$ in P. Looking at the formula for the lower sums, all the other subintervals contribute the same to the lower sum for P as they do for Q. We therefore have

$$L(f,P) - L(f,Q) = \left((x^* - x_{n-1}) \inf_{x\in[x_{n-1},x^*]} f(x)\right) + \left((x_n - x^*) \inf_{x\in[x^*,x_n]} f(x)\right) - \left((x_n - x_{n-1}) \inf_{x\in[x_{n-1},x_n]} f(x)\right)$$

Let

$$y = \inf_{x\in[x_{n-1},x_n]} f(x)$$

Then, because they are infima over smaller sets,

$$\inf_{x\in[x_{n-1},x^*]} f(x) \geq y \qquad \inf_{x\in[x^*,x_n]} f(x) \geq y$$

We now have

$$L(f,P) - L(f,Q) \geq (x^* - x_{n-1})y + (x_n - x^*)y - (x_n - x_{n-1})y = 0$$

so $L(f,P) \geq L(f,Q)$ as claimed. The proof that $U(f,P) \leq U(f,Q)$ is similar — see Figure 12 for an a illustration. □

29.8 Corollary. *Suppose $f : [a,b] \to \mathbb{R}$ is bounded. Then:*

1. *If P, Q are partitions of $[a,b]$ then $L(f,P) \leq U(f,Q)$*
2. $L_a^b f \leq U_a^b f$

Proof.

1. Let R be a common refinement of P and Q. It is clear from the defining formulae that $L(f,R) \le U(f,R)$ (each inf is no larger than the corresponding sup). Now, from the previous result,

$$L(f,P) \le L(f,R) \le U(f,R) \le U(f,Q)$$

2. By the first part, for any partition P of $[a,b]$, $L(f,P)$ is a lower bound for the set of all the upper sums of f. The greatest lower bound for these is $U_a^b f$, so $L(f,P) \le U_a^b f$. This says that $U_a^b f$ is an upper bound for the set of all lower sums of f; the least upper bound is $L_a^b f$ so $L_a^b f \le U_a^b f$.

□

With this fundamental result established, we can work out a few examples directly from the definition (which is rarely a good way to do it!)

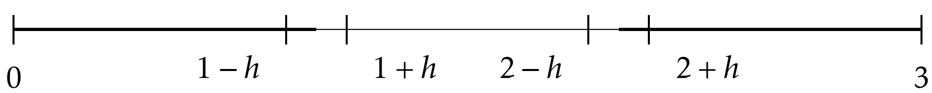
29.9 Example (A step function). Define $f : [0,3] \to \mathbb{R}$ by

$$f(x) = \begin{cases} 0 & \text{if } 0 \le x < 1 \\ 1 & \text{if } 1 \le x \le 2 \\ 0 & \text{if } 2 < x \le 3 \end{cases}$$

The idea here is to construct a partition which has very small intervals capturing the points of discontinuity. To that end, consider

$$P_h = (0, 1-h, 1+h, 2-h, 2+h, 3)$$

To make this a partition, we need $0 < h < 1/2$; this is not a problem, because we only care about very small h.



Now, we can write down the lower and upper sums:

$$L(f,P_h) = (1-h)\times 0 + 2h \times 0 + (1-2h)\times 1 + (2h)\times 0 + (1-h)\times 0 = 1-2h$$

$$U(f,P_h) = (1-h)\times 0 + 2h \times 1 + (1-2h)\times 1 + (2h)\times 1 + (1-h)\times 0 = 1+2h$$

The only difference between the upper and lower sums is on the two intervals of width $2h$, where the minimum and the maximum values of f are 0 and 1; on the longer intervals, f is constant.

Now, for all $h \in (0,1/2)$ we have

$$L_0^3 f \ge L(f,P_h) = 1-2h$$

which implies $L_0^3 f \geq 1$. Similarly, for all $h \in (0, 1/2)$ we have

$$U_0^3 f \leq U(f, P_h) = 1 + 2h$$

which implies $U_0^3 f \leq 1$. Because $L_0^3 f \leq U_0^3 f$ (Corollary 29.8(2)), we have

$$1 \leq L_0^3 f \leq U_0^3 f \leq 1$$

so $L_0^3 f = U_0^3 f = 1$; by definition, f is integrable and $\int_0^3 f = 1$.

29.10 Example. For some $k \in \mathbb{N}$, we shall try to integrate $f(x) = x^k$ from 0 to 1, by dividing $[0, 1]$ into N equal subintervals. Our partition is

$$P_N = (0, 1/N, 2/N, \ldots, 1)$$

Because f is increasing, on an interval $[(n-1)/N, n/N]$ of this partition, the infimum of f is attained at the left-hand endpoint and the supremum is attained at the right-hand endpoint. We can use this to find the lower and upper sums:

$$L(f, P_N) = \sum_{n=1}^{N} \frac{1}{N} \left(\frac{n-1}{N} \right)^k$$
$$U(f, P_N) = \sum_{n=1}^{N} \frac{1}{N} \left(\frac{n}{N} \right)^k$$

Here, the $1/N$ factor is the width of the subintervals and $[(n-1)/N]^k$ and $[n/N]^k$ are the minimal and maximal values of f over the subinterval $[(n-1)/N, n/N]$, attained at the left and right endpoints (see Exercise 12.5 for a more general application of this technique). We can slightly simplify:

$$L(f, P_N) = \frac{1}{N^{k+1}} \sum_{n=1}^{N} (n-1)^k$$
$$U(f, P_N) = \frac{1}{N^{k+1}} \sum_{n=1}^{N} n^k$$

At this point, we specialise to particular values of k. If $k = 1$, then we have an arithmetic series and

$$L(f, P_N) = \frac{1}{N^2} \frac{1}{2} N(N-1)$$
$$U(f, P_N) = \frac{1}{N^2} \frac{1}{2} N(N+1)$$

Incorporating the $1/N^2$ with the other factors gives

$$L(f, P_N) = \frac{1}{2}(1 - 1/N)$$
$$U(f, P_N) = \frac{1}{2}(1 + 1/N)$$

Now, we have

$$L(f,P_N) \le L_0^1 f \le U_0^1 f \le U(f,P_N)$$

and, as $N \to \infty$, both $L(f,P_N)$ and $U(f,P_N)$ converge to 1/2. We can therefore let $N \to \infty$ to conclude that

$$\frac{1}{2} \le L_0^1 f \le U_0^1 f \le \frac{1}{2}$$

and hence that

$$L_0^1 f = U_0^1 f = \frac{1}{2}$$

which by definition means that f is Riemann integrable and that

$$\int_0^1 f = \frac{1}{2}$$

It is now possible to continue in this way for other values of k, using e.g.

$$\sum_{n=1}^{N} n^2 = \frac{1}{6}N(N+1)(2N+1)$$

to integrate x^2, etc. See Exercise 12.3.

The reasoning at the end of this example calculation can be abstracted to work with general functions. The following is a kind of Cauchy criterion for integrals, able to prove integrability without using the actual value of the integral.

29.11 Theorem (Cauchy Criterion for Riemann Integrability)**.** *A bounded function $f : [a,b] \to \mathbb{R}$ is integrable if and only if for any $h > 0$ there exists a partition P of $[a,b]$ such that*

$$U(f,P) - L(f,P) < h$$

Proof. If this condition holds, then we can write

$$U_a^b f - L_a^b f \le U(f,P) - L(f,P) < h$$

This being true for all $h > 0$, we must have $U_a^b f - L_a^b f \le 0$. But we already know that $U_a^b f - L_a^b f \ge 0$, so we must have $U_a^b f = L_a^b f$.

Conversely, if f is integrable then for any $h > 0$ we can find partitions Q and R such that

$$L(f,Q) > \left(\int_a^b f\right) - h/2$$

$$U(f,R) < \left(\int_a^b f\right) + h/2$$

If we let P be a common refinement of Q and R (Theorem 29.6) then we have (Theorem 29.7)

$$L(f,P) \ge L(f,Q) > \left(\int_a^b f\right) - h/2$$

$$U(f,P) \le U(f,R) < \left(\int_a^b f\right) + h/2$$

Subtracting these gives $U(f,P) - L(f,P) < h$. □

30 Properties of the Integral

30.1 Theorem (Linearity). *Suppose $f,g:[a,b]\to\mathbb{R}$ are integrable and $s,t\in\mathbb{R}$. Then $sf+tg$ is also integrable and*

$$\int_a^b (sf+tg) = s\int_a^b f + t\int_a^b g$$

We tackle the proof in two stages.

First stage: $\int_a^b sf = s\int_a^b f$. Consider first the case $s>0$ where we have (Exercise 2.4) for any non-empty, bounded set Y

$$\inf\{sy : y\in Y\} = s\inf(Y)$$

and similarly for sup. This means that, for any partition P of $[a,b]$, we have $L(sf,P) = sL(f,P)$ and $U(sf,P) = sU(f,P)$. It follows that $L_a^b sf = sL_a^b f$ and $U_a^b sf = sU_a^b f$ and hence (given that f is integrable) that sf is integrable and $\int_a^b sf = s\int_a^b f$.

Now consider the same integral, with $s<0$. In this case we have (Exercise 2.4 again, combined with Exercise 2.7) for any bounded set Y,

$$\inf\{sy : y\in Y\} = s\sup(Y)$$

and similarly for sup. This means that, for any partition P of $[a,b]$, we have $L(sf,P) = sU(f,P)$ and $U(sf,P) = sL(f,P)$. It follows that $L_a^b sf = sU_a^b f$ and $U_a^b sf = sL_a^b f$ and hence (given that f is integrable) that sf is integrable and $\int_a^b sf = s\int_a^b f$.

The case $s=0$ is covered by Example 29.4. □

Second stage: $\int_a^b (f+g) = \int_a^b f + \int_a^b g$. Note first that for any set $X\subseteq[a,b]$

$$\inf_{x\in X}(f(x)+g(x)) \ge \inf_{x\in X} f(x) + \inf_{x\in X} g(x)$$

(because, for any $x\in X$, $f(x)\ge \inf_X f$ and $g(x)\ge\inf_X g$). Applied to lower sums, this gives for any partition P

$$L(f+g,P)\ge L(f,P)+L(g,P) \qquad (*)$$

Now, given $h>0$ there are partitions P_f and P_g of $[a,b]$ such that

$$L(f,P_f) > \left(\int_a^b f\right) - \frac{h}{2}$$
$$L(g,P_g) > \left(\int_a^b g\right) - \frac{h}{2}$$

Let P be a common refinement of P_f and P_g, so (Theorem 29.7),

$$L(f,P) > \left(\int_a^b f\right) - \frac{h}{2}$$
$$L(g,P) > \left(\int_a^b g\right) - \frac{h}{2}$$

Now,

$$\begin{aligned}
L_a^b(f+g) &\geq L(f+g,P) \\
&\geq L(f,P) + L(g,P) \text{ by } (*) \\
&> \left(\int_a^b f\right) - \frac{h}{2} + \left(\int_a^b g\right) - \frac{h}{2} \\
&= \left(\int_a^b f\right) + \left(\int_a^b g\right) - h
\end{aligned}$$

Since this is true for all $h > 0$, we must have

$$L_a^b(f+g) \geq \int_a^b f + \int_a^b g$$

An exactly similar argument gives

$$U_a^b(f+g) \leq \int_a^b f + \int_a^b g$$

We also have $L_a^b(f+g) \leq U_a^b(f+g)$, so

$$\int_a^b f + \int_a^b g \leq L_a^b(f+g) \leq U_a^b(f+g) \leq \int_a^b f + \int_a^b g$$

It now follows that

$$L_a^b(f+g) = U_a^b(f+g) = \int_a^b f + \int_a^b g$$

so $f+g$ is integrable on $[a,b]$ and

$$\int_a^b f + g = \int_a^b f + \int_a^b g$$

□

30.2 Corollary. *If f, g are integrable on $[a,b]$ and $f(x) \geq g(x)$ for $x \in [a,b]$ then $\int_a^b f \geq \int_a^b g$.*

Proof. Looking at $\int_a^b (f-g)$, all values of $f-g$ are non-negative, so all the upper and lower sums are non-negative, so the integral is non-negative, i.e.

$$\int_a^b (f-g) \geq 0$$

Now, by linearity,

$$\int_a^b f \geq \int_a^b g.$$

□

The remaining general decomposition property is:

30.3 Theorem. *Suppose $f : [a,b] \to \mathbb{R}$ and $a < c < b$. The the following are equivalent:*

1. *f is integrable over $[a,b]$*
2. *f is integrable over both $[a,c]$ and $[c,b]$*

In this case, we have

$$\int_a^b f = \int_a^c f + \int_c^b f$$

Proof of (1) $\Longrightarrow$ (2). Given $h > 0$, we can (Theorem 29.11) find a partition P of $[a,b]$ such that $U(f,P) - L(f,P) < h$. Refine this by adding (if not already present) the point c, giving a new partition

$$P' = (\underbrace{x_0}_{=a}, \dots, \underbrace{x_M}_{=c}, \dots, \underbrace{x_N}_{=b})$$

which also satisfies (Theorem 29.7) $U(f,P') - L(f,P') < h$. It also gives rise to partitions

$$Q = (\underbrace{x_0}_{=a}, \dots, \underbrace{x_M}_{=c}); \qquad R = (\underbrace{x_M}_{=c}, \dots, \underbrace{x_N}_{=b})$$

of $[a,c]$ and $[c,b]$, respectively. Note that (see Figure 11)

$$\begin{aligned} L(f,P') &= L(f,Q) + L(f,R) \\ U(f,P') &= U(f,Q) + U(f,R) \end{aligned}$$

from which it follows that

$$\Big(U(f,Q) - L(f,Q)\Big) + \Big(U(f,R) - L(f,R)\Big) = U(f,P') - L(f,P') < h$$

Since both terms on the LHS are non-negative, we have

$$U(f,Q) - L(f,Q) < h; \qquad U(f,R) - L(f,R) < h$$

showing that f is integrable on $[a,c]$ and on $[c,b]$ (Theorem 29.11 again). □

Proof of (2) $\Longrightarrow$ (1) and decomposition formula. Given integrability of f on $[a,c]$ and $[c,b]$ and any $h > 0$, we can find a partition Q of $[a,c]$ and a partition R of $[c,b]$ such that

$$L(f,Q) > \left(\int_a^c f\right) - \frac{h}{2}; \qquad L(f,R) > \left(\int_c^b f\right) - \frac{h}{2}$$

Suppose

$$Q = (\underbrace{x_0}_{=a}, \dots, \underbrace{x_M}_{=c}); \qquad R = (\underbrace{y_0}_{=c}, \dots, \underbrace{y_N}_{=b})$$

Then we can combine the two to give a partition P of $[a,b]$:

$$P = (\underbrace{x_0}_{=a}, \dots, \underbrace{x_M}_{=c}, y_1, \dots, \underbrace{y_N}_{=b})$$

We have

$$\begin{aligned} L_a^b f &\geq L(f,P) \\ &= L(f,Q) + L(f,R) \\ &> \left(\int_a^c f\right) - h/2 + \left(\int_c^b f\right) - h/2 \\ &= \left(\int_a^c f\right) + \left(\int_c^b f\right) - h \end{aligned}$$

Because this holds for all $h > 0$, we must have

$$L_a^b f \geq \int_a^c f + \int_c^b f$$

An exactly similar argument gives

$$U_a^b f \leq \int_a^c f + \int_c^b f$$

We also have $L_a^b f \leq U_a^b f$, so

$$\int_a^c f + \int_c^b f \leq L_a^b f \leq U_a^b f \leq \int_a^c f + \int_c^b f$$

It now follows that

$$L_a^b f = U_a^b f = \int_a^c f + \int_c^b f$$

so f is integrable on $[a,b]$ and

$$\int_a^b f = \int_a^c f + \int_c^b f$$

$\square$

30.4 Definition. Suppose $a > b$ and f is integrable on $[b,a]$. We define

$$\int_a^b f = -\int_b^a f$$

and also, for any function defined at a,

$$\int_a^a f = 0$$

With these conventions in place, we find (by enumerating all possible cases and applying Theorem 30.3) that

$$\int_a^b f = \int_a^c f + \int_c^b f$$

holds whatever order a, b, c appear on the line, provided two of the integrals are known to exist (in which case the third is guaranteed to exist), or f is integrable on the largest interval in question, namely $[\min\{a,b,c\}, \max\{a,b,c\}]$.

31 Integrating Continuous Functions: the Mean Value Theorem for Integrals and the Fundamental Theorem of Calculus

So far, we have hardly any examples of integrable functions. We shall now show that if $f : [a,b] \to \mathbb{R}$ is continuous, then it is integrable. This is surprisingly tricky, and requires a new idea.

31.1 Definition. Suppose I is an interval and $f : I \to \mathbb{R}$. We say that f is *uniformly continuous* on I if for any $\varepsilon > 0$ there exists $\delta_\varepsilon > 0$ such that if $x, y \in I$ with $|x-y| < \delta_\varepsilon$ then $|f(x) - f(y)| < \varepsilon$.

The difference between continuity and uniform continuity is that, for a uniformly continuous function, the dependence of δ on ε has to work across the whole interval. For a continuous function, the dependence of δ on ε can be different at each point.

31.2 Example. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = ax + b$, for some $a \neq 0$. Then $|f(x) - f(y)| = |a||x-y|$ so, if we let $\delta_\varepsilon = \varepsilon/|a|$ then, for any $x, y \in \mathbb{R}$ with $|x-y| < \delta_\varepsilon$, we have $|f(x)-f(y)| = |a||x-y| < |a|\varepsilon/|a| = \varepsilon$. This shows that f is uniformly continuous on $\mathbb{R}$.

Now define $g : \mathbb{R} \to \mathbb{R}$ by $g(x) = x^2$. If g were uniformly continuous on $\mathbb{R}$, then we would have relationships of the form $|g(x) - g(y)| < \varepsilon$ for all x, y with $|x-y| < \delta$; to see that this is impossible, note that for any $x \in \mathbb{R}$, $|x - (x + \delta/2)| < \delta$ but

$$|g(x) - g(x + \delta/2)| = |x^2 - (x + \delta/2)^2| = |\delta^2/4 + \delta x|$$

tends to ∞ as $x \to \pm\infty$, so certainly cannot be less than ε. This function is thus not uniformly continuous. What is happening here is that ε-δ relationship is different at every point, with δ having to be made smaller for larger values of x; a closer analysis using the Mean Value Theorem shows that, to prove continuity at x, we need $\delta < \varepsilon/(2|x|)$ for $x \neq 0$ and small ε.

In this example of non-uniform continuity, the interval ($\mathbb{R}$) is unbounded. This is not essential, e.g. $h(x) = 1/x$ on $(0,1)$ would also work. However, on closed, bounded intervals, such examples cannot be built.

31.3 Theorem. *Suppose $a, b \in \mathbb{R}$ with $a < b$ and $f : [a,b] \to \mathbb{R}$ is continuous. Then f is uniformly continuous on $[a,b]$.*

Proof. We shall prove the contrapositive: that if f is not uniformly continuous, then it is not continuous.

To that end, suppose f is not uniformly continuous. Then there exists $\varepsilon > 0$ such that for all $\delta > 0$ there exist $x, y \in [a,b]$ with $|x-y| < \delta$ but $|f(x)-f(y)| \geq \varepsilon$. In particular ($\delta = 1/n$) there exist $x_n, y_n \in [a,b]$ such that $|x_n - y_n| < 1/n$ but $|f(x_n) - f(y_n)| > \varepsilon$. By the Bolzano-Weierstrass Theorem (Corollary 11.6), $(x_n)_{n\in\mathbb{N}}$ has a convergent subsequence, say $x_{n_k} \to x$ as $k \to \infty$. Because $|x_{n_k} - y_{n_k}| < 1/n_k \to 0$ as $k \to \infty$, we also have $y_{n_k} \to x$ as $k \to \infty$. Now, as $k \to \infty$, we have x_{n_k} and y_{n_k} both tend to x but $|f(x_{n_k}) - f(y_{n_k})| > \varepsilon > 0$; we therefore cannot have $f(x_{n_k}) \to f(x)$ and $f(y_{n_k}) \to f(x)$, both of which would be true if f were continuous at x. We conclude that f is not continuous at x, and hence not continuous on I. □

31.4 Theorem. *Suppose $f : [a,b] \to \mathbb{R}$ is continuous. Then f is integrable.*

Proof. By the previous result, f is uniformly continuous. For any $\varepsilon > 0$, there exists $\delta_\varepsilon > 0$ such that if $x, y \in [a,b]$ then $|f(x)-f(y)| < \varepsilon$. Choose a partition $P = (x_0, x_1, \ldots, x_N)$ such that $x_n - x_{n-1} < \delta_\varepsilon$ for $1 \leq n \leq N$. For this partition, we have

$$U(f,P) - L(f,P) = \sum_{n=1}^{N} (x_n - x_{n-1}) \left(\max_{x\in[x_{n-1},x_n]} f(x) - \min_{x\in[x_{n-1},x_n]} f(x) \right)$$

Here, we have replaced the sup and inf of the definition with max and min, because f is continuous (Theorem 19.13) For any $x, y \in [x_{n-1}, x_n]$, we have (because $x_n - x_{n-1} < \delta_\varepsilon$) $|x-y| < \delta_\varepsilon$ and hence $|f(x)-f(y)| < \varepsilon$. In particular, the difference between the maximal and minimal values of f on $[x_{n-1}, x_n]$ is less than ε. We now have

$$U(f,P) - L(f,P) < \sum_{n=1}^{N} (x_n - x_{n-1})\varepsilon = (b-a)\varepsilon$$

So, given $h > 0$ we can put $\varepsilon = h/(b-a)$ in the above argument to find a partition P such that

$$U(f,P) - L(f,P) < \sum_{n=1}^{N} (x_n - x_{n-1})\varepsilon = (b-a)\varepsilon = h$$

and conclude from the Cauchy criterion (Theorem 29.11) that f is integrable. □

31.5 Corollary (Mean Value Theorem for Integrals)**.** *Suppose $a \neq b$, let*

$$I = [\min\{a,b\}, \max\{a,b\}]$$

and suppose $f : I \to \mathbb{R}$ is continuous. Then for some $c \in I$ we have

$$f(c) = \frac{1}{b-a} \int_a^b f$$

If $a < b$ then, by analogy with means (averages) of finite sets, the RHS is called the mean *of f over I.*

Proof. Consider first the case $a < b$. Applying the Really Rough Estimate (Corollary 29.3), replacing sup and inf by max and min because f is continuous, we see that

$$\min_{x\in[a,b]} f(x) \leq \frac{1}{b-a}\int_a^b f \leq \max_{x\in[a,b]} f(x)$$

(i.e., the mean is caught between the minimum and the maximum). Now, by the Intermediate Value Theorem (Corollary 19.9), f takes on every value between its minimum and its maximum; in particular, the mean of f is equal to $f(c)$ for some $c \in I$.

If $a > b$, the same argument with a and b exchanged yields $c \in I$ such that

$$f(c) = \frac{1}{a-b}\int_b^a f$$

Now, exchanging the limits on the integral changes the sign, which can be absorbed into the $a-b$ factor giving

$$f(c) = \frac{1}{b-a}\int_a^b f$$

□

31.6 Theorem (Fundamental Theorem of Calculus (I))**.** *Suppose $a, b \in \mathbb{R}$, $a < b$ and $f : [a,b] \to \mathbb{R}$ is continuous. Fix $p \in [a,b]$ and define $F : [a,b] \to \mathbb{R}$ by*

$$F(x) = \int_p^x f$$

Then F is differentiable on $[a,b]$ and $F' = f$.

Proof. Fix $x_0 \in [a,b]$. For any $h \neq 0$ such that $x_0 + h \in [a,b]$, we have

$$\begin{aligned}\frac{F(x_0+h)-F(x_0)}{h} &= \frac{1}{h}\left(\int_p^{x_0+h} f - \int_p^{x_0}\right)\\ &= \frac{1}{h}\int_{x_0}^{x_0+h} f \text{ (Definition 30.4)}\\ &= f(c(h))\end{aligned}$$

for some $c(h)$ between x_0 and $x_0 + h$, in particular with $|c(h) - x_0| \leq |h|$ (Mean Value Theorem for integrals, Corollary 31.5).

Now, as $h \to 0$, $c(h) \to x_0$ (because $|c(h) - x_0| \leq |h|$) and $f(c(h)) \to f(x_0)$ (because f is continuous). This shows that $F'(x_0) = f(x_0)$. □

31.7 Theorem (Fundamental Theorem of Calculus (II))**.** *Suppose $F : [a,b] \to \mathbb{R}$ is continuously differentiable (i.e. F' exists and is continuous on $[a,b]$), and let $f = F'$. Then*

$$\int_a^b f = \int_a^b F' = F(b) - F(a)$$

Proof. Let $\Phi(x) = \int_a^x f$ (for $x \in [a,b]$). Then, by the previous version of the FTC, $\Phi' = f$. We now have $F' = f$ and $\Phi' = f$, so $(F - \Phi)' = 0$; it follows (Theorem 23.6) that $F - \Phi$ is constant, so in particular $\Phi(b) - F(b) = \Phi(a) - F(a)$. Now, $\Phi(b) = \int_a^b f$ and $\Phi(a) = \int_a^a f = 0$, so $\int_a^b f = F(b) - F(a)$, as claimed. $\square$

We can now use any integration methods based on the FTC, e.g. integration by substitution (based on the chain rule) and integration by parts (based on the product rule). Note too that the second version respects the convention that reversing the limits of integration reverses the sign: exchanging a and b changes $F(b) - F(a)$ for $F(a) - F(b)$.

32 Part VI: Key Points

The rigorous description of integration is (as presented here) somewhat different from the way that limits have been presented, lacking the "given ε there exists δ" character and using instead the ideas of supremum and infimum.

Given a bounded function f on an interval $[a,b]$, the idea (Definition 29.1) is to divide the interval into N subintervals with endpoints $(x_n)_{n=0}^N$ written in increasing order; here $x_0 = a$, $x_N = b$ and the subintervals are $[x_{n-1}, x_n]$. We call this subdivision a *partition* of $[a,b]$. On each subinterval, we consider the largest rectangle fitting under the curve and the smallest rectangle sitting over the curve, and think of their areas as lower and upper bounds for the integral of f from x_{n-1} to x_n. Adding these areas together gives lower and upper bounds for the integral of f over the whole of $[a,b]$, called the *lower sum* and the *upper sum* of f using the given partition P.

$$L(f,P) = \sum_{n=1}^{N} (x_n - x_{n-1}) \inf_{x\in[x_{n-1},x_n]} f(x)$$
$$U(f,P) = \sum_{n=1}^{N} (x_n - x_{n-1}) \sup_{x\in[x_{n-1},x_n]} f(x)$$

Now, if we think of the lower sums as lower bounds for the integral, then their supremum is also a lower bound for the integral. Similarly, if we think of the upper sums as upper bounds for the integral, then their infimum is also an upper bound for the integral. These bounds (the best that can be obtained from the partitioning process) are the *lower integral* and *upper integral* of f from a to b:

$$L_a^b f = \sup\{L(f,P) : \ P \text{ is a partition of } [a,b]\}$$
$$U_a^b f = \inf\{U(f,P) : \ P \text{ is a partition of } [a,b]\}$$

Now, the idea is that the integral itself is somewhere between the lower and upper integrals. If these differ, then we can go no further, but if they coincide then their common value is the only candidate for the integral and we *define*

$$\int_a^b f = L_a^b f = U_a^b f$$

provided the lower and upper integrals are equal.

The main technical tool for understanding the integral is the idea of *refinement*: roughly, we refine a partition by adding to it one or more extra points, after which the associated lower sum increases and the associated upper sum decreases, making them both better approximations to the integral. See Definition 29.5, Theorem29.6 and Theorem29.7 for the precise statements.

Using the idea of refinement, we establish the (intuitively obvious, not quite so easy to prove) fact that for any bounded $f : [a,b] \to \mathbb{R}$ we have $L_a^b f \le U_a^b f$ (Corollary 29.8) and the Cauchy criterion that f is integrable on $[a,b]$ if and only if for any $h > 0$ we can find a partition P of $[a,b]$ such that $U_a^b f - L_a^b f < h$ (Theorem 29.11). This last result allows us to prove that a function is integrable, without actually evaluating the integral.

Thus defined, the integral has the basic properties that we expect. These can be summarised as:

Normalisation (Example 29.4) Thinking of a real number c as a constant function,

$$\int_a^b c = (b-a)c$$

Linearity (Theorem 30.1) If f, g are integrable on $[a,b]$ and $s,t \in \mathbb{R}$ then $sf + tg$ is integrable on $[a,b]$ and

$$\int_a^b sf + tg = s\int_a^b f + t\int_a^b g$$

Positivity, or order-preserving property (Corollary 30.2) If f, g are integrable on $[a,b]$ and $f(x) \geq g(x)$ for $x \in [a,b]$ then

$$\int_a^b f \geq \int_a^b g$$

Additivity, or domain decomposition (Theorem 30.3) If $a < c < b$ then

$$\int_a^b f = \int_a^c f + \int_c^b f$$

in the sense that if either side exists, then so does the other and they are equal. More generally, if for $a > b$ we define $\int_a^b f = -\int_b^a f$ and $\int_a^a f = 0$, then this holds for any a, b, c, in the sense that (e.g.) if any two of the integrals exist then so does the third and the identity holds.

The remaining and supremely important result is the Fundamental Theorem of Calculus (FTC) which is an exact statement of the idea that differentiation and integration are, in some sense, inverses to each other. In fact, the FTC comes in two versions, one for differentiating an integral, one for integrating a derivative.

FTC(I) (Theorem 31.6) If $f : [a,b] \to \mathbb{R}$ is continuous and $F : [a,b] \to \mathbb{R}$ is defined by $F(x) = \int_a^x f$ then F is differentiable and $F' = f$.

FTC(II) (Theorem 31.7) If $f : [a,b] \to \mathbb{R}$ is continuous and $F : [a,b] \to \mathbb{R}$ is differentiable on $[a,b]$ with $F' = f$ then $\int_a^b f = F(b) - F(a)$

Given this, as well as being able to integrate many examples directly, we can also use well-known techniques of integration (e.g. substitution, parts) by referring to the corresponding property of differentiation (e.g. chain rule, product rule).

The requirement that f is continuous can actually be relaxed a little, even in the world of the Riemann integral, and in the more powerful Lebesgue integration theory (taught in the third year) can be relaxed quite a lot.

C Riemann Sums

This is a technical appendix, for anyone who would like a deeper understanding of the way that integration works; in particular, none of this is examinable!

Instead of upper and lower sums, Riemann's equivalent definition of the integral involves *Riemann sums*, which you might have seen in Calculus. These are technically a little more complicated than Darboux's upper and lower sums, but have some advantages (they also work for complex- or vector-valued integration, and are more often seen in applications than Darboux's sums).

In this appendix, we give a brief sketch of how Riemann's definition can be connected to Darboux's.

C.1 Definition. Suppose $a, b \in \mathbb{R}$, $a < b$. If $P = (x_n)_{n=0}^N$ is a partition of $[a, b]$, a *compatible list of tags* is a sequence $(t_n)_{n=1}^N$ such that $x_{n-1} \leq t_n \leq x_n$ $(1 \leq n \leq N)$. The pair (P, T) is then called a *tagged partition* of $[a, b]$. A *Riemann sum* of a function $f : [a, b] \to \mathbb{R}$ is a sum of the form

$$R(f, P, T) = \sum_{n=1}^{N} (x_n - x_{n-1}) f(t_n)$$

So, instead of looking at the infimum and supremum of f over each subinterval, a Riemann sum looks at the value of f at an individual point in the subinterval; the tags are the points at which f is evaluated. There is an immediate connection between lower and upper sums, and Riemann sums:

C.2 Lemma. *Suppose $a, b \in \mathbb{R}$, $a < b$, and $f : [a, b] \to \mathbb{R}$ is bounded. Then for any partition P of $[a, b]$,*

$$\begin{aligned} L(f, P) &= \inf\{R(f, P, T) : T \textit{ is a list of tags compatible with } P\} \\ U(f, P) &= \sup\{R(f, P, T) : T \textit{ is a list of tags compatible with } P\} \end{aligned}$$

Proof. We consider only the first equation, for lower sums; the proof for upper sums is a minor modification.

Notation: let $P = (x_0, x_1, \ldots, x_N)$ and

$$m_n = \inf_{x \in [x_{n-1}, x_n]} f(x) \qquad (1 \leq n \leq N)$$

First we show that $L(f, P)$ is a lower bound for the set in question. If $T = (t_1, \ldots, t_N)$ is a list of tags compatible with P then for $1 \leq n \leq N$, we clearly have $f(t_n) \geq m_n$; multiplying by $(x_n - x_{n-1})$ and summing gives $L(f, P) \leq R(f, P, T)$, which is the bound we require.

Now we need to show that $L(f, P)$ is the greatest lower bound, i.e. that if $h > 0$ then $L(f, P) + h$ is not a lower bound. We know that $m_n + h/(b - a)$ is not a lower bound for f on $[x_{n-1}, x_n]$ so there exists $t_n \in [x_{n-1}, x_n]$ with $f(t_n) < m_n + h/(b - a)$. Let $T = (t_1, \ldots, t_N)$, then

$$R(f, P, T) = \sum_{n=1}^{N} (x_n - x_{n-1}) f(t_n) < \sum_{n=1}^{N} (x_n - x_{n-1}) m_n + \frac{h}{b - a} \sum_{n=1}^{N} (x_n - x_{n-1}) = L(f, P) + h$$

which indeed shows that $L(f, P) + h$ is not a lower bound. $\square$

We can now establish a basic connection between Riemann sums and the integral: integrability, with integral y, is equivalent to the existence of a partition, all of whose Riemann sums are close to y. Precisely:

C.3 Theorem. *Suppose $a, b \in \mathbb{R}$, $a < b$, $f : [a,b] \to \mathbb{R}$ and $y \in \mathbb{R}$. The following are equivalent:*

1. *f is integrable over $[a,b]$, with $\int_a^b f = y$.*
2. *For any $\varepsilon > 0$ there exists a partition P_ε of $[a,b]$ such that for any list T of tags compatible with P_ε we have $|R(f, P_\varepsilon, T) - y| < \varepsilon$.*

(1) $\implies$ (2). If $\varepsilon > 0$, we can use the Cauchy criterion for integrability (Theorem 29.11) to find a partition P_ε of $[a,b]$ such that $U(f, P_\varepsilon) - L(f, P_\varepsilon) < \varepsilon$. By definition of the integral, $L(f, P_\varepsilon) \le y \le U(f, P_\varepsilon)$ and by Lemma C.2, $L(f, P_\varepsilon) \le R(f, P_\varepsilon, T) \le U(f, P_\varepsilon)$. We now have that y and $R(f, P_\varepsilon, T)$ both lie between $L(f, P_\varepsilon)$ and $U(f, P_\varepsilon)$, which are less than ε apart. It follows that $|R(f, P_\varepsilon, T) - y| < \varepsilon$. □

(2) $\implies$ (1). We begin with a slightly technical point: to apply the definition of integrability, f must be bounded, so we need to check first that (2) implies boundedness. Choose a list $(t_n)_{n=1}^N$ of tags compatible with P. For m with $1 \le m \le N$ and $t \in [x_{m-1}, x_m]$, consider the list of tags formed from T by replacing t_m by some number $t \in [x_{m-1}, x_m]$. Put $\varepsilon = 1$ in (2) and expand the absolute value in $|R(f, P_\varepsilon, T) - y| < \varepsilon$ to give $y - 1 < R(f, P, T) < y + 1$. Now, we can expand the definition of $R(f, P, T)$ and separate out the $n = m$ term to give

$$y - 1 < S_m + (x_m - x_{m-1}) f(t) < y + 1$$

where

$$S_m = \sum_{\substack{n=1 \\ n \ne m}}^{N} (x_n - x_{n-1}) f(t_n)$$

Now we can find upper bounds on $f(t)$ for $t \in [x_{m-1}, x_m]$:

$$\frac{y - 1 - S_m}{x_m - x_{m-1}} < f(t) < \frac{y + 1 - S_m}{x_m - x_{m-1}}$$

Since f is bounded on each of the subintervals $[x_{m-1}, x_m]$, it is bounded on $[a,b]$.

Having disposed of that, the rest of the argument is simpler:

Given $\varepsilon > 0$, let P_ε be as in (2). By Lemma C.2, $y - \varepsilon \le L(f, P_\varepsilon) \le U(f, P_\varepsilon) \le y + \varepsilon$. By definition of lower and upper integral, $y - \varepsilon \le L_a^b f \le U_a^b f \le y + \varepsilon$. Combining these inequalities gives $|L_a^b f - y| \le \varepsilon$ and $|U_a^b f - y| \le \varepsilon$. Since this is true for all $\varepsilon > 0$, we have $L_a^b f = U_a^b f = y$, so f is integrable over $[a,b]$ and $\int_a^b f = y$. □

The usual definition of integrability using Riemann sums looks somewhat different from this, and uses the idea of the mesh of a partition: the length of the smallest interval.

C.4 Definition. Suppose $P = (x_0, x_1, \dots, x_N)$ is a partition of an inteval $[a, b]$. The *mesh* of the partition is defined by

$$\text{mesh}(P) = \max_{1 \le n \le N} (x_n - x_{n-1})$$

With this notation, we can give the classical definition of Riemann integral, framed here as a theorem because we are taking the Darboux formulation as the definition.

C.5 Theorem. *Suppose $a, b \in \mathbb{R}$, $a < b$, $f : [a, b] \to \mathbb{R}$ and $y \in \mathbb{R}$. The following are equivalent:*

1. *f is integrable over $[a, b]$, with $\int_a^b f = y$.*
2. *For any $\varepsilon > 0$ there exists δ_ε such that if (P, T) is a tagged partition of $[a, b]$ with* $\text{mesh}(P) < \delta_\varepsilon$ *then $|R(f, P, T) - y| < \varepsilon$.*

C.6 Remark.

1. This shows that the integral can be approached by Riemann sums using any partition of sufficiently small mesh, e.g. N equal-width subintervals for large N. It follows that the same holds for upper and lower sums; this is not immediately apparent from the Darboux presentation, where the nature of the partitions giving close lower and upper sums is not clear.
2. The order relation on the codomain of f is not expicitly used here, only the absolute value function in the inequality $|R(f, P, T) - y| < \varepsilon$. This makes this formulation more flexible: it can be used to describe the integral of complex-valued functions, where $|\cdot|$ represents complex magnitude, or of vector-valued functions, where $|\cdot|$ is replaced by vector length, or indeed in much more general situations (Banach spaces).

Half of the proof is easy.

Proof of Theorem C.5, (2) $\Longrightarrow$ (1). Given $\varepsilon > 0$, any partition with mesh less than δ_ε from (2) will work as P_ε in (2) of Theorem C.3. □

The other half is less easy, and would merit a "more demanding proof" label if it wasn't already in a technical appendix!

Proof of Theorem C.5, (1) $\Longrightarrow$ (2). Note first that f is bounded: this can be proved exactly as in the first part of the proof of (2) $\Longrightarrow$ (1) in Theorem C.3. Choose $C > 0$ such that $|f(x)| \le C$ for all $x \in [a, b]$.

Let $y = \int_a^b f$ and suppose $\varepsilon > 0$. By the Cauchy criterion for integrability (Theorem 29.11) there is a partition $P_\varepsilon = (x_0, x_1, \dots, x_N)$ of $[a, b]$ such that $U(f, P_\varepsilon) - L(f, P_\varepsilon) < \varepsilon/2$.The use of $\varepsilon/2$ is not deeply important here, but gives some elbow-room to allow the final estimate for $|R(f, P, T) - y|$ to be made smaller than ε. By definition of the integral, $L(f, P_\varepsilon) \le y \le U(f, P_\varepsilon)$. Let

$$\mu = \min_{1 \le n \le N} (x_n - x_{n-1})$$

(NB this is the minimum subinterval width, not the mesh!) and, for $1 \leq n \leq N$, let

$$c_n = \inf_{x \in [x_{n-1}, x_n]} f(x); \qquad C_n = \sup_{x \in [x_{n-1}, x_n]} f(x)$$

Now suppose $P = (y_0, y_1, \ldots, y_M)$ is any partition of $[a,b]$ with $\mathrm{mesh}(P) < \mu$. An interval $[y_{m-1}, y_m]$ from this partition must either:

1. be contained in a (uniquely determined) interval of P_ε; or

2. not be contained in any single interval of P_ε, but be contained in the union of two (uniquely determined) adjacent intervals of P_ε.

This is because $[y_{m-1}, y_m]$ is contained in $[a,b]$, which is the union of all the intervals from P_ε, and $[y_{m-1}, y_m]$ cannot intersect two non-adjacent intervals of P_ε without containing points more than μ apart, which would contradict $y_m - y_{m-1} \leq \mathrm{mesh}(P) < \mu$.

We encode this information by defining, for $1 \leq m \leq M$,

$$k_m = \begin{cases} n & \text{if } [y_{m-1}, y_m] \subseteq [x_{n-1}, x_n] \\ 0 & \text{otherwise.} \end{cases}$$

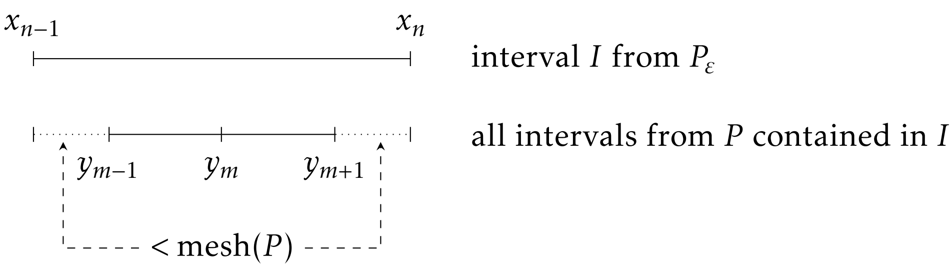
Now let $T = (t_1, \ldots, t_M)$ be a list of tags compatible with P and consider the Riemann sum

$$R(f, P, T) = \sum_{m=1}^{M} (y_m - y_{m-1}) f(t_m)$$

We break this up by the intervals from P_ε that contain those from P:

$$R(f, P, T) = \left[\sum_{n=1}^{N} \sum_{\substack{m=1 \\ k_m = n}}^{M} (y_m - y_{m-1}) f(t_m) \right] + \sum_{\substack{m=1 \\ k_m = 0}}^{M} (y_m - y_{m-1}) f(t_m) \tag{*}$$

In the double sum, the outer sum is over intervals of P_ε; the inner sum is over intervals of P contained in each interval of P_ε. The remaining single sum covers the intervals of P not contained in any interval of P_ε, i.e those straddling two intervals of P_ε. The nth term from the outer sum of the double sum in (*) can be illustrated as follows:



The top line represents a subinterval $[x_{n-1}, x_n]$ from the partition P_ε; the bottom line represents all the subintervals from P contained in it. The total length of the intervals from P is clearly no more than $x_n - x_{n-1}$. The dotted sections to the left and right of the intervals from P must have length less than $\operatorname{mesh}(P)$, otherwise the next interval from P would also have been contained in $[x_{n-1}, x_n]$; the length of the intervals from P is therefore greater than $x_n - x_{n-1} - 2\operatorname{mesh}(P)$.

So far, we have assumed that $\operatorname{mesh}(P) < \mu$. For the next part of the argument, we need the stronger assumption that $\operatorname{mesh}(P) < \mu/2$, so $x_n - x_{n-1} - 2\operatorname{mesh}(P) > 0$.

Now look at the inner sum from the double sum in (*):

$$S_n = \sum_{\substack{m=1\\k_m=n}}^{M} (y_m - y_{m-1}) f(t_m)$$

Here, $t_m \in [y_{m-1}, y_m] \subseteq [x_{n-1}, x_n]$ so $f(t_m) \le C_n$ and we can estimate this above by

$$S_n \le C_n \sum_{\substack{m=1\\k_m=n}}^{M} (y_m - y_{m-1}) \le C_n (x_n - x_{n-1})$$

because, as observed, the sum of the lengths of $[y_m, y_{m-1}]$ contained in $[x_{n-1}, x_n]$ is no more than $x_n - x_{n-1}$. Similarly, we can estimate S_n below by

$$S_n \ge c_n (x_n - x_{n-1} - 2\operatorname{mesh}(P))$$

(this is where the restriction $\operatorname{mesh}(P) < \mu/2$ is used). Now we can estimate the double sum from (*) above by

$$S_n \le \sum_{n=1}^{N} c_n (x_n - x_{n-1}) = U(f, P_\varepsilon)$$

which is our first connection to a Darboux sum. We can also estimate it below by

$$\begin{aligned} S_n \ge \sum_{n=1}^{N} c_n (x_n - x_{n-1} - 2\operatorname{mesh}(P)) &= L(f, P_\varepsilon) - 2\operatorname{mesh}(P) \sum_{n=1}^{N} c_n \\ &\ge L(f, P_\varepsilon) - 2CN \operatorname{mesh}(P) \end{aligned}$$

Finally, we look at the remaining single sum from (*):

$$S_0 = \sum_{\substack{m=1\\k_m=0}}^{M} (y_m - y_{m-1}) f(t_m)$$

When $k_m = 0$, this means that $[y_{m-1}, y_m]$ straddles two, uniquely determined, adjacent intervals from P_ε. But there are only N pairs of adjacent intervals from P_ε, so there are only N terms in this sum. Each term is bounded below and above by $-C\operatorname{mesh}(P)$ and $C\operatorname{mesh}(P)$ respectively, so we have

$$-CN \operatorname{mesh}(P) \le S_0 \le CN \operatorname{mesh}(P)$$

All that remains is to stitch together the estimate for the various parts of (*) and find a suitable δ_ε. We have

$$L(f,P_\varepsilon) - 3CN\,\mathrm{mesh}(P) \le R(f,P,T) \le U(f,P_\varepsilon) + CN\,\mathrm{mesh}(P)$$

provided $\mathrm{mesh}(P) < \mu/2$. We also have $L(f,P_\varepsilon) \le y \le U(f,P_\varepsilon)$ so, since $C > 0$, we have

$$L(f,P_\varepsilon) - 3CN\,\mathrm{mesh}(P) \le y \le U(f,P_\varepsilon) + CN\,\mathrm{mesh}(P)$$

So, $R(f,P,T)$ and y both lie in an interval of width $U(f,P_\varepsilon) - L(f,P_\varepsilon) + 4CN\,\mathrm{mesh}(P)$; it follows that

$$|R(f,P,T) - y| \le U(f,P_\varepsilon) - L(f,P_\varepsilon) + 4CN\,\mathrm{mesh}(P) < \frac{\varepsilon}{2} + 4CN\,\mathrm{mesh}(P)$$

where the $\varepsilon/2$ comes from the original choice of the partition P_ε. To get a bound of ε, as required, we now need to make $4CN\,\mathrm{mesh}(P) < \varepsilon/2$, i.e. $\mathrm{mesh}(P) < \varepsilon/(8CN)$. This gives us δ_ε: taking into account the requirement $\mathrm{mesh}(P) < \mu/2$ to make the above arguments work,

$$\delta_\varepsilon = \min\left\{\frac{\mu}{2}, \frac{\varepsilon}{8CN}\right\}$$

□

Exercises 1

Questions (or question parts) marked with a star ★ constitute Assignment 1; please submit on Moodle by the due date. Assignments make up 5% of the module mark.

Here are four basic properties of finite sums:

$$\text{(A)}\quad \sum_{j=p}^{q}(\alpha a_j+\beta b_j)=\alpha\sum_{j=p}^{q}a_j+\beta\sum_{j=p}^{q}b_j \qquad \text{(B)}\quad \sum_{j=p+k}^{q+k}a_j=\sum_{j=p}^{q}a_{j+k}$$

$$\text{(C)}\quad \sum_{j=p}^{q}a_j=a_p+\left(\sum_{j=p+1}^{q}a_j\right)=\left(\sum_{j=p}^{q-1}a_j\right)+a_q \qquad \text{(D)}\quad \sum_{j=p}^{q}(a_{j+1}-a_j)=a_{q+1}-a_p$$

1.1. Property (A) (called *linearity*) is analogous to the familiar fact from calculus that

$$\int_p^q \alpha f(x)+\beta g(x)\,\mathrm{d}x=\dots$$

What is the missing RHS? What is the fact from calculus corresponding to (B)? How about (D)? ((C) doesn't have a calculus analogue.)

★ 1.2. The geometric series: fix some real number $x\neq 1$ and let

$$S=\sum_{j=p}^{q}x^j$$

Now, indicating in your solution where properties (A)–(C) above are used, show that

$$xS=S+x^{q+1}-x^p$$

Solve this equation to find the well-known formula for S. What happens if $x=1$?

1.3. The arithmetic series: let $a_j=j(j-1)$. Use property (D) (and property (A), in a supporting role) to find a simple, and probably familiar, formula for

$$\sum_{j=1}^{n}j$$

★ 1.4. By converting the summand into partial fractions, use (D) to find a simple formula for

$$\sum_{j=2}^{n}\frac{1}{j(j-1)}$$

Verify that your answer is correct in the simplest possible case, $n=2$. *If your formula does not give the correct answer, you might have made a sign or offset error, both of which are very easy to make in this question. Check that you haven't confused a_j with $-a_j$ or a_{j+1} with a_j.*

⚠ 1.5. Binomial coefficients $\binom{n}{j}$ are defined recursively for $n, j \in \mathbb{N}_0$ with $0 \leq j \leq n$ by

$$\binom{n}{0} = \binom{n}{n} = 1; \qquad \binom{n+1}{j} = \binom{n}{j-1} + \binom{n}{j}$$

(this is just Pascal's triangle, expressed in formulae). Write out a proof of the Binomial Theorem: if $n \in \mathbb{N}_0$ and $a, b \in \mathbb{R}$ then

$$(a+b)^n = \sum_{j=0}^{n} \binom{n}{j} a^j b^{n-j}$$

You should work by induction on n, indicating where properties (A)–(C) are used.

★ Hand in.
⚠ Harder.

Exercises 1: Suggested Solutions

1.1. The complete formula is:

$$\int_p^q \alpha f(x) + \beta g(x)\,\mathrm{d}x = \alpha \int_p^q f(x)\,\mathrm{d}x + \beta \int_p^q g(x)\,\mathrm{d}x$$

The most obvious analogue to (B) is

$$\int_{p+k}^{q+k} f(x)\,\mathrm{d}x = \int_p^q f(x+k)\,\mathrm{d}x$$

which is a simple example of integration by substitution: x has been changed to $x+k$, and the limits altered appropriately; because $\mathrm{d}(x+k)/\mathrm{d}x = 1$, there is no further change. The full formula for a general substitution could also be said to be the calculus analogue of (B); more general substitution rules don't really work for sums because we need to map integers to integers, in continuous blocks.

The analogue of (D) is possibly less obvious. It's the Fundamental Theorem of Calculus:

$$\int_p^q f'(x)\,\mathrm{d}x = f(q) - f(p)$$

In the sum version, we take a difference of consecutive terms instead of a derivative (and we end up with a_{q+1}, not a_q).

1.2. Start with

$$xS = x\sum_{j=p}^{q} x^j = \sum_{j=p}^{q} xx^j = \sum_{j=p}^{q} x^{j+1}$$

using property (A) (with only one term; e.g. let $\alpha = x$ and $\beta = 0$). Now, we want to manipulate the RHS until we can find a copy of S inside it, to give us an equation to solve for S.

First, we use property (B) to change the x^{j+1} back to x^j, at the cost of changing the limits. This makes the summand on the RHS the same as that in S, namely x^j.

$$= \sum_{j=p+1}^{q+1} x^j$$

Now, we can take one term off the end using property (C) to make the upper limits match:

$$= \left(\sum_{j=p+1}^{q} x^j\right) + x^{q+1}$$

Next, we add on an extra term to recreate the missing $j = p$ term; having added it, we need to subtract it to keep our equality in order

$$= x^p + \left(\sum_{j=p+1}^{q} x^j\right) + x^{q+1} - x^p$$

Finally, we can combine the first two terms using (C) again:

$$= \left(\sum_{j=p}^{q} x^j \right) + x^{q+1} - x^p = S + x^{q+1} - x^p$$

This is the equation we were asked to derive. Solved for S, it yields

$$S = \frac{x^{q+1} - x^p}{x - 1}$$

which is the standard geometric sum formula. Note that $x = 1$ was ruled out in the question, so the division by $x - 1$ is valid.

In the case $x = 1$, the equation obtained above is correct by not useful: it just says $S = S$. However, we can easily enough evaluate the sum: $x^j = 1$ for all j and there are $q - p + 1$ terms so we have $S = q - p + 1$.

1.3. If we let $a_j = j(j-1)$ then

$$a_{j+1} - a_j = (j+1)j - j(j-1) = j(j+1-j+1) = 2j$$

Property (D) tells us that

$$\sum_{j=1}^{n} (2j) = a_{n+1} - a_1 = n(n+1)$$

amd dividing by 2 (formally, property (A)) gives the standard geometric sum formula

$$\sum_{j=1}^{n} j = \frac{1}{2} n(n+1)$$

1.4. Write $1/[j(j-1)]$ in partial fractions:

$$\frac{1}{j(j-1)} = \frac{1}{j-1} - \frac{1}{j}$$

This fits (D), with $a_j = -1/(j-1)$, $p = 2$, $q = n$:

$$\sum_{j=2}^{n} \frac{1}{j(j-1)} = \sum_{j=2}^{n} \left(\frac{1}{j-1} - \frac{1}{j} \right) = 1 - \frac{1}{n}$$

Check with $n = 2$: the sum consists of one term, with $j = 2$, and is $1/2$. The formula is $1 - n/2$ with $n = 2$, which is also $1/2$, so we have agreement.

1.5. This is somewhat similar to the geometric series question, but a bit more complicated. We begin by observing that the Binomial Theorem as stated is clearly true for $n = 0$ (or we could start at $n = 1$, it doesn't make much difference). Now assume that for some $n \in \mathbb{N}_0$ we have

$$(a+b)^n = \sum_{j=0}^{n} \binom{n}{j} a^j b^{n-j}$$

Then

$$(a+b)^{n+1} = (a+b)\sum_{j=0}^{n} \binom{n}{j} a^j b^{n-j} = \sum_{j=0}^{n} \binom{n}{j} a^{j+1} b^{n-j} + \sum_{j=0}^{n} \binom{n}{j} a^j b^{n+1-j}$$

using (A). Our task now is to recombine these sums to give us the binomial formula with n replaced by $n+1$. We'll need the powers of a and b to match in the two sums; to this end, use (B) to change the limits in the first sum to give

$$= \sum_{j=1}^{n+1} \binom{n}{j-1} a^j b^{n+1-j} + \sum_{j=0}^{n} \binom{n}{j} a^j b^{n+1-j}$$

(Why change the first sum, not the second? Because the second sum is a much closer match to the formula we're aiming for.) Now, these two sums can't be combined using (A) because the limits don't quite match: the first goes from 1 to $n+1$, the second from 0 to n. But the summation ranges have terms 1 to n in common. We therefore use both parts of (C) to break off the last term from the first sum and the first term from the second sum, to leave us with two sums over the same range, and two other terms.

$$= \left(\sum_{j=1}^{n} \binom{n}{j-1} a^j b^{n+1-j} \right) + \binom{n}{n} a^{n+1} b^0 + \binom{n}{0} a^0 b^{n+1} + \sum_{j=1}^{n} \binom{n}{j} a^j b^{n+1-j}$$

The two loose terms simplify to a^{n+1} and b^{n+1} and we can combine the sums using (A) to give

$$= b^{n+1} + \left(\sum_{j=1}^{n} \left[\binom{n}{j-1} + \binom{n}{j} \right] a^j b^{n+1-j} \right) + a^{n+1}$$

Now, the Pascal triangle formula allows us to combine the two binomial coefficients together to give $\binom{n+1}{j}$

$$= b^{n+1} + \left(\sum_{j=1}^{n} \binom{n+1}{j} a^j b^{n+1-j} \right) + a^{n+1}$$

Finally, if we put $j = 0$ and $j = n+1$ in the summand we get b^{n+1} and a^{n+1}, respectively: exactly the two terms left outside the sum. We can therefore use both parts of (C) to bring these back inside the sum:

$$(a+b)^{n+1} = \sum_{j=0}^{n+1} \binom{n+1}{j} a^j b^{n+1-j}$$

That is, the binomial formula is true when n is replaced by $n+1$. It therefore holds for all $n \in \mathbb{N}_0$ by induction.

Exercises 2

Questions (or question parts) marked with a star ★ constitute Assignment 2; please submit on Moodle by the due date. Assignments make up 5% of the module mark.

In the first three questions, the sets S^+ and S^- are probably most clearly expressed as a union of intervals (notation for intervals was introduced in Calculus last term).

★ 2.1. Let $P(x) = x^3 - x^2 - 2x$. By factorising P, find the sets

$$S^+ = \{x \in \mathbb{R} : P(x) > 0\}; \qquad S^- = \{x \in \mathbb{R} : P(x) < 0\}.$$

For both of these sets, answer the following (don't bother with a theoretical argument, just write down the answer):

(a) Does it have a maximum element? If so, what is it?

(b) Is it bounded above? If so, what is its supremum?

(c) Does it have a minimum element? If so, what is it?

(d) Is it bounded below? If so, what is its infimum?

2.2. In the previous question, what happens if $>$ and $<$ are changed to $\geq$ and $\leq$ in the definitions of S^+ and S^-?

2.3. Repeat the previous two questions, but with $P(x) = x^4 - 5x^2 + 4$.

2.4. Suppose $S \subseteq \mathbb{R}$ with $S \neq \emptyset$ and $c \in \mathbb{R}$ with $c > 0$. Let $cS = \{cx : x \in S\}$. Prove the following:

(a) If S has a maximal element, then $c\max(S)$ is the maximal element of cS.

(b) If S has a minimal element, then $c\min(S)$ is the minimal element of cS.

(c) If $b \in \mathbb{R}$ then b is an upper bound for S if and only if cb is an upper bound for cS.

(d) If $a \in \mathbb{R}$ then a is a lower bound for S if and only if ca is a lower bound for cS.

(e) If S is bounded above then $\sup(cS) = c\sup(S)$.

(f) If S is bounded below then $\inf(cS) = c\inf(S)$.

The last two parts are most easily done by using the earlier parts.

★ 2.5. Suppose S_1 and S_2 are bounded, non-empty subsets of $\mathbb{R}$. Show that

$$\sup(S_1 \cup S_2) = \max\{\sup(S_1), \sup(S_2)\}$$

State without proof the corresponding formula expressing $\inf(S_1 \cup S_2)$ in terms of $\inf(S_1)$ and $\inf(S_2)$.

2.6. Continuing from the previous question, show that there is no corresponding result for intersections by constructing examples of sets S_1, S_2, T_1, T_2 such that

$$\begin{aligned}
\sup(S_1) &= \sup(T_1) & \inf(S_1) &= \inf(T_1) \\
\sup(S_2) &= \sup(T_2) & \inf(S_2) &= \inf(T_2) \\
\sup(S_1 \cap S_2) &\neq \sup(T_1 \cap T_2) & \inf(S_1 \cap S_2) &\neq \inf(T_1 \cap T_2)
\end{aligned}$$

2.7. Suppose $S \subseteq \mathbb{R}$ is non-empty and bounded below and let

$$S' = \{-x : x \in S\}$$

(a) Show that b is an upper bound for S' if and only $-b$ is a lower bound for S.

(b) Show that $\sup(S') = -\inf(S)$.

(c) If S is bounded below, express $\inf(S')$ in terms of $\sup(S)$, justifying your answer.

⚠ 2.8. Suppose S_1 and S_2 are bounded, non-empty subsets of $\mathbb{R}$, and let

$$S = S_1 + S_2 = \{x + y : x \in S_1,\ y \in S_2\}.$$

Show that

$$\sup S = \sup S_1 + \sup S_2; \qquad \inf S = \inf S_1 + \inf S_2$$

Find the sup and inf of

$$S_1 - S_2 = \{x - y : x \in S_1,\ y \in S_2\}.$$

2.9. As part of the construction of $\sqrt{2}$, we noted that if $x, y > 0$ then $x < y$ if and only if $x^2 < y^2$. Deduce that $\sqrt{x} < \sqrt{y}$ if and only if $x < y$.

★ Hand in.
⚠ Harder.

Exercises 2: Suggested Solutions

2.1. We can factorise

$$P(x) = P(x) = x^3 - x^2 - 2x = x(x^2 - x - 2) = x(x+1)(x-2)$$

When $x < -1$, all three factors are negative, so $P(x) < 0$. When $x \in (-1, 0)$, one factor is positive and two negative, so $P(x) > 0$. When $x \in (0, 2)$, two factors are positive and one negative, so $P(x) < 0$. When $x > 2$, all factors are positive, so $P(x) > 0$. We thus have

$$S^+ = \{x \in \mathbb{R} : P(x) > 0\} = (-1, 0) \cup (2, \infty)$$
$$S^- = \{x \in \mathbb{R} : P(x) < 0\} = (-\infty, -1) \cup (0, 2)$$

It is now easy to read off the answers:

S^+ is bounded below, with infimum -1, but not above. It has no maximum or minimum.

S^- is bounded above, with supremum 2, but not below. It has no maximum or minimum.

2.2. Changing the inequalities from strict ($<$, $>$) to weak ($\leq$, $\geq$) attaches the zeros (roots) of P to both S^+ and S^-. This gives

$$S^+ = \{x \in \mathbb{R} : P(x) \geq 0\} = [-1, 0] \cup [2, \infty)$$
$$S^- = \{x \in \mathbb{R} : P(x) \leq 0\} = (-\infty, -1] \cup [0, 2]$$

This has no effect on boundedness, or on the supremum and infimum. However, it does change the maximum/minimum properties: S^+ now has a minimum element and S^- a maximum element, -1 and 2, respectively.

2.3. Here we can factorise

$$P(x) = x^4 - 5x^2 + 4 = (x^2 - 4)(x^2 - 1) = (x+2)(x+1)(x-1)(x+2)$$

With S^+ and S^- defined by strict inequalities ($<$, $>$), we apply similar reasoning to the earlier question to conclude that:

$$S^+ = (-\infty, -2) \cup (-1, 1) \cup (2, \infty); \qquad S^- = (-2, -1) \cup (1, 2)$$

from which we can read off the answers: S^+ is bounded neither above nor below, and consequently has no maximum, minimum, supremum or infimum.

S^- is bounded both above and below, with infimum -2 and supremum $+2$. It has neither a maximum nor a minimum.

Changing the inequalities from strict ($<$, $>$) to weak ($\leq$, $\geq$) attaches the zeros of P to both S^+ and S^-, giving

$$S^+ = (-\infty, -2] \cup [-1, 1] \cup [2, \infty); \quad S^- = [-2, -1] \cup [1, 2]$$

Once again, this has no effect on boundedness, suprema or infima, but it changes the maximum/minimum properties. S^+, being unbounded in both directions, still has no maximum or minimum, but S^- now has a minimum of -2 and a maximum of $+2$.

2.4.

(a) If $b = \max(S)$ is the maximal element of S, then $b \in S$ and $x \leq b$ for all $x \in S$. Multiplying by $c > 0$, we see that $cx \leq cb$ for all $x \in S$, which is to say that cb is an upper bound for cS. Since $b \in S$, $cb \in cS$, so cb is an upper bound for cS which is contained in cS; that is, cb is the maximal element of cS.

(b) Almost identical to the previous part: if $a = \min(S)$ then $a \in S$ and $a \leq x$ for all $x \in S$. It follows that $ca \in cS$ and $ca \leq cx$ for all $x \in S$, so ca is a lower bound for cS and, since it lies in cS, is the minimal element of cS.

(c) Suppose b is an upper bound for S. Then $x \leq b$ for all $x \in S$. Since $c > 0$ we can multiply by c to give $cx \leq cb$; since cx is an arbitrary member of cS, this shows that cb is an upper bound for cS.

Now suppose cb is an upper bound for cS, so for all $x \in S$, $cx \leq cb$. Divide by c to see that for all $x \in S$, $x \leq b$, showing that b is an upper bound for S.

(d) Almost identical to the previous part.

(e) Let U be the set of upper bounds of S. We see from (c) that the set of upper bounds of cS is exactly the set cU. Now, using (b) and the definition of sup as "least upper bound", we have

$$\sup(cS) = \min(cU) = c\min(U) = \sup(S).$$

(f) Again, very similar to the previous part. Let L be the set of lower bounds of S. Then using earlier parts and the definition of inf as "greatest lower bound"

$$\inf(cS) = \max(cL) = c\max(L) = c\inf(S).$$

2.5. Let $b_1 = \sup(S_1)$ and $b_2 = \sup(S_2)$ and suppose $x \in S_1 \cup S_2$. Then either $x \in S_1$, in which case $x \leq b_1$, or $x \in S_2$, in which case $x \leq b_2$. In either case, $x \leq \max\{b_1, b_2\}$, showing that $\max\{b_1, b_2\}$ is an upper bound for $S_1 \cup S_2$.

Here are two alternative ways of showing that it is the least upper bound.

(a) Suppose b is another upper bound for $S_1 \cup S_2$, so $x \leq b$ for all $x \in S_1 \cup S_2$. Then in particular $x \leq b$ for all $x \in S_1$ and $x \leq b$ for all $x \in S_2$, so b is an upper bound for both S_1 and S_2. Since b_1 and b_2 are the least upper bounds of these sets, $b_1 \leq b$ and $b_2 \leq b$; hence $\max\{b_1, b_2\} \leq b$. This shows that all other bounds for $S_1 \cup S_2$ are larger than $\max\{b_1, b_2\}$.

(b) Alternatively, suppose $b < \max\{b_1, b_2\}$. Then either $b < b_1$ or $b < b_2$. If $b < b_1$ then b is not an upper bound for S_1, so there exists $x \in S_1$ such that $x > b$. If $b < b_2$ then b is not an upper bound for S_2, so there exists $x \in S_2$ such that $x > b$. In either case, there exists $x \in S_1 \cup S_2$ such that $x > b$, showing that any $b < \max\{\sup(S_1), \sup(S_2)\}$ is not an upper bound for $S_1 \cup S_2$.

Whichever way you do the second part, together these show that $\max\{b_1, b_2\}$ is the least upper bound of $S_1 \cup S_2$, or in symbols

$$\sup(S_1 \cup S_2) = \max\{\sup(S_1), \sup(S_2)\}$$

The corresponding result for inf is that

$$\inf(S_1 \cup S_2) = \min\{\inf(S_1), \inf(S_2)\}$$

The question doesn't ask for a proof, but the argument is very similar to the one above:

Let $a_1 = \inf(S_1)$ and $a_2 = \inf(S_2)$ and suppose $x \in S_1 \cup S_2$. Then either $x \in S_1$, in which case $x \geq a_1$, or $x \in S_2$, in which case $x \geq a_2$. In either case, $x \geq \min\{a_1, a_2\}$, showing that $\min\{a_1, a_2\}$ is a lower bound for $S_1 \cup S_2$.

Again, here are two different ways of showing that it is the greatest lower bound.

(a) Suppose a is another lower bound for $S_1 \cup S_2$, so $a \leq x$ for all $x \in S_1 \cup S_2$. Then in particular $a \leq x$ for all $x \in S_1$ and $a \leq x$ for all $x \in S_2$, so a is a lower bound for both S_1 and S_2. But a_1 and a_2 are the greatest lower bounds for these sets, so $a \leq a_1$ and $a \leq a_2$; hence $a \leq \min\{a_1, a_2\}$. This shows that any other lower bound for $S_1 \cup S_2$ is less than $\min\{a_1, a_2\}$.

(b) Alternatively, suppose $a > \min\{a_1, a_2\}$. Then either $a > a_1$ or $a > a_2$. If $a > a_1$ then a is not a lower bound for S_1, so there exists $x \in S_1$ such that $x < a$. If $a > a_2$ then a is not a lower bound for S_2, so there exists $x \in S_2$ such that $x < a$. In either case, there exists $x \in S_1 \cup S_2$ such that $x < a$, showing that any $a > \min\{a_1, a_2\}$ is not a lower bound for $S_1 \cup S_2$.

These show that $\min\{a_1, a_2\}$ is the greatest lower bound of $S_1 \cup S_2$, or in symbols

$$\inf(S_1 \cup S_2) = \min\{\inf(S_1), \inf(S_2)\}$$

2.6. One way to construct such examples is to consider

$$S = \{-1, c, 1\}; \qquad T = (-1, 1)$$

Here c is any number between -1 and 1; S consists of just three points, while T is an open interval. Whatever the value of c, both S and T have supremum $+1$ and infimum -1. However, $S \cap T = \{c\}$ which has supremum and infimum c. We can construct the required examples by using two different values of c.

2.7.

(a)
$$\begin{aligned} b \text{ is an upper bound for } S' &\iff \text{for all } y \in S',\ y \leq b \\ &\iff \text{for all } x \in S,\ -x \leq b \\ &\iff \text{for all } x \in S,\ x \geq -b \\ &\iff -b \text{ is a lower bound for } S. \end{aligned}$$

(b) $\inf(S)$ is a lower bound for S so, by (a), $-\inf(S)$ is an upper bound for S'. Here are two different, equally valid, ways to show that $-\inf(S)$ is the least upper bound for S', i.e. that $\sup(S') = -\inf(S)$.

i. If b is an upper bound for S' then, by the above, $-b$ is a lower bound for S and hence (because the infimum is the greatest lower bound) $-b \leq \inf(S)$; multiply by -1 to give $b \geq -\inf(S)$. This shows that $-\inf(S)$ is the least upper bound of S', i.e. the supremum.

ii. Suppose $h > 0$. There exists $x \in S$ such that $x < \inf(S) + h$ (otherwise, $\inf(S) + h$ would be a lower bound for S, greater than the greatest lower bound). Now, $-x \in S'$ and $-x > -\inf(S) - h$, showing that $-\inf(S) - h$ is not an upper bound for S' and hence that $-\inf(S)$ is the least upper bound of S', i.e. the supremum.

(c) The easiest way to do this is to let $S'' = \{-x : x \in S'\}$ and apply part (b) to S' and S'', noting that $S'' = S$. In order to apply (b), note that if S is bounded above then S'' is bounded above so by (a) S' is bounded below. Now apply (b) with S replaced by S' we see that $\sup(S'') = -\inf(S')$. But $S'' = S$, so $\sup(S) = -\inf(S')$ and we conclude that $\inf(S') = -\sup(S)$. Alternatively, work through parts (a) and (b) again but with reversed inequalities for a direct proof.

2.8. Suppose $z \in S$, so $z = x + y$ for some $x \in S_1$ and $y \in S_2$. Since $x \leq \sup(S_1)$ and $y \leq \sup(S_2)$, we have $z \leq \sup(S_1) + \sup(S_2)$, showing that $\sup(S_1) + \sup(S_2)$ is an upper bound for S.

Now, we need to show that $\sup(S_1) + \sup(S_2)$ is the *least* upper bound for S. Here are two equally valid arguments:

Firstly, we can use the definition directly. Suppose b is an upper bound for S. Then for all $x \in S_1$ and $y \in S_2$, we have $x+y \leq b$, i.e. $x \leq b-y$; that is, each $b-y$ is an upper bound for S_1. Since $\sup(S_1)$ is the least upper bound for S_1 we must have $b - y \geq \sup(S_1)$, i.e. $y \leq b - \sup(S_1)$; that is, $b - \sup(S_1)$ is an upper bound for S_2. Since $\sup(S_2)$ is the least upper bound for S_2 we must have $b - \sup(S_1) \geq \sup(S_2)$ and hence $b \geq \sup(S_1)+\sup(S_2)$. This shows that any upper bound for S is greater tham or equal to $\sup(S_1) + \sup(S_2)$, so $\sup(S) = \sup(S_1) + \sup(S_2)$.

Alternatively, we can use Remark 3.16. For any $h > 0$ there exist $x \in S_1$ such that $x > \sup(S_1) - h/2$ and $y \in S_2$ such that $y > \sup(S_2) - h/2$. If we let $z = x + y$ then $z \in S$ and $z > \sup(S_1) + \sup(S_2) - h$. This shows that $\sup(S_1) + \sup(S_2)$ is the least upper bound of S.

To show that $\inf(S) = \inf(S_1) + \inf(S_2)$, we can write down a similar argument. Or, we can use Exercise 2.7: let

$$S_1' = \{-x : x \in S_1\}; \qquad S_2' = \{-x : x \in S_2\}; \qquad S' = \{-x : x \in S\}$$

and note that

$$S' = \{x + y : x \in S_1', y \in S_2'\}$$

By the result just established, $\sup(S') = \sup(S_1')+\sup(S_2')$. By Exercise 2.7, $-\inf(S) = -\inf(S_1) - \inf(S_2)$, so $\inf(S) = \inf(S_1) + \inf(S_2)$.

The remaining results can also be derived directly as above, or from Exercise 2.7:

$$S_1 - S_2 = S_1 + S_2'$$

where we know from Exercise 2.7 that

$$\sup(S_2') = -\inf(S_2); \qquad \inf(S_2') = -\sup(S_2)$$

(the second of these follows by exchanging the roles of S and S') In combination with the earlier parts of this question,

$$\sup(S_1 - S_2) = \sup(S_1) - \inf(S_2); \qquad \inf(S_1 - S_2) = \inf(S_1) - \sup(S_2)$$

2.9. This is easy (as well as useful):

$$\sqrt{x} < \sqrt{y} \iff (\sqrt{x})^2 < (\sqrt{y})^2 \iff x < y$$

using the result quoted in the question.

Exercises 3

Questions (or question parts) marked with a star ★ constitute Assignment 3; please submit on Moodle by the due date. Assignments make up 5% of the module mark.

Make free use of square roots wherever necessary, including the monotonicity property in Exercise 2.9. Note that the stars refer to individual lettered parts of questions, not whole questions.

3.1. Use the Archimedean property to establish the following (Remark 3.16 (3) is useful here):

★ (a) $\sup_{n\in\mathbb{N}} \dfrac{3n+1}{n+3} = 3$

(b) $\inf_{n\in\mathbb{N}} \dfrac{1}{n^2} = 0$

(c) $\sup_{n\in\mathbb{N}} \dfrac{n^2}{n^2+2} = 1$

(d) $\inf_{n\in\mathbb{Z}} \dfrac{1}{1+n^2} = 0$

Notation: $\sup_{n\in\mathbb{N}} a_n$ *means the same as* $\sup\{a_n : n \in \mathbb{N}\}$.

3.2. Use the Archimedean property to show that the following sets are not bounded above:

(a) $\{\sqrt{n} : n \in \mathbb{N}\}$

(b) $\{n^2 : n \in \mathbb{N}\}$

★ (c) $\left\{\dfrac{3n^3+4n+1}{n^2} : n \in \mathbb{N}\right\}$

(d) $\{n + 1/n : n \in \mathbb{N}\}$

3.3. Show by induction that $2^n > n$ for all $n \in \mathbb{N}$. Deduce from this and the property of Archimedes that the set

$$\{2^n : n \in \mathbb{N}\}$$

is not bounded above.

3.4. Each of the following sequences $(a_n)_{n\in\mathbb{N}}$ converges to a (finite) limit a as $n \to \infty$. In each case:

(i) Decide what you think the limit a is (using any technique you like);

(ii) Write down the inequality $|a_n - a| < \varepsilon$ (where $\varepsilon > 0$);

(iii) Find $N_\varepsilon \in \mathbb{N}$ such that if $n > N_\varepsilon$ then $|a_n - a| < \varepsilon$.

If you can make the last step work, then it doesn't matter how you arrived at the prospective limit a in the first step. Contrapositive, if you make the wrong guess in the first step then you'll never make the last step work!

(a) $a_n = \dfrac{1}{\sqrt{n}}$

(b) $a_n = \dfrac{n}{n+1}$

★ (c) $a_n = \dfrac{3n^2+1}{2n^2-1}$

(d) $a_n = \dfrac{1}{n} + \dfrac{1}{n^2}$

3.5. Try deliberately putting the wrong value of a into one of the examples from the previous question, and see how the last step goes wrong; e.g., try to show that $n/(n+1) \to 2$, and see how far you get!

3.6. Suppose $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ are real sequences,$a \in \mathbb{R}$, $b_n \to 0$ as $n \to \infty$ and $|a_n - a| \le b_n$ for all $n \in \mathbb{N}$. Show directly from the definition of convergence that $a_n \to a$ as $n \to \infty$. *This is a simpler, less powerful, version of the Sandwich Theorem, Theorem 7.1 in the notes.*

⚠ 3.7. Suppose $(a_n)_{n\in\mathbb{N}}$ is a real sequence and that ϕ is an injective map from $\mathbb{N}$ to itself, and consider the sequence $(a_{\phi(k)})_{k\in\mathbb{N}}$. Show that if $a_n \to a \in \mathbb{R}$ as $n \to \infty$ then $a_{\phi(k)} \to a$ as $k \to \infty$. *This shows that discarding terms from a sequence and changing the order of terms in a sequence does not change the limit. It includes Theorem 11.3 in the notes as a special case.*

★ Hand in.
⚠ Harder.

Exercises 3: Suggested Solutions

3.1. Strategy: the result mentioned in the question is that if b is an upper bound for S then it is the least upper bound if and only if for any $h > 0$ there exists $x \in S$ with $x > b - h$. In all these examples, the set is indexed by the natural numbers, so the inequality has the form $f(n) > b - h$ for some function f. This can be rearranged into the form $n > g(h)$ for some function g, which is satisfied for some $n \in \mathbb{N}$ by Archimedes property. A similar approach is followed for infima.

(a) We have $(3n+1)/(n+3) < (3n+9)/(n+3) = 3$, so 3 is an upper bound. To show that 3 is the least upper bound, we must show that for any $h > 0$ there exists $n \in \mathbb{N}$ such that $(3n+1)/(n+3) > 3-h$. This inequality can be rearranged into the equivalent form $n > 8/h - 3$; a suitable n now exists by Archimedes.

(b) Certainly 0 is a lower bound (since $n^2 > 0$, $1/n^2 > 0$). To show that is the greatest lower bound, it is enough to show that for any $h > 0$ there exists $n \in \mathbb{N}$ such that $1/n^2 < h$, or equivalently

(c) Clearly 1 is an upper bound, and to show that is the least upper bound we need to show that for any $h > 0$ there exists $n \in \mathbb{N}$ such that $n^2/(n^2+2) > 1-h$. This can be rearranged to the equivalent form $n^2 > 2/h - 2$. We want to take square roots here, but need to take care as well: the RHS might is negative if $h > 1$. In this case, however, any $n \in \mathbb{N}$ satisfies the inequality. If $h \leq 1$, then the RHS is non-negative and we can take square roots to give the equivalent $n > (2/h - 2)^{1/2}$, which is true for some $n \in \mathbb{N}$ by the Archimedean property.

(d) 0 is clearly a lower bound. To show that it is the greatest lower bound we need to show that for any $h > 0$ there exists $n \in \mathbb{Z}$ such that $1/(1 + n^2) < h$, or equivalently $n^2 > 1/h - 1$. Since $n^2 = (-n)^2$, we might as well assume that $n \geq 0$. As in the previous part, we need to eliminate the case $h > 1$, in which case the RHS is negative and any $n \in \mathbb{Z}$ will do the job. Assuming $h \leq 1$, we can take square roots to give the equivalent inequality $n > (1/h - 1)^{1/2}$, which is true for some $n \in \mathbb{N}$ by Archimedes property.

3.2. In each case, we posit the existence of an upper bound b for the given set, and derive from this an upper bound for $\mathbb{N}$, in contradiction to Archimedes' axiom. This contradiction shows that the given set has no upper bounds.

(a) If we have an upper bound b for $\{\sqrt{n} : n \in \mathbb{N}\}$, then $\sqrt{n} \leq b$ for all $n \in \mathbb{N}$. This implies $n \leq b^2$ for all $n \in \mathbb{N}$, which contradicts Archimedes.

(b) If we have an upper bound b for $\{n^2 : n \in \mathbb{N}\}$, then $n^2 \leq b$ for all $n \in \mathbb{N}$. This implies $n \leq \sqrt{b}$ for all $n \in \mathbb{N}$, which contradicts Archimedes.

(c) Suppose $(3n^3 + 4n + 1)/n^2 < b$ for all $n \in \mathbb{N}$. Rather than trying to solve this inequality exactly, observe that the LHS is greater than $3n$, so if this were true then we would have $3n < b$, i.e. $n < b/3$, for all $n \in \mathbb{N}$, immediately contradicting Archimedes.

(d) Suppose $n + 1/n \leq b$ for all $n \in \mathbb{N}$. Then $n \leq b$ for all $n \in \mathbb{N}$, immediately contradicting Archimedes.

3.3. Certainly $2^1 > 1$. If, for some n, $2^n > n$, then

$$2^{n+1} = 2.2^n > 2n = n + n \geq n + 1.$$

It follows that $2^n >$ for all n. Now suppose that there is an upper bound b for the set $\{2^n : n \in \mathbb{N}\}$, so $2^n \leq b$ for all $n \in \mathbb{N}$. Then $n < b$ for all $n \in \mathbb{N}$, so b is an upper bound for $\mathbb{N}$, in contradiction to Archimedes' axiom. This contradiction shows that no such bound b exists.

3.4.

(a)

i. $a_n = 1/\sqrt{n}$. Guess $a = 0$, because if n is large then so is $\sqrt{n}$, and hence $1/\sqrt{n}$ is small.

ii. The statement $|a_n - a| < \varepsilon$ now boils down to $1/\sqrt{n} < \varepsilon$.

iii. This can be rewritten as $n > 1/\varepsilon^2$. For any $\varepsilon > 0$, let N_ε be some natural number greater than $1/\varepsilon^2$ (note use of Archimedes property). Now, if $n > N_\varepsilon$ then $n > 1/\varepsilon^2$ so $1/\sqrt{n} < \varepsilon$ and hence $|a_n - a| < \varepsilon$.

The fact that we have found N_ε proves that the informal guess that started the process was correct.

(b)

i. $a_n = n/(n+1)$. Guess $a = 1$ because when n is large, 1 is small in comparison to n.

ii. The statement $|a_n - a| < \varepsilon$ now gives

$$\left|\frac{n}{n+1} - 1\right| < \varepsilon$$

iii. The expression inside $|\cdot|$ is always negative, so we can rewrite this as

$$1 - \frac{n}{n+1} < \varepsilon$$

which rearranges to the equivalent statement $n > 1/\varepsilon + 1$. For any $\varepsilon > 0$, let N_ε be some natural number greater than $1/\varepsilon - 1$ so, if $n > N_\varepsilon$, then $n > 1/\varepsilon - 1$ and $|a_n - a| < \varepsilon$

Again, finding N_ε justifies the informal way of guessing a.

(c)

i. $a_n = (3n^2 + 1)/(2n^2 - 1)$. Guess $a = 3/2$ because when n is large, ± 1 are small in comparison to n^2.

ii. The statement $|a_n - a| < \varepsilon$ now gives

$$\left|\frac{3n^2 + 1}{2n^2 - 2} - \frac{3}{2}\right| < \varepsilon$$

iii. We can rewrite this as

$$\left|\frac{5}{2(2n^2-1)}\right| < \varepsilon$$

The expression inside $|\cdot|$ is always positive for $n \in \mathbb{N}$, so we can rewrite this as

$$\frac{5}{2(2n^2-1)} < \varepsilon$$

Because both sides are positive, we can invert each side and move factors of 5 and 2 around to give the equivalent form

$$2n^2 - 1 > \frac{5}{2\varepsilon}$$

Now, still equivalently, we can add 1, divide by 2 and take the square root to reach

$$n > \sqrt{\frac{5}{4\varepsilon} + \frac{1}{2}}$$

We can therefore choose $N_\varepsilon \in \mathbb{N}$ such that

$$N_\varepsilon > \sqrt{\frac{5}{4\varepsilon} + \frac{1}{2}}$$

If $n > N_\varepsilon$ then the above calculation (because all steps are equivalences, i.e. "if and only if") shows that $|a_n - a| < \varepsilon$.

(d)

i. $a_n = 1/n + 1/n^2$. Guess $a = 0$ because both $1/n$ and $1/n^2$ are small when n is large.

ii. $|a_n - a| < \varepsilon$ now reads

$$\frac{1}{n} + \frac{1}{n^2} < \varepsilon$$

(no need for absolute value since the LHS is plainly positive)

iii. We could, at this point, derive an equivalent statement $n >$ (some function of ε), as in the previous parts. This is a bit messy, though: the answer turns out to be $n > (1 + \sqrt{1 + 4\varepsilon})/(2\varepsilon)$ (try it!). A better idea is to observe that $1/n + 1/n^2 \le 2/n$, so if we have $2/n < \varepsilon$ then $|a_n - a| < \varepsilon$. It is easy to arrange to have $2/n < \varepsilon$ if $n > N_\varepsilon$: just choose $N_\varepsilon > 2/\varepsilon$. This is an example of an important idea in Analysis: *estimation*. We can often save work by replacing a complicated quantity with a simpler upper or lower bound. Although it hasn't given us the smallest possible N_ε, it has given us an N_ε that works; and that's all we need.

3.5. Suppose we wrongly guess that $n/(n+1) \to 2$ as $n \to \infty$. Then we can certainly write down the inequality

$$\left|\frac{n}{n+1} - 2\right| < \varepsilon$$

and, since the quantity inside the absolute value is negative, this is the same as

$$2 - \frac{n}{n+1} < \varepsilon$$

which we can rearrange to give

$$(1-\varepsilon)n < \varepsilon - 2$$

Now we have a problem: we need to divide by $1-\varepsilon$, so we need to know its sign. We only really care about small ε, so the important case is $1-\varepsilon > 0$ which gives $n < (\varepsilon-2)/(1-\varepsilon)$. But we wanted to end up with a statement of the form $n >$ (some function of ε), not $n <$ (some function of ε). We have thus failed to prove that $n/(n+1) \to 2$ as $n \to \infty$ (and a good thing too!). Moreover, as the last inequality was equivalent to $|n/(n+1)-1| < \varepsilon$ (at least in the case $\varepsilon < 1$), we have proved that $n/(n+1)$ does not converge to 2.

3.6. Since $b_n \to 0$ as $n \to \infty$, we know from the definition of convergence that for any $\varepsilon > 0$ there exists $N_\varepsilon \in \mathbb{N}$ such that if $n > N_\varepsilon$ then $|b_n - 0| < \varepsilon$, i.e. $|b_n| \le \varepsilon$. Now

$$|a_n - a| \le b_n \le |b_n| < \varepsilon$$

so, again by definition, $a_n \to a$ as $n \to \infty$.

3.7. Suppose $\varepsilon > 0$. Then, because $a_n \to a$ as $n \to \infty$, there exists $N_\varepsilon \in \mathbb{N}$ such that if $n > N_\varepsilon$ then $|a_n - a| > \varepsilon$.

Now consider

$$S_\varepsilon = \{k \in \mathbb{N} : |a_{\phi(k)} - a| \ge \varepsilon\}$$

If $k \in S_\varepsilon$ then $\phi(k) \le N_\varepsilon$. Because ϕ is injective, there are at most N_ε elements of S_ε (each equation $\phi(k) = n$ has at most one solution, and there are N_ε natural numbers n with $n \le \mathbb{N}_\varepsilon$). Every finite set is bounded, so we can let $K_\varepsilon \in \mathbb{N}$ be an upper bound for S_ε.

Finally, if $k > K_\varepsilon$ then $k \notin S_\varepsilon$ then so $|a_{\phi(k)} - a| < \varepsilon$, showing that $a_{\phi(k)} \to a$ as $k \to \infty$.

Exercises 4

Questions (or question parts) marked with a star ★ constitute Assignment 4; please submit on Moodle by the due date. Assignments make up 5% of the module mark.

4.1. It was shown in Theorem 5.11 that if $a_n \to a$ as $n \to \infty$ and $k \in \mathbb{N}$ then $a_{n+k} \to a$ as $n \to \infty$ (the shift property). Prove the reverse implication: that if $a_{n+k} \to a$ as $n \to \infty$ then $a_n \to a$ as $n \to \infty$.

⚠ 4.2. Suppose $x > 1$. Use the binomial theorem to show that, for any $n \in \mathbb{N}$, $x^n \geq 1 + n(x-1)$. Use the Archimedes property to show that the set $\{x^n : n \in \mathbb{N}\}$ is unbounded. Does x^n converge to a limit as $n \to \infty$? What about x^{-n}? *This is marked as harder because it's intended at this point for a bare-handed attack from the definitions; with more technology, available later in the course, it becomes easier.*

4.3. (a) Suppose $(a_n)_{n\in\mathbb{N}}$ is a sequence such that $a_n \to a$ as $n \to \infty$. Show that $a_{2k} \to a$ as $k \to \infty$ and $a_{2k-1} \to a$ as $k \to \infty$. *This is very similar to Theorem 5.11(3). All of this is really to do with* subsequences, *which we will cover later in the course.*

(b) *A bit trickier*: prove the converse to the previous part, i.e. show that if $(a_n)_{n\in\mathbb{N}}$ is a sequence such that $a_{2k} \to a$ as $k \to \infty$ and $a_{2k-1} \to a$ as $k \to \infty$, then $a_n \to a$ as $n \to \infty$. *This will be used later, in the proof of the Leibniz alternating series test.*

★ 4.4. Suppose $f : \mathbb{R} \to \mathbb{R}$ is a function with the property that $|f(x) - f(y)| \leq |x - y|$ for all $x, y \in \mathbb{R}$ (e.g. sin and cos both have this property). Give a direct "ε" argument to show that if $(a_n)_{n\in\mathbb{N}}$ is a real sequence and $a_n \to a$ as $n \to \infty$ then $f(a_n) \to f(a)$ as $n \to \infty$

4.5. Suppose $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ are two sequences, one convergent, the other divergent. Show that $a_n + b_n$ diverges as $n \to \infty$.

4.6. Show by means of examples that if $a_n \to 0$ as $n \to \infty$ and $b_n \to +\infty$ as $n \to \infty$ then $a_n b_n$ could tend to zero, or tend to a finite, non-zero limit, or tend to $+\infty$, or be bounded and divergent.

4.7. Use the algebra of limits / combination rules to find the following:

★ (a) $\lim_{n\to\infty} \frac{2n+3}{n+1}$

(b) $\lim_{m\to\infty} \frac{1-2m^2}{2+m+m^2}$

★ (c) $\lim_{j\to\infty} \frac{aj+b}{cj^2+dj+e}$

(d) $\lim_{k\to\infty} \frac{k^2}{ck^3+dk^2+e}$

Here, a, b, c, d, e are real constants with $c \neq 0$; for simplicity, assume also that they are chosen so that no division by zero issues arise in defining the terms of the sequence.

4.8. Show that if $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ are real sequences, $a_n \to \infty$ as $n \to \infty$ and $a_n \leq b_n$ for all n then $b_n \to \infty$ as $n \to \infty$. *This analogue of the Sandwich Theorem for infinite limits appears without proof as Theorem 7.4 in the notes.* Show also that $-a_n \to -\infty$ as $n \to \infty$. *This was mentioned without proof after Definition 5.14 in the notes. For both of these, the algebraic work is almost trivial; it's all about carefully writing down the definitions and their immediate consequences.*

4.9. Similar to the previous question, but set up in slightly more generality: show that if $(a_n)_{n\in\mathbb{N}}$ is bounded below and $b_n \to +\infty$ as $n \to \infty$, then $a_n + b_n \to \infty$ as $n \to \infty$. What similar hypothesis on a_n would guarantee that if $b_n \to -\infty$ then $a_n + b_n \to -\infty$?

⚠ 4.10. Suppose $(a_j)_{j\in\mathbb{N}}$ is a real sequence and for $n \in \mathbb{N}$ let

$$s_n = \frac{1}{n}\sum_{j=1}^{n} a_j$$

(known as the *Cesàro means* of the a_j). Show that if $a_j \to 0$ as $j \to \infty$ then $s_n \to 0$ as $n \to \infty$. Deduce that if $a_j \to a$ as $j \to \infty$ then $s_n \to a$ as $n \to \infty$. Show also that if $a_j \to \infty$ as $j \to \infty$ then $s_n \to \infty$ as $n \to \infty$. Finally, give a concrete example of a divergent sequence $(a_j)_{j\in\mathbb{N}}$ such that the s_n converge as $n \to \infty$.

★ Hand in.
⚠ Harder.

Exercises 4: Suggested Solutions

4.1. Suppose $a_{n+k} \to a$ as $n \to \infty$ and $\varepsilon > 0$. There exists $N_\varepsilon \in \mathbb{N}$ such that if $n > N_\varepsilon$ then $|a_{n+k} - a| < \varepsilon$. If $n > N_\varepsilon + k$ then $n - k > N_\varepsilon$. We can therefore replace n by $n - k$ in the previous statement to see that $|a_{(n-k)+k} - a| < \varepsilon$, which is to say that $|a_n - a| < \varepsilon$. Since this is true for all $n > N_\varepsilon + k$, we have shown that $a_n \to a$ as $n \to \infty$.

4.2. Write $x = 1 + (x - 1)$. Expanding x^n using the binomial theorem gives

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

These terms are all positive, so we can estimate

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

Now, if we had a bound b such that $x^n \leq b$ for all $n \in N$ then we would have $1 + n(x-1) \leq b$ for all n; solving this, we see that $n \leq (b-1)/(x-1)$. So, existence of an upper bound for the x^n would lead to existence of an upper bound for $\mathbb{N}$ in contradiction to Archimedes. No such b can therefore exist.

Every convergent sequence is bounded, so x^n cannot converge to a finite limit as $n \to \infty$. However, the inequality $x^n \geq 1 + n(x-1)$ can be used to show that $x^n \to \infty$ as $n \to \infty$: given $K \in \mathbb{R}$, we can make $x^n > K$ by making n such that $1 + n(x-1) > K$, i.e. $n > (K-1)/(x-1)$. We can thus satisfy Definition 5.14 by choosing (Archimedes) $N_K \in \mathbb{N}$ with $N_K > (K-1)/(x-1)$.

It now follows from Theorem 5.15 that $x^{-n} \to 0$ as $n \to \infty$. Alternatively, we can prove this directly from the Archimedes property: given $\varepsilon > 0$, we want to make $|x^{-n} - 0| < \varepsilon$, or equivalently $x^{-n} < \varepsilon$. Because $x^n \geq 1 + n(x-1)$, it is enough to make $1/(1 + n(x-1)) < \varepsilon$. This is equivalent to $1 + n(x-1) > 1/\varepsilon$, which in turn is equivalent to $n > (1/\varepsilon - 1)/(x-1)$. Now we can finish in the usual way: given $\varepsilon > 0$, choose $N_\varepsilon \in \mathbb{N}$ with $N_\varepsilon\ (1/\varepsilon - 1)/(x-1)$. If $n > N_\varepsilon$, the above reasoning gives $|x^{-n} - 0| < \varepsilon$, showing that $x^n \to 0$ as $n \to \infty$.

4.3.

(a) Suppose $\varepsilon > 0$. Since $a_n \to a$ as $n \to \infty$, there exists $N_\varepsilon \in \mathbb{N}$ such that if $n > N_\varepsilon$ then $|a_n - a| < \varepsilon$. Now, if $k > N_\varepsilon$ then $2k > k > N_\varepsilon$, so $|a_{2k} - a| < \varepsilon$; this shows that $a_{2k} \to a$ as $k \to \infty$. Similarly, if $k > N_\varepsilon$ then $2k - 1 \geq k > N_\varepsilon$, so $a_{2k-1} \to a$ as $n \to \infty$.

(b) Suppose $\varepsilon > 0$. Since $a_{2k} \to a$ and $a_{2k-1} \to a$ as $k \to \infty$, there exist $K_\varepsilon^{(e)}$ and $K_\varepsilon^{(o)}$ such that if $k > K_\varepsilon^{(e)}$ then $|a_{2k} - a| < \varepsilon$ and if $k > K_\varepsilon^{(o)}$ then $|a_{2k-1} - a| < \varepsilon$.

Now suppose $n > N_\varepsilon := 2\max(K_\varepsilon^{(e)}, K_\varepsilon^{(o)})$. If n is even then $n/2 \in \mathbb{N}$ and $n/2 > K_\varepsilon^{(e)}$ so $|a_{2(n/2)} - a| < \varepsilon$. If n is odd then $(n+1)/2 \in \mathbb{N}$ and $(n+1)/2 > K_\varepsilon^{(o)}$, so $|a_{2(n+1)/2-1} - a| < \varepsilon$. In either case, this reduces to $|a_n - a| < \varepsilon$; since this is true for all $n > N_\varepsilon$, we have $a_n \to a$ as $n \to \infty$.

4.4. We are give $|f(x)-f(y) \leq |x-y|$, so $|f(a_n)-f(a)| \leq |a_n - a|$. For any $\varepsilon > 0$ there exists $N_\varepsilon \in \mathbb{N}$ such that if $n > N_\varepsilon$ then $|a_n - a| < \varepsilon$ and hence $|f(a_n) - f(a)| \leq |a_n - a| < \varepsilon$. This shows that $f(a_n) \to f(a)$ as $n \to \infty$.

4.5. Assume for definiteness that $a_n \to a$ and (b_n) diverges, and suppose for a contradiction that $a_n + b_n \to c$. Then $b_n = (a_n + b_n) - a_n \to c - a$; contradiction.

4.6. If $a_n = 1/n^2$ and $b_n = n$ then $a_n b_n = 1/n \to 0$.

If $a_n = 1/n$ and $b_n = n$ then $a_n b_n = n/n = 1 \to 1$

If $a_n = 1/n$ and $b_n = n^2$ then $a_n b_n = n \to \infty$

If $a_n = 1/n$ and $b_n = n(2 + (-1)^n)$ then $a_n b_n = 2 + (-1)^n$ which is bounded and divergent.

4.7.

(a) $\dfrac{2n+3}{n+1} = \dfrac{2+3/n}{1+1/n} \to \dfrac{2+0}{1+0} = 2$ as $n \to \infty$.

(b) $\dfrac{1-2m^2}{2+m+m^2} = \dfrac{1/m^2 - 2}{2/m^2 + 1/m + 1} \to \dfrac{0-2}{0+0+1} = -2$ as $m \to \infty$

(c) $\dfrac{aj+b}{cj^2+dj+e} = \dfrac{a/j + b/j^2}{c + d/j + e/j^2} \to \dfrac{0+0}{c+0+0} = 0$ as $j \to \infty$; note that $c \neq 0$, otherwise this does not work!

(d) $\dfrac{k^2}{ck^3+dk^2+e} = \dfrac{1/k}{c+d/k+e/k^3} \to \dfrac{0+0}{c+0+0} = 0$ as $k \to \infty$; note again that $c \neq 0$, otherwise this does not work!

4.8. Given $K \in \mathbb{R}$ there exists $N_K \in \mathbb{N}$ such that if $n > N_K$ then $a_n > K$. Since $b_n \geq a_n$, we have $b_n > K$. This shows that $b_n \to \infty$ as $n \to \infty$.

To show that $-a_n \to -\infty$, suppose $K \in \mathbb{R}$. Then there exists $N_K \in \mathbb{N}$ such that if $n > N_K$ then $a_n > -K$, and hence $-a_n < K$; this shows that $-a_n \to -\infty$ as $n \to \infty$.

4.9. Since $(a_n)_{n\in\mathbb{N}}$ is bounded below then there exists $a \in \mathbb{N}$ such that $a_n > a$ for all n. Since $b_n \to \infty$ as $n \to \infty$, for any $K > 0$ there exists $N_K \in \mathbb{N}$ such that if $n > N_K$ then $b_n > K - a$. It follows that $a_n + b_n > K$ if $n > N_K$, so $a_n + b_n \to +\infty$ as $n \to \infty$.

The corresponding result with $b_n \to -\infty$ as $n \to \infty$ is that if $(a_n)_{n\in\mathbb{N}}$ is bounded above then $a_n + b_n \to -\infty$ as $n \to \infty$. The proof is similar to the first part, or can be deduced from the first part by considering $-b_n$ and $-a_n - b_n$.

4.10. Suppose $a_j \to 0$ as $j \to \infty$. Given $\varepsilon > 0$, we want to show that, for sufficiently large n, $|s_n| < \varepsilon$. Firstly, there exists J_ε such that if $j > J_\varepsilon$, then $|a_j| < \varepsilon/2$ (don't worry about $\varepsilon/2$ vs. ε at this point; it represents me doing a rough calculation and realising that I need some elbow-room in the final steps, to get a neat ε out at the end). The points beyond J_ε will cause no problems in making s_n smaller than ε: roughly, if individual points are smaller than $\varepsilon/2$, then so is their average. The problem is the first J_ε points, over which we have no control. Start off by giving these points a name: let

$$S = \sum_{j=1}^{J_\varepsilon} a_j$$

For $n > J_\varepsilon$, we have

$$s_n = \frac{1}{n}\sum_{j=1}^{n} a_j = \frac{1}{n}\left(S + \sum_{j=J_\varepsilon+1}^{n} a_j\right)$$

We can estimate this using the triangle inequality:

$$|s_n| \le \frac{1}{n}\left(|S| + \sum_{j=J_\varepsilon+1}^{n} |a_j|\right)$$

and use the fact that $|a_j| < \varepsilon/2$ to estimate further

$$|s_n| \le \frac{1}{n}\left(|S| + (n - J_\varepsilon)\frac{\varepsilon}{2}\right) \tag{*}$$

Now, the RHS tends to $\varepsilon/2$ as $n \to \infty$ (algebra of limits). If we put $\varepsilon/2$ in the definition of that limit, we find N_ε such that if $n > N_\varepsilon$ then

$$\left|\frac{1}{n}\left(|S| + (n - J_\varepsilon)\frac{\varepsilon}{2}\right) - \frac{\varepsilon}{2}\right| < \varepsilon/2$$

(Again, don't worry about why $\varepsilon/2$ at this point; in a first draft, put ε, see what happens, fix it up later). Now, we can drop the absolute value on the LHS and add $\varepsilon/2$ to both sides to give

$$\frac{1}{n}\left(|S| + (n - J_\varepsilon)\frac{\varepsilon}{2}\right) < \varepsilon$$

But the LHS of this inequality is the RHS of (*) above so, provided $n > \max\{J_\varepsilon, N_\varepsilon\}$, we have $|s_n| < \varepsilon$. This shows that $s_n \to 0$ as $n \to \infty$. Now, you can see why I put $\varepsilon/2$ in the two uses of the limit definition: in a first draft of the argument above, using ε, I got 2ε at the end; I wanted ε, so I went back and changed ε to $\varepsilon/2$ so I ended up with ε at the end.

Now suppose $a_j \to a$ as $j \to \infty$. Then $a_j - a \to 0$ as $j \to \infty$. By the above result,

$$\frac{1}{n}\sum_{j=1}^{n}(a_j - a) \to 0$$

as $n \to \infty$. Subtracting a from each term subtracts a from the mean, so we have

$$\frac{1}{n}\left(\sum_{j=1}^{n} a_j\right) - a \to 0$$

and adding a to both sides gives (combination rules again)

$$\frac{1}{n}\sum_{j=1}^{n} a_j \to a$$

as $n \to \infty$. It is also possible to prove this directly, by a similar argument to the one in the first part of the question, but keeping track of the subtractions of a is a bit fiddly.

Next, suppose $a_j \to \infty$ as $n \to \infty$. This is a very similar argument to the first one. Given $K > 0$, we want $s_n > K$ for sufficiently large n. Instead of the $\varepsilon/2$ move for the zero limit, we use a $2K$ move: there exists $J_K \in \mathbb{N}$ such that if $j > J_K$ then $a_j > 2K$. Just as in the first part, the terms beyond J_K will not cause any trouble; but the terms before J_K are not under control (this is why we need the extra elbow-room of the $2K$). Let

$$S = \sum_{j=1}^{J_K} a_j$$

For $n > J_K$, we have

$$s_n = \frac{1}{n}\sum_{j=1}^{n} a_j = \frac{1}{n}\left(S + \sum_{j=J_K+1}^{n} a_j\right)$$

which we can estimate using $a_j > 2K$ for $j > J_K$:

$$s_n \geq \frac{1}{n}(S + (n - J_K)2K) \tag{**}$$

The RHS tends to $2K$ as $n \to \infty$. Putting $\varepsilon = K$ in the definition of that limit, we have N_K such that for $n > N_K$

$$\left|\frac{1}{n}(S + (n - J_K)2K) - 2K\right| < K$$

Expanding this to

$$-K \leq \frac{1}{n}(S + (n - J_K)2K) - 2K < K$$

and adding $2K$ to both sides of the LH inequality gives

$$K \leq \frac{1}{n}(S + (n - J_K)2K)$$

But, for $n > J_K$, the RHS is by (**) a lower bound for s_n. This gives, provided $n > \max\{J_K, N_K\}$, $s_n > K$, showing that $s_n \to \infty$ as $n \to \infty$.

For a concrete example, let $a_j = (-1)^j$, which diverges. Then

$$s_n = \frac{1}{n}\sum_{j=1}^{n}(-1)^j = \frac{1}{n}\frac{(-1)^{n+1} - (-1)}{-1-1} = \frac{1}{n}\frac{-1 + (-1)^{n+1}}{2}$$

Now, the right-hand fraction takes on only the values -1 (n odd) or 0 (n even). We therefore have the sandwich inequality

$$-\frac{1}{n} \leq s_n \leq 0$$

and it follows from the Sandwich Theorem that $s_n \to 0$ as $n \to \infty$. Or, without recourse to the Sandwich Theorem, given $\varepsilon > 0$, choose $N_\varepsilon \in \mathbb{N}$ with $N_\varepsilon > 1/\varepsilon$.

Exercises 5

Questions (or question parts) marked with a star ★ constitute Assignment 5; please submit on Moodle by the due date. Assignments make up 5% of the module mark.

5.1. Give a direct "ε" argument to show that if $a_n \geq 0$ and $a_n \to 0$ as $n \to \infty$ then $\sqrt{a_n} \to 0$ as $n \to \infty$. *This completes the proof of Theorem 7.3 in the notes.*

5.2. Suppose $x \in \mathbb{R}$. For any $n \in \mathbb{N}$, the interval $(x - 1/n, x + 1/n)$ contains a rational number q_n (Theorem 3.21). Show that $q_n \to x$ as $n \to \infty$. *This is another form of rational density: every real number is the limit of a sequence of rational numbers.*

★ 5.3. Let

$$a_n = \frac{\sin(n) - 2\cos(n)}{2n}; \qquad a = 0$$

Find a constant C such that the sandwiching inequality

$$|a_n - a| \leq \frac{C}{n} \qquad (n \in \mathbb{N})$$

(from which it follows that $a_n \to a$ as $n \to \infty$) is true. Here, sin and cos can be any two functions whose domain includes $\mathbb{N}$ that satisfy $|\sin(n)| \leq 1$ and $|\cos(n)| \leq 1$ for $n \in \mathbb{N}$.

5.4. Let

$$a_n = \frac{(n!)^2}{(2n)!}$$

for $n \in \mathbb{N}$. By expanding out all the factorials as products (e.g., $n! = 1.2 \ldots n$) and comparing terms in the numerator and denominator, show that $a_n \leq 1/2^n$. Deduce that $a_n \to 0$ as $n \to \infty$. What can we now say about $(2n)!/(n!)^2$?

⚠ 5.5. Double factorials: if $n \in \mathbb{N}$, define

$$n!! = n(n-2)(n-4)\ldots 1 \ (n \text{ odd}); \qquad n!! = n(n-2)(n-4)\ldots 2 \ (n \text{ even})$$

Show that if $x \in \mathbb{R}$ then $x^n/n!! \to 0$ as $n \to \infty$. *You might find the results from Exercise 4.3 useful here; if so, feel free to use them.*

5.6. Use combination rules and the hierarchy of limits to evaluate the following. *Don't try to use an "ε" argument!*

★ (a) $\displaystyle\lim_{n\to\infty} \frac{3^n + 2^n}{3^n + n^2}$

(b) $\displaystyle\lim_{j\to\infty} \frac{2^j + j^3}{j! + 2^j + 3j}$

★ (c) $\displaystyle\lim_{m\to\infty} \frac{m^m + m!}{m^{m+1} + 2m}$

(d) $\displaystyle\lim_{k\to\infty} \frac{x^{k+1} - k^3}{x^k + 1} \quad (|x| > 1)$

5.7. Which of the following sequences (for $n \in \mathbb{N}$) are increasing, decreasing, strictly increasing, strictly decreasing? Which are bounded? Which converge?

$$a_n = \frac{1}{n^2} \qquad b_n = 1 + n^{1/3}$$
$$c_n = \frac{(-1)^n}{n} \qquad d_n = \frac{n-1}{n+1}$$

★ 5.8. Let $a_n = (2n+1)/(n+3)$ ($n \in \mathbb{N}$). Show that $(a_n)_{n\in\mathbb{N}}$ is an increasing sequence and find its limit. Use the Principle of Bounded Monotone Convergence to deduce that

$$\sup\left\{\frac{2n+1}{n+3} : n \in \mathbb{N}\right\} = 2$$

★ Hand in.
⚠ Harder.

Exercises 5: Suggested Solutions

5.1. Given $\varepsilon > 0$ we find $N_\varepsilon \in \mathbb{N}$ such that if $n > N_\varepsilon$ then $0 \le a_n < \varepsilon^2$, from which it follows that $0 \le \sqrt{a_n} < \varepsilon$, so $\sqrt{a_n} \to 0$ as $n \to \infty$.

5.2. We have $x - 1/n < q_n < x + 1/n$, $x - 1/n \to x$ and $x + 1/n \to x$ as $n \to \infty$. By the Sandwich Theorem, $q_n \to x$ as $n \to \infty$.

5.3. $|\sin(n) - 2\cos(n)| \le 3$ by the triangle inequality; $C = 3/2$ will do.

5.4. Expanding out the factorials gives us

$$a_n = \frac{(n!)^2}{(2n)!} = \frac{1.2\ldots n.1.2\ldots n}{1.2\ldots n.(n+1)(n+2)\ldots(2n)}$$

The first n terms are identical on top and bottom, so they cancel giving us

$$a_n = \frac{1.2\ldots n}{(n+1)(n+2)\ldots 2n} = \frac{1}{n+1}\frac{2}{n+2}\cdots\frac{1}{2}$$

Each term is of the form $j/(n+j)$ where $1 \le j \le n$. Now, $j/(n+j) \le 1/2$, because

$$\frac{1}{2} - \frac{j}{n+j} = \frac{n-j}{2(n+j)} \ge 0$$

since $j \le n$. There are n term $\le 1/2$, so $a_n \le 1/2^n \to 0$ as $n \to \infty$. We also have $a_n > 0$, so by the Sandwich Theorem $a_n \to 0$.

It follows that $1/|a_n| \to \infty$; but $a_n > 0$, so $1/a_n \to \infty$; that is, $(2n)!/(n!)^2 \to \infty$ as $n \to \infty$.

$$\underbrace{\sum_{j=1}^{n} a_j}_{(1)} + \underbrace{\sum_{j=n+1}^{\infty} a_j}_{(2)} = \underbrace{\sum_{j=1}^{\infty} a_j}_{(3)}$$

As $n \to \infty$, term (1) converges to term (3) by definition; it follows from the algebra of limits that term (2) must tend to zero.

5.5. This can be done like the $x^n/n!$ example in the notes, but it's a bit fiddly.

A neater line is to observe that, looking only at the even terms,

$$(2n)!! = (2n)(2n-2)(4n-4)\ldots 2$$

There are n terms on the RHS, all of which have a factor of 2. Grouping these together, we have

$$(2n)! = 2^n n!$$

Now,

$$\frac{x^{2n}}{(2n)!!} = \frac{x^{2n}}{2^n n!} = \frac{(x^2/2)^n}{n!}$$

which tends to zero; it's one of the standard example in the notes.

The odd terms are slightly more awkward but can be handled with the inequality $(2n+1)!! > (2n)!!$. For $x \geq 0$ we have

$$0 \leq \frac{x^{2n+1}}{(2n+1)!!} < x\frac{x^{2n}}{(2n)!!}$$

and we have already seen that the fraction on the right tends to 0, so the odd terms tend to zero by the sandwich theorem.

Now, because the odd and even terms of the sequences both tends to zero, the sequence itself tends to zero. The case $x < 0$ is handled using the sandwich theorem exactly as in the $x^n/n!$ case in the notes.

5.6. First one, properly laid out: the dominant term (term tending most quickly to infinity, in the hierarchy of limits) is 3^n Divide top and bottom by this term to give

$$\frac{3^n + 2^n}{3^n + n^2} = \frac{1 + (2/3)^n}{1 + n^2/3^n} \to \frac{1+0}{1+0} = 1$$

as $n \to \infty$, using the hierarchy and algebra of limits.

All solutions, abbreviated:

(a) Dominant term is 3^n, limit is $(1+0)/(1+0) = 1$
(b) Dominant term is $j!$; limit is $(0+0)/(1+0+0) = 0$
(c) Dominant term is m^{m+1}; limit is $(0+0)/(1+0) = 0$
(d) Dominant term is x^k (or x^{k+1}, if you prefer); limit is $(x-0)/(1+0) = x$

5.7. (a_n) is strictly decreasing and bounded (below by 0, above by 1). It converges to 0, as we saw long ago. Alternatively, the principle of bounded monotone convergence tells us that it converges to the infimum of its terms, which we could verify is 0 using Archimedes.

(b_n) is strictly increasing but not bounded above (Archimedes). It does not converge, because all convergent sequences are bounded.

(c_n) is neither increasing nor decreasing, because the terms alternate in sign. It converges to 0.

It might not be immediately obvious whether (d_n) is increasing or decreasing. We test it with the definition by looking at $d_{n+1} - d_n$ and seeing what sign it has:

$$d_{n+1} - d_n = \frac{n}{n+2} - \frac{n-1}{n+1} = \frac{n^2 + n - (n^2 + n - 2)}{(n+1)(n+2)} = \frac{2}{(n+1)(n+2)} > 0$$

so (d_n) is strictly increasing. It also follows from the standard use of combination rules that $d_n \to 1$ as $n \to \infty$, and hence that it is bounded.

Another manoeuvre, which is sometimes easier, is to start with the limit and subtract it:

$$d_n - 1 = -\frac{2}{n+1} \implies d_n = 1 - \frac{2}{n+1}$$

In this form, it is clear that the sequence is strictly increasing.

5.8. Consider the difference between two consecutive terms:

$$\frac{2n+3}{n+4} - \frac{2n+1}{n+3} = \cdots = \frac{5}{(n+3)(n+4)} > 0$$

Since this is positive, the sequence is increasing. We also have

$$\frac{2n+1}{n+3} = \frac{2+1/n}{1+3/n} \to 2$$

as $n \to \infty$ (showing that the sequence is bounded, in passing). The principle of bounded monotone convergence now shows that the limit and the supremum of the terms are equal; we know that the limit is 2, so the supremum is also 2.

Exercises 6

Questions (or question parts) marked with a star ★ constitute Assignment 6; please submit on Moodle by the due date. Assignments make up 5% of the module mark.

6.1. For $x \in \mathbb{R}$, let $f(x) = (x^2 - x + 2)/2$.

(a) By factorising $f(x) - f(y)$, show that f is strictly increasing on $(1/2, \infty)$, i.e. that if $1/2 < x < y$ then $f(x) < f(y)$. Deduce that if $1 < x < 2$ then $1 < f(x) < 2$.

(b) Let $x_1 = 3/2$ and (for $n \in \mathbb{N}$) $x_{n+1} = f(x_n)$. Show that $(x_n)_{n\in\mathbb{N}}$ is a bounded monotonic sequence, and find its limit.

★ 6.2. Let $a_1 = 1$ and $a_{n+1} = \sqrt{2a_n}$ ($n \in \mathbb{N}$).

(a) Show that if $x \in \mathbb{R}$ and $0 < x < 2$ then $0 < \sqrt{2x} < 2$.

(b) Show that if $x \in \mathbb{R}$ and $0 < x < 2$ then $x < \sqrt{2x}$.

(c) Show that $0 < a_n < 2$ for all n (induction).

(d) Show that $(a_n)_{n\in\mathbb{N}}$ is monotonic.

(e) Deduce that $(a_n)_{n\in\mathbb{N}}$ converges to a limit a; find and solve an equation to determine the value of a.

Later parts depend on earlier parts; but if you can't solve one part, all the information needed for subsequent parts can be found in the question.

6.3. Use the Limit Comparison Test to determine whether or not the following series converge or diverge:

$$\sum_{j=1}^{\infty} \frac{j^2 - 2}{j^3 + j^2}; \qquad \sum_{j=1}^{\infty} \frac{j - 1}{j^3 - 3}$$

Remember to check for positive terms!

6.4. Which of the following series satisfy the hypotheses of the Leibniz test?

(a) $\displaystyle\sum_{j=1}^{\infty} \frac{(-1)^{j+1}}{1 + 2^j}$

(b) $\displaystyle\sum_{j=1}^{\infty} (-1)^{j+1} \frac{j}{1 + j}$

(c) $\displaystyle\sum_{j=1}^{\infty} (-1)^{j+1} \frac{j}{1 + j^2}$

(d) $\displaystyle\sum_{j=1}^{\infty} (-1)^{j+1} \frac{1 + (-1)^j}{j^2}$

★ 6.5. Any Cauchy sequence $(a_n)_{n\in\mathbb{N}}$ has the property that $|a_{n+1} - a_n| \to 0$ as $n \to \infty$ (this can be seen directly from the definition, or by using the fact that a_n and a_{n+1} converge to the same limit).

Show that if $b_n = n^{1/3}$ ($n \in \mathbb{N}$) then $b_{n+1} - b_n \to 0$ as $n \to \infty$ but that $(b_n)_{n\in\mathbb{N}}$ is not a Cauchy sequence. *The identity $x^3 - y^3 = (x - y)(x^2 + xy + y^2)$, applied to b_{n+1} and b_n, is useful here.*

6.6. Let

$$a_j = \begin{cases} 1/j & \text{if } j \text{ is odd} \\ 1/j^2 & \text{if } j \text{ is even} \end{cases}$$

Show that the series

$$\sum_{j=1}^{\infty} (-1)^{j+1} a_j$$

diverges. Why does this not contradict the Leibniz test?

⚠ 6.7. Consider the series

$$\sum_{j=1}^{\infty} \frac{(-1)^{j+1}}{j} \tag{†}$$

which converges by the Leibniz test. Another way to show that this converges is to put the terms together in pairs, i.e. to consider

$$\sum_{k=1}^{\infty} \underbrace{\frac{1}{2k-1} - \frac{1}{2k}}_{=b_k} \tag{‡}$$

This is a sum of positive terms; use the (limit) comparison test to show directly that it converges, and explain why (†) and (‡) sum to the same value. Now try adding the same terms together in a different order: from (*), add in two positive (odd-numbered) terms then one negative (even-numbered) term and repeat: the partial sums begin as

$$(1 + 1/3 - 1/2) + (1/5 + 1/7 - 1/4) + (1/9 + 1/11 - 1/6) + \cdots$$

Let c_k be the kth bracketed term in this sum. Write down an explicit formula for c_k. Show that $c_k > b_k$ for all k and that $\sum_{k=1}^{\infty} c_k$ converges to a value strictly greater than that of (†) and (‡). Look up the *Riemann rearrangement theorem* for the whole story!

★ Hand in.
⚠ Harder.

Exercises 6: Suggested Solutions

6.1.

(a) Easy to check that

$$f(x) - f(y) = \frac{1}{2}(x + y - 1)(x - y)$$

If $1/2 < x < y$ then $x+y-1 > 0$ and $x-y < 0$, so $f(x)-f(y) < 0$, i.e. $f(x) < f(y)$. It follows that if $1 < x < 2$ then $f(1) < f(x) < f(2)$. But $f(1) = 1$ and $f(2) = 2$ so $1 < f(x) < 2$.

(b) We know that $x_1 \in (1, 2)$ and it follows from the above that if $x_n \in (1, 2)$ then $x_{n+1} \in (1, 2)$; by induction, $x_n \in (1, 2)$ for all n. This establishes boundedness. To establish monotonicity, consider

$$x_{n+1} - x_n = \frac{x_n^2 - x_n + 2}{2} - x_n = \frac{1}{2}(x_n^2 - 3x_n + 2) = \frac{1}{2}(x_n - 1)(x_n - 2)$$

Since $1 < x_n < 2$, this is always negative, so $(x_n)_{n\in\mathbb{N}}$ is decreasing. Since it is bounded, it must converge, say $x_n \to x$ as $n \to \infty$. We also have $x_{n+1} \to x$ as $n \to \infty$; but

$$x_{n+1} = \frac{1}{2}(x_n^2 - x_n + 2) \to \frac{1}{2}(x^2 - x + 2)$$

as $n \to \infty$. By uniqueness of limits, $x = (x^2 - x + 2)/2$. Solving this quadratic gives $x = 1$ or $x = 2$. Since the sequence is decreasing, we must have $x \le x_1 = 3/2$, so $x = 1$.

6.2.

(a) Multiply $0 < x < 2$ by 2 to give $0 < 2x < 4$; take square roots to give $0 < \sqrt{2x} < 2$.

(b) Multiply $0 < x < 2$ by x to give $0 < x^2 < 2x$; take square roots to give $0 < x < \sqrt{2x}$.

(c) Note that a_1 satisfies $0 < a_1 < 2$. Suppose $0 < a_n < 2$. Then (a) gives $0 < a_{n+1} < 2$. By induction, $0 < a_n < 2$ for all $n \in \mathbb{N}$.

(d) By (c), each a_n lies between 0 and 2. If follows from (b) that $a_{n+1} = \sqrt{2a_n} > a_n$, so $(a_n)_{n\in\mathbb{N}}$ is an increasing sequence.

(e) We know from (c) that $(a_n)_{n\in\mathbb{N}}$ is bounded above by 2 and from (d) that $(a_n)_{n\in\mathbb{N}}$ is increasing, so by bounded monotone convergence we know that $a_n \to a$ as $n \to \infty$, for some $a \le 2$. Now, if $a_n \to a$ as $n \to \infty$ then also $a_{n+1} \to a$; also, $a_{n+1} = \sqrt{2a_n} \to \sqrt{2a}$ as $n \to \infty$. By uniqueness of limits, we must have $a = \sqrt{2a}$, so $a^2 = 2a$ and hence either $a = 0$ or $a = 2$. But $a_1 = 1$ and $(a_n)_{n\in\mathbb{N}}$ is increasing, so the limit cannot be 0; we conclude that $a = 2$.

6.3. First, we check the signs of the terms. In the first series, the $j = 1$ term is negative, but all others are strictly positive. One term does not affect convergence or divergence, so we can proceed with the limit comparison test. We have degrees 2 and 3 on the numerator and denominator, so compare with $\sum j^{2-3} = \sum 1/j$, which diverges (harmonic series). We consider

$$\frac{j^2-2}{j^3+j^2}\Big/\frac{1}{j} = \frac{j^3-2j}{j^3+j^2} = \frac{1-2/j^2}{1+1/j} \to 1$$

as $j \to \infty$; the limit is non-zero so by, the Limit Comparison Test, convergence of the series in question is equivalent to convergence of the comparison series. Since the harmonic series diverges, the first series in the question diverges.

In the second series, all terms are strictly positive except the first, which is zero and has no effect on anything. We have degrees 1 and 3 on the numerator and denominator, so compare with $\sum j^{1-3} = \sum 1/j^2$, which converges (standard example, can be established using e.g. the condensation test). We consider

$$\frac{j-1}{j^3-3}\Big/\frac{1}{j^2} = \frac{j^3-j^2}{j^3-3} = \frac{1-1/j}{1-3/j^2} \to 1$$

as $j \to \infty$. Since the limit exists and the comparison series converges, the Limit Comparison Test shows that the second series in the question converges.

6.4.

(a) Satisfies hypotheses.

(b) Does not satisfy hypotheses: $j/(j+1) \not\to 0$ as $j \to \infty$ (hence divergent).

(c) Satisfies hypotheses (though it takes a little algebra to verify that $j/(1+j^2)$ is decreasing).

(d) Does not satisfy hypotheses: $(1+(-1)^j)/j^2$ not decreasing.The series is, however, convergent (by comparison with $2/j^2$).

6.5. Suppose (a_n) is a Cauchy sequence. Here are two ways to show $a_{n+1} - a_n \to 0$ as $n \to \infty$.

(a) Directly from the Cauchy property: for any $\varepsilon > 0$ there exists $N\varepsilon \in \mathbb{N}$ such that if $m, n > N_\varepsilon$ then $|a_m - a_n| < \varepsilon$. In particular, if $n > N_\varepsilon$ then also $n+1 > N_\varepsilon$, so $|a_{n+1} - a_n| < \varepsilon$. Since this is true for all $n > N_\varepsilon$, $a_{n+1} - a_n \to 0$.

(b) Via convergence: since (a_n) is Cauchy, it converges, say $a_n \to a$ as $n \to \infty$. Then we also have $a_{n+1} \to a$ as $n \to \infty$, so (combination rules) $a_{n+1} - a_n \to a - a = 0$.

Substituting $x = b_{n+1} = (n+1)^{1/3}$ and $y = b_n = n^{1/3}$ into the stated identity $x^3 - y^3 = (x-y)(x^2+xy+y^2$ gives

$$1 = (b_{n+1} - b_n)(n^{2/3} + n^{1/3}(n+1)^{1/3} + (n+1)^{2/3})$$

and hence

$$b_{n+1} - b_n = \frac{1}{n^{2/3} + n^{1/3}(n+1)^{1/3} + (n+1)^{2/3}}$$

By standard hierarchy of limits / algebra of limits arguments, the RHS tends to zero as $n \to \infty$ but $b_n \to \infty$ as $n \to \infty$. Since $(b_n)_{n\in\mathbb{N}}$ does not converge to a finite limit, it is not a Cauchy sequence.

6.6. Consider the sum of the first $2n$ terms. Writing the odd and even terms separately, we have

$$\left(1+\frac{1}{3}+\frac{1}{5}+\cdots+\frac{1}{2n-1}\right)-\left(\frac{1}{4}+\frac{1}{16}+\cdots+\frac{1}{4n^2}\right)$$

The first bracket tends to ∞ as $n \to \infty$, for the same reasons as the harmonic series. The second bracket remains bounded as $n \to \infty$, since $\sum_{j=1}^{\infty} 1/j^2$ converges. It follows that the $(2n)$th partial sum tends to ∞ as $n \to \infty$, and hence that the series diverges. This does not contradict the Leibniz test, because (a_j) is not decreasing.

6.7. The series (‡) converges by limit comparison with the standard example $\sum 1/k^2$:

$$\frac{1}{2k-1}-\frac{1}{2k}=\frac{1}{2k(k-1)}; \qquad \frac{1}{2k(2k-1)}\Bigg/\frac{1}{k^2}=\frac{k^2}{2k(2k-1)}\to\frac{1}{4}$$

as $k \to \infty$. The kth partial sum of (‡) is the $2k$th partial sum of (†), so the partial sums of (‡) form a subsequence of the partial sums of (†), and hence converge to the same limit.

We can write down c_k explicitly:

$$c_k=\frac{1}{4k-3}+\frac{1}{4k-1}-\frac{1}{2k}$$

To compare with b_k, subtract; the $1/(2k)$ terms cancel, leaving

$$c_k-b_k=\frac{1}{4k-3}+\frac{1}{4k-1}-\frac{1}{2k-1}$$

which simplifies to

$$c_k-b_k=\frac{1}{(4k-3)(4k-1)(2k-1)}>0$$

To show that the $\sum c_k$ converge, we can directly apply the limit comparison test to b_k, with comparison series $\sum 1/k^2$, or save a little algebra by noticing from the above formula for c_k-b_k that $\sum(c_k-b_k)$ converges (limit comparison with $\sum 1/k^3$); combined with the fact that $\sum b_k$ converges, this tells us that $\sum c_k$ converges. Now, to conclude that $\sum_{k=1}^{\infty} b_k > \sum_{k=1}^{\infty} c_k$, we note that the partial sums of $\sum_{k=1}^{\infty}(c_k-b_k)$ are an increasing sequence whose first term is $c_1-b_1>0$; it follows that

$$\sum_{k=1}^{\infty}(c_k-b_k)\geq c_1-b_1>0$$

so, breaking the sum on the LHS into two pieces using linearity

$$\sum_{k=1}^{\infty} b_k-\sum_{k=1}^{\infty} c_k>0$$

Exercises 7

Questions (or question parts) marked with a star ★ constitute Assignment 7; please submit on Moodle by the due date. Assignments make up 5% of the module mark.

7.1. We will define logarithms later in the course; for now, using any of their standard properties that you might need, apply the condensation test to the series

$$\sum_{j=2}^{\infty} \frac{1}{j(\log(j))^p}; \qquad \sum_{j=3}^{\infty} \frac{1}{j\log(j)(\log(\log(j)))^p}$$

to find for what values of p they converge. Why are the sums from $j = 2$ and $j = 3$, not $j = 1$?

7.2. For which $x \in \mathbb{R}$ is the geometric series $\sum_{j=0}^{\infty} x^j$ absolutely convergent?

★ 7.3. Use the convergence tests and examples in the notes to determine for which values of $p \in \mathbb{Q}$ the series

$$\sum_{j=1}^{\infty} \frac{(-1)^j}{j^p}$$

is

(a) convergent

(b) absolutely convergent

(c) divergent

7.4. Show that

$$\sum_{j=0}^{\infty} \frac{(j!)^2}{(2j)!}$$

converges. *Remarkably, there is an exact expression for this sum:* $2\sqrt{3}\pi/27 + 4/3$.

7.5. The aim of this question is to show that

$$\sum_{n=1}^{\infty} \frac{n!}{n^n} \tag{†}$$

converges, using techniques similiar to those used in the proof of the ratio test.

(a) For $n \in \mathbb{N}$, we have by the Bionomial Theorem

$$\left(1 + \frac{1}{n}\right)^n = \sum_{k=0}^{n} \binom{n}{k} \frac{1}{n^k} \tag{‡}$$

Use the first two terms of (‡) to find a simple lower bound for $(1 + 1/n)^n$.

(b) Letting $a_n = n!/n^n$ ($n \in \mathbb{N}$), use the lower bound from the previous part to show that for any $n \in \mathbb{N}$, $a_{n+1}/a_n \le 1/2$.

(c) Write down an induction argument showing that, for all $n \in \mathbb{N}$, $a_n \le 2^{2-n}$.

(d) Use an appropriate convergence test (but not the Ratio Test!) to deduce that the sum in (†) converges. Find an upper bound for the sum.

⚠ 7.6. The Root Test: suppose $(a_j)_{j\in\mathbb{N}}$ is a sequence with the property that $|a_j|^{1/j} \to r$ as $j \to \infty$. Show that $\sum_{j=1}^{\infty} a_j$ converges absolutely if $r < 1$ and diverges if $r > 1$. *The proof is similar to that of the ratio test, but a bit simpler!*

Use this test to show that the power series

$$\sum_{n=1}^{\infty} \frac{x^n}{n^n}$$

has radius of convergence ∞. Use the ratio test to derive the same conclusion, and compare the amount of work needed to apply the two different techniques.

7.7. Use the Ratio test to find the radius of convergence of the following power series:

★ (a) $\displaystyle\sum_{n=0}^{\infty} \frac{(x+3)^n}{(n!)^2}$

(b) $\displaystyle\sum_{n=0}^{\infty} (2^n + 1)x^n$

★ (c) $\displaystyle\sum_{n=0}^{\infty} \frac{2^n - 1}{3^n - 2} x^n$

(d) $\displaystyle\sum_{n=1}^{\infty} \frac{3^n - 1}{n} x^{2n}$

7.8. Find the radius of convergence of the following power series:

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

(b) $\displaystyle\sum_{n=0}^{\infty} \frac{(3n)!}{(n!)^3} x^n$

7.9. Show that if P is any polynomial then $P(n+1)/P(n) \to 1$ as $n \to \infty$. Deduce that if P and Q are polynomials (such that $Q(n) \ne 0$ if $n \in \mathbb{N}_0$) then the power series

$$\sum_{n=0}^{\infty} \frac{P(n)}{Q(n)} (x - x_0)^n$$

has radius of convergence 1.

⚠ 7.10. Suppose $(a_n)_{n\in\mathbb{N}_0}$ is a bounded sequence and $a_n \not\to 0$ as $n \to \infty$. Show that the power series

$$\sum_{n=0}^{\infty} a_n x^n$$

has radius of convergence 1.

★ Hand in.
⚠ Harder.

Exercises 7: Suggested Solutions

7.1. For the first example, the condensed series is

$$\sum_{k=1}^{\infty} 2^k \frac{1}{2^k(\log(2^k))^p} = (\log 2)^p \sum_{k=2}^{\infty} \frac{1}{k^p}$$

which we know converges if $p > 1$ and diverges if $p \leq 1$. The series starts at 2, since $\log(1) = 0$, so the formula for the terms makes no sense when $k = 0$.

For the second example, the condensed series is

$$\begin{aligned}\sum_{k=2}^{\infty} 2^k \frac{1}{2^k \log(2^k)(\log(\log(2^k)))^p} &= \frac{1}{\log(2)} \sum_{k=2}^{\infty} \frac{1}{k(\log(k\log(2)))^p} \\ &= \frac{1}{\log(2)} \sum_{k=2}^{\infty} \frac{1}{k(\log(k)+\log(2))^p}\end{aligned}$$

This is almost, but not quite, the same as the first example. We can show that they have the same convergence properties by using the limit comparison test:

$$\frac{1}{j(\log(j))^p} \Bigg/ \frac{1}{j(\log(j)+\log(2))^p} = \left(1 + \frac{\log(2)}{\log(j)}\right)^p \to 1$$

as $j \to \infty$ (because $\log(j) \to \infty$ as $j \to \infty$). The two series therefore converge or diverge together so, applying the result of the previous question, we have convergence if $p > 1$ and divergence if $p \leq 1$.

The series starts at $j = 3$ because this is the first j for which $\log(\log(j)) > 0$, and the condensation test requires non-negative terms.

7.2. By definition, the geometric sum is absolutely convergent if and only if $\sum_{j=0}^{\infty} |x^j|$ converges. But $|x^j| = |x|^j$, so this is equivalent to $\sum_{j=1}^{\infty} |x|^j$ converging. This is also a geometric series, with common ratio $|x|$; it converges (Example 8.7) if and only if $|x| < 1$.

Or, we could apply the general fact that a power series converges absolutely within its radius of convergence; but since the proof of that fact used (implicitly) the result stated in this question, such a solution could be aaccused of circular reasoning.

7.3.

(a) If $p > 0$ then the series is convergent by the Leibniz test.

(b) If $p > 1$ then the series is absolutely convergent (Example 12.4)

(c) If $p \leq 0$ then the series is divergent because then $(-1)^j/n^j$ does not tend to zero as $j \to \infty$.

7.4. This is almost the same as an example in the notes, but with the numerator and denominator swapped. Ratio test:

$$\frac{[(j+1)!]^2}{(2j+2)!}\Bigg/\frac{(j!)^2}{(2j)!} = \frac{(j+1)^2}{(2j+1)(2j+2)} \to \frac{1}{4} < 1.$$

We can also make a more bare-handed attack, using the inequality from Exercise 5.4, where we saw that

$$\frac{(j!)^2}{(2j)!} \le 2^{-j}$$

for $j \in \mathbb{N}$, and we can note that this is also trivially true if $j = 0$. Here are two slightly different ways of concluding the problem:

(a) Using Bounded Monotone Convergence: let

$$s_n = \sum_{j=0}^{n} \frac{(j!)^2}{(2j)!} \le \sum_{j=0}^{n} 2^{-j} = \frac{1-(1/2)^{n+1}}{1-1/2} < 2$$

Since (s_n) is an increasing sequence which is bounded above by 2, it converges.

(b) Using the comparison test (which is just a pre-packaged version of the above argument: we know that $\sum_{j=0}^{\infty} 2^{-j}$ converges, so the series

$$\sum_{j=0}^{\infty} \frac{(j!)^2}{(2j)!}$$

converges by the comparison test.

7.5. (a) The first two terms are

$$\binom{n}{0}\frac{1}{n^0} + \binom{n}{1}\frac{1}{n} = 1\frac{1}{1} + n\frac{1}{n} = 2$$

Since all the terms in the binomial expansion are non-negative, we have

$$\left(1+\frac{1}{n}\right)^n \ge 2$$

(b) We have

$$\frac{a_{n+1}}{a_n} = \frac{(n+1)!}{(n+1)^{n+1}}\frac{n^n}{n!} = \frac{(n+1)^n}{n^n} = \left(\frac{n}{n+1}\right)^n = \left(1+\frac{1}{n}\right)^{-n} \le \frac{1}{2}$$

using the estimate from the previous part (which turns into a lower bound when we take reciprocals)

(c) We have $a_1 = (1+1)^1 = 2 = 2^{2-1}$ so the inequality $a_n \le 2^{2-n}$ is true for $n = 1$. Assume that for some $n \in \mathbb{N}$ we have $a_n \le 2^{2-n}$; then since $a_{n+1} \le a_n/2$, we have $a_{n+1} \le 2^{2-n}/2 = 2^{2-(n+1)}$,so the required inequality is true for $n+1$. Now, by induction, we have $a_n \le 2^{2-n}$ for all $n \in \mathbb{N}$.

(d) Now, we use the Comparison Test. Since $\sum_{n=1}^{\infty} 2^{2-n}$ converges (geometric sum, common ratio 1/2), we see that the series in (*) converges and that

$$\sum_{n=1}^{\infty} \frac{n!}{n^n} \leq \sum_{n=1}^{\infty} 2^{2-n} = \frac{2}{1-1/2} = 4$$

7.6. Assume first that $r < 1$ and, as in the proof of the ratio test, choose s with $1 < s < r$. Then there exists $N \in /N$ such that if $j > N$ then $|a_j|^{1/j} < s$ and hence $|a_j| < s^j$. Since $0 < s < 1$, $\sum_{j=N+1}^{\infty} s^j$ converges. By the comparison test, $\sum_{j=N+1}^{\infty} a_j$ is absolutely convergent and hence $\sum_{j=1}^{\infty} a_j$ is absolutely convergent.

Now assume that $r > 1$ and choose s with $1 < r < s$, so there exists $N \in \mathbb{N}$ such that if $j > N$ then $|a_j|^{1/j} > s$ and hence $|a_j| > s^j$. Since $s > 1$, $s^j \to \infty$, so we also have $|a_j| \to \infty$, and hence $\sum_{j=1}^{\infty} a_j$ diverges.

We need to look at

$$\left|\frac{x^n}{n^n}\right|^{1/n} = \frac{|x|}{n} \to 0$$

as $n \to \infty$, regardless of the value of x. Since the limit is less than 1, the series converges for all x; that is, the radius of convergence is ∞.

The ratio test is a little more complicated to use:

$$\frac{n^n}{(n+1)^{n+1}}|x| = \frac{1}{n+1}\left(\frac{n}{n+1}\right)^n |x|$$

Now, $1/(n+1) \to 0$, $(n/(n+1))^n < 1$ and $|x|$ is constant, so this tends to 0 as $n \to \infty$ and again we can conclude that the radius of convergence is ∞.

The root test is, for this example, a little simpler to use: in particular, it does not rely on considering $(n/(n+1))^n$ and involves a bit less algebraic manipulation.

7.7. Use the ratio test in all cases. Note that this test cannot be applied at the centre point; at that point, we always have convergence.

(a)
$$\left|\frac{(x+3)^{n+1}}{((n+1)!)^2}\frac{(n!)^2}{(x+3)^n}\right| = \frac{|x+3|}{n^2} \to 0$$

as $n \to \infty$. Since the limit is 0, regardless of x, we have convergence for all $x \in \mathbb{R}$; that is, radius of convergence $R = \infty$.

(b)
$$\left|\frac{(2^{n+1}+1)x^{n+1}}{(2^n+1)x^n}\right| = \frac{2^{n+1}+1}{2^n+1}|x| \to 2|x|$$

We have convergence if $2|x| < 1$ and divergence if $2|x| > 1$, so we have radius of convergence $R = 1/2$.

(c)

$$\left|\frac{(2^{n+1}-1)x^{n+1}}{3^{n+1}-2}\frac{3^n-2}{(2^n-1)x^n}\right| = \frac{(2^{n+1}-1)(3^n-2)}{(3^{n+1}-2)(2^n-1)}|x|$$

It is possible, of course, to expand out the brackets in this, but slightly easier to divide top and bottom by both 2^n and 3^n, giving

$$\frac{(2-2^{-n})(1-2\cdot 3^{-n})}{(3-2\cdot 3^{-n})(1-2^{-n})}|x|$$

which converges to $2|x|/3$ as $n \to \infty$. We have convergence if $2|x|/3 < 1$, i.e $|x| < 3/2$, and divergence if $2|x|/3 > 1$, i.e $|x| > 3/2$. The radius of convergence is thus $3/2$.

(d)

$$\left|\frac{3^{n+1}-1}{n+1}\frac{n}{3^n-1}\right||x^2| \to 3|x|^2$$

so we have convergence if $3|x|^2 < 1$ and divergence if $3|x|^2 > 1$; that is, radius of convergence $R = \sqrt{3}/3$.

7.8. These are both calculated using the ratio test. All the coefficients are positive, except the first two in the first example, so absolute values can be ignored except on the x terms. In the first example, which has only odd terms present, we have $|x^{2n+3}/x^{2n+1}| = x^2$; in the second example, we have $|x^{n+1}/x^n| = |x|$. The rest of the calculation is, for the first example:

$$\frac{(n+1)^2+1}{2^{n+1}-3}\frac{2^n-3}{n^2+1}x^2 \to x^2/2$$

as $n \to \infty$ (dominant term in the limit calculation is 2^n or 2^{n+1}). We therefore have convergence if $x^2/2 < 1$, i.e. $|x| < \sqrt{2}$ and divergence if $x^2/2 > 1$, i.e. $|x| > \sqrt{2}$. The radius of convergence is therefore $\sqrt{2}$. And for the second example:

$$\frac{[3(n+1)]!}{[(n+1)!]^3}\frac{(n!)^3}{(3n)!}|x| = \frac{(3n+3)(3n+2)(3n+1)}{(n+1)^3}x^2 \to 27|x|$$

as $n \to \infty$. We therefore have convergence if $27|x| < 1$, i.e. $|x| < 1/27$ and divergence if $27|x| > 1$, i.e. $|x| > 1/27$. The radius of convergence is therefore $1/27$.

7.9. Suppose $P(n) = \sum_{j=0}^N \alpha_j n^j$ where $\alpha_N \neq 0$, so

$$\frac{P(n+1)}{P(n)} = \frac{\sum_{j=0}^N \alpha_j (n+1)^j}{\sum_{j=0}^N \alpha_j n^j}$$

The numerator and denominator both have leading term $\alpha_N n^N$ so the ratio tends to 1 as $n \to \infty$. Now, if we apply the ratio test to the series

$$\sum_{n=0}^{\infty} \frac{P(n)}{Q(n)}(x-x_0)^n$$

then we have to consider

$$\left|\frac{P(n+1)Q(n)}{Q(n+1)P(n)}\right| |x - x_0|$$

as $n \to \infty$. Applying the first part of the question to P and Q shows that the limit is $|x - x_0|$, so we have convergence if $|x - x_0| < 1$ and divergence if $|x - x_0| > 1$; that is, radius of convergence $R = 1$.

7.10. Firstly, we are given that the coefficient sequence is bounded, say $|a_n| \leq K$ for all n. This gives us $|a_n x^n| \leq K|x|^n$ for all n. Since the series $\sum_{n=0}^{\infty} K|x|^n$ converges for any x with $|x| < 1$, then comparison text shows that the power series in question converges if $|x| < 1$; that is, its radius of convergence R satisfies $R \geq 1$.

Secondly, we are given that $a_n \nrightarrow 0$ as $n \to \infty$. We can think of this as saying that $a_n 1^n \nrightarrow 0$ as $n \to \infty$, which implies that the series $\sum_{n=0}^{\infty} a_n 1^n$ does not converge. Since the power series diverges when $x = 1$, its radius of convergence cannot be larger than 1, i.e. $R \leq 1$.

We now have $R \geq 1$ and $R \leq 1$, so $R = 1$.

Exercises 8

Questions (or question parts) marked with a star ★ constitute Assignment 8; please submit on Moodle by the due date. Assignments make up 5% of the module mark.

8.1. Show that if $n \in \mathbb{N}$ then

$$\lim_{x \to 1} \frac{x^{n+1} - 1}{x - 1} = n + 1$$

This can be done using only the results of Section 18; no differentiation is needed! Think geometric series.

★ 8.2. Use the ε-δ definition, Definition 18.1 (and no theorems about limits!) to show that if $x_0 \in \mathbb{R}$ then $|x| \to |x_0|$ as $x \to x_0$.

8.3. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^2$. Use the ε-δ definition, Definition 18.1 (and no theorems about limits!) to show that $f(x) \to 4$ as $x \to 2$.

8.4. The analogues of Theorem 5.12 for limits at a point: suppose $S \subseteq \mathbb{R}$ and that $f(x) \to L$ as $x \to x_0$ through S. Show that:

(a) Provided (*) below holds, if $f(x) \geq M$ for all $x \in S \setminus \{x_0\}$ then $L \geq M$; if $f(x) \leq M$ for all $x \in S$ then $L \leq M$.

(b) If $L < M$ then there exists $\delta > 0$ such that if $x \in S$ with $0 < |x - x_0| < \delta$ then $f(x) < M$; if $L > M$ then there exists $\delta > 0$ such that if $x \in S$ with $0 < |x - x_0| < \delta$ then $f(x) > M$.

The first one requires the technical assumption that x_0 is a limit point of S (see Remark 18.8); precisely, that

$$\text{if } \delta > 0 \text{ then there exists } x \in S \text{ with } 0 < |x - x_0| < \delta \qquad (*)$$

These are most easily done using ε and δ. The first can also be done straightforwardly using the sequential characterisation; the second part can be done using sequences, but it's tricky.

8.5. The Sandwich Theorem for limits at a point: suppose $S \subseteq \mathbb{R}$, that $f, g, h : S \to \mathbb{R}$ are such that $f(x) \leq g(x) \leq h(x)$ for all $x \in S$, and that $L \in \mathbb{R}$. Show that if $f(x) \to L$ as $x \to x_0$ through S and $h(x) \to L$ as $x \to x_0$ through S then $g(x) \to L$ as $x \to x_0$ through S. *The quickest way to do this is to use the sequential characterisation of limits and the Sandwich Theorem for sequences, but it can also be done using ε and δ.*

★ 8.6. Suppose $f : \mathbb{R} \to \mathbb{R}$ has the property that $f(x) \to a$ as $x \to \infty$ and $f(x) \to b$ as $x \to -\infty$, where $a, b \in \mathbb{R}$. Let $g(x) = f(1/x)$ for $x \neq 0$. Find the left and right limits of g at 0. Under what circumstances does g have a limit at 0? *You can do this with ε and δ, but it's easier to use the results in the notes, particularly about compositions of functions, in this case f with $x \mapsto 1/x$.*

⚠ 8.7. The sine function was defined by means of a power series in Definition 15.1. Use this definition to show that, as $x \to 0$,

$$\frac{\sin(x)}{x} \to 1$$

Adapt your argument to find the limits as $x \to 0$ of

$$\frac{\cos(x)-1}{x^2}; \qquad \frac{\exp(x)-1}{x}$$

★ Hand in.
⚠ Harder.

Exercises 8: Suggested Solutions

8.1. We know that if $x \neq 1$ then

$$\frac{x^{n+1}-1}{x-1} = \sum_{j=0}^{n} x^j$$

As $x \to 1$, each term in the (finite!) sum tends to 1; there are $n+1$ terms, so the sum tends to $n+1$.

8.2. Given $\varepsilon > 0$, from the inequality $0 < |x - x_0| < \delta_\varepsilon$, we need to reach the inequality $||x_0| - |x|| < \varepsilon$. Recall the reverse triangle inequality: $||x| - |x_0|| \le |x - x_0|$. Comparing these, we see that $\delta_\varepsilon = \varepsilon$ will work. Formally, given $\varepsilon > 0$, let $\delta_\varepsilon = \varepsilon$. If $0 < |x - x_0| < \delta_\varepsilon$ then $||x| - |x_0|| \le |x - x_0| < \delta_\varepsilon = \varepsilon$. This shows that $|x| \to |x_0|$ as $x \to x_0$. *In language introduced a little later in the course, this shows that the absolute value function is continuous on $\mathbb{R}$; in fact, since δ_ε does not depend on x_0, it shows that the absolute value function is uniformly continuous on $\mathbb{R}$.*

8.3. Idea: from the inequality $|x - 2| < \delta$, we need to reach the inequality $|x^2 - 4| < \varepsilon$, or equivalently $|x + 2||x - 2| < \varepsilon$. If we restrict δ to being no larger than 1, then we will have $x \in (1, 3)$ so $|x + 2| < 5$. This suggests that $\delta_\varepsilon = \min\{1, \varepsilon/5\}$ should work.

Formal proof: given $\varepsilon > 0$, let $\delta_\varepsilon = \min\{1, \varepsilon/5\}$. If $|x - 2| < \delta_\varepsilon$ then $|x - 2| < 1$, so $x \in (1, 3)$ and $|x + 2| < 5$. We also have $|x - 2| < \varepsilon/5$ so

$$|f(x) - 4| = |x^2 - 4| = |x + 2||x - 2| < 5\frac{\varepsilon}{5} = \varepsilon$$

This shows that $f(x) \to 4$ as $x \to 2$.

8.4. As in the proof of Theorem 5.12, which the following argument follows very closely, we only give a proof for the first statement; the second one is then a minor modification.

(a) For any $\varepsilon > 0$ there exists $\delta > 0$ such that if $x \in S$ with $0 < |x - x_0| < \delta$ then $|f(x) - L| < \varepsilon$, or equivalently $L - \varepsilon < f(x) < L + \varepsilon$.
<u>Choose and fix some x with $0 < |x - x_0| < \delta$.</u>
Since $f(x) \ge M$, we have $M < L + \varepsilon$. Because this is true for every $\varepsilon > 0$, we have (Lemma 3.5) $M \le L$.

(b) Put $\varepsilon = M - L > 0$ in the definition of limit to find $\delta > 0$ such that if $x \in S$ with $0 < |x - x_0| < \delta$ then $|f(x) - L| < M - L$. In particular, $f(x) - L < M - L$, so $f(x) < M$.

The technical point about x_0 being a limit point of S is required to make the underlined "Choose and fix ..." step possible. Without this step, the two inequalities could not usefully be combined.

As remarked, the sequential characterisation can also be used.

(a) Apply (*) with $\delta = 1/n$ to find a sequence $(x_n)_{n\in\mathbb{N}}$ in $S \setminus \{x_0\}$ with $|x - x_0| < 1/n$, so $x_n \to x_0$ as $n \to \infty$. Then, by the sequential characterisation of limits, $f(x_n) \to L$. But $f(x_n) \ge M$ so, by Theorem 5.12, $L \ge M$.

(b) The contrapositive of the required statement is:
"If for all $\delta > 0$ there exists $x \in S$ with $0 < |x - x_0| < \delta$ such that $f(x) \geq M$ then $L \geq M$."
Apply this with $\delta = 1/n$ ($n \in \mathbb{N}$) to find $x_n \in S \setminus \{x_0\}$ with $|x_0 - x_n| < 1/n$ and $f(x_n) \geq M$. Then $x_n \to x_0$ as $n \to \infty$, so (sequential characterisation of limits) $f(x_n) \to L$ as $n \to \infty$. Now apply Theorem 5.12, to give $L \geq M$.

8.5. As suggested, we use the sequential characterisation of limits. Suppose $x_n \to x_0$ as $n \to \infty$ in $S \setminus \{x_0\}$. Then, by the sequential characterisation, $f(x_n) \to L$ and $h(x_n) \to L$ as $n \to \infty$. Since $f(x_n) \leq g(x_n) \leq h(x_n)$, the Sandwich Theorem for sequences shows that $g(x_n) \to L$ as $n \to \infty$. Since this is true for every sequence in $S \setminus \{0\}$ converging to x_0, we conclude that $g(x) \to L$ as $x \to x_0$ through S.

Or, we can use ε and δ. For any $\varepsilon > 0$ there exists $\delta_\varepsilon > 0$ such that if $x \in S$ and $0 < |x - x_0| < \delta_\varepsilon$ then $|f(x) - L| < \varepsilon$ and $|h(x) - L| < \varepsilon$. Expand these as

$$L - \varepsilon < f(x) < L + \varepsilon; \qquad L - \varepsilon < h(x) < L + \varepsilon$$

and use also the sandwich inequality $f(x) \leq g(x) \leq h(x)$ to give

$$L - \varepsilon < f(x) \leq g(x) \leq h(x) < L + \varepsilon.$$

From this extract

$$L - \varepsilon < g(x) < L + \varepsilon$$

or equivalently $|g(x) - L| < \varepsilon$. Since this is true for all $x \in S$ with $0 < |x - x_0| < \delta_\varepsilon$, we conclude that $g(x) \to L$ as $x \to x_0$.

8.6. If $x \to 0+$ then $1/x \to \infty$ so $g(x) = f(1/x) \to a$; this gives the right limit $g(0+) = a$. Similarly, if $x \to 0-$ then $1/x \to -\infty$ so $g(x) = f(1/x) \to b$ and we have $g(0-) = b$. We know that a function has a left limit at a point if and only if it has equal left and right limits there, so g has a limit at 0 if and only if $a = b$ (and the limit is this common value).

OR, with full eplisonics: given $\varepsilon > 0$ there exists $X_\varepsilon > 0$ such that if $x > X_\varepsilon$ then $|f(x) - a| < \varepsilon$; if $0 < x < 1/X_\varepsilon$ then $1/x > X_\varepsilon$ so $|f(1/x) - a| < \varepsilon$; this is the definition of $f(1/x) \to a$ as $x \to 0+$, with $\delta_\varepsilon = 1/X_\varepsilon$. The argument for the left limit is similar.

8.7. By definition, for $x \neq 0$ we have

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

We can divide the x into the terms of the series, break off the leading term and bring out another factor of x to give

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

If $|x| \leq 1$ then $|x^{2n-1}| \leq 1$ for all $n \in \mathbb{N}$. We can estimate the sum on the RHS by

$$\left|\sum_{n=1}^{\infty}(-1)^n \frac{x^{2n-1}}{(2n+1)!}\right| \leq \sum_{n=1}^{\infty} \frac{1}{(2n+1)!} =: C$$

(this series converges by e.g. the ratio test) so we have

$$\left|\frac{\sin(x)}{x} - 1\right| \leq C|x|$$

The required limit now follows by the Sandwich Theorem.

Similar arguments gives $(\cos(x)-1)/x^2 \to -1/2$ and $(\exp(x)-1)/x \to 1$ as $x \to 0$.

Exercises 9

Questions (or question parts) marked with a star ★ constitute Assignment 9; please submit on Moodle by the due date. Assignments make up 5% of the module mark.

9.1. Recall that a function $f:\mathbb{R}\to\mathbb{R}$ is called *odd* if $f(-x)=-f(x)$ for all $x\in\mathbb{R}$ and *even* if $f(-x)=f(x)$ for all $x\in\mathbb{R}$. Show that if f is odd and f is continuous at x_0 then f is continuous at $-x_0$. Repeat with f even.

9.2. If $I\subseteq\mathbb{R}$ is an interval, a function $f:I\to\mathbb{R}$ is said to be a *Lipschitz function* or to satisfy a *Lipschitz condition* if there is a constant $k\geq 0$ (a *Lipschitz constant*) such that $|f(x)-f(y)|\leq k|x-y|$ for all $x,y\in I$. Show that every Lipschitz function $f:I\to\mathbb{R}$ is continuous. *Be careful of the case* $k=0$.

9.3. Show that $g:\mathbb{R}\to\mathbb{R}$ defined by $g(x)=x^2$ (which is certainly continuous) is not a Lipschitz function (as defined in Exercise 9.2).

9.4. Show that the absolute value function, $f:\mathbb{R}\to\mathbb{R}$ defined by $f(x)=|x|$, is a Lipschitz function (Exercise 9.2) and is hence continuous.

9.5. (The Dirichlet function)

(a) We know (Theorem 3.21) that any interval (a,b) with $a<b$ contains a rational number. Show that it also contains an irrational number.

(b) Define $f:\mathbb{R}\to\mathbb{R}$ by

$$f(x)=\begin{cases}1 & \text{if } x \text{ is rational}\\ 0 & \text{if } x \text{ is irrational}\end{cases}$$

Show that if $x_0\in\mathbb{R}$ and $\delta>0$ then there exists $x\in\mathbb{R}$ such that $|x-x_0|<\delta$ but $|f(x)-f(x_0)|=1$. Deduce that f is not continuous at x_0 (so f is everywhere discontinuous).

⚠ 9.6. Suppose $I\subseteq\mathbb{R}$ is an interval and $f:I\to\mathbb{R}$ is such that, for every $x\in I$, $f(y)$ converges to a limit as $y\to x$ (through I). Define $g:I\to\mathbb{R}$ by

$$g(x)=\lim_{y\to x}f(y)$$

Show that g is continuous (on I).

9.7. Suppose $f:[a,b]\to[a,b]$ is continuous. Apply the Intermediate Value Theorem to $g(x)=f(x)-x$ to show that there exists $x_0\in[a,b]$ such that $f(x_0)=x_0$.

9.8. We have used on several occasions the algebraic identity

$$\sqrt{x}-\sqrt{y}=\frac{x-y}{\sqrt{x}+\sqrt{y}}$$

(for $x, y \geq 0$, not both zero). The corresponding identity for cube roots is

$$x^{1/3} - y^{1/3} = \frac{x - y}{x^{2/3} + (xy)^{1/3} + y^{2/3}}$$

(valid whenever the denominator on the RHS is non-zero). Use this to show, without using the Inverse Function Theorem, that $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^{1/3}$ is a continuous function. *This is intended to illustrate how much easier it is to use the Inverse Function Theorem than to grapple bare-handedly even with a fairly simple example! Suggestion: in finding the limit at some point x_0, treat the three cases $x_0 < 0$, $x_0 > 0$ and $x_0 = 0$ separately. Theorem 7.3 and Exercise 5.1 are relevant. The results from Exercise 9.1 can reduce the overall amount of work.*

★ 9.9. Define $f : [0, \infty) \to \mathbb{R}$ by $f(x) = x^{1/2} + x^{1/3}$.

(a) Explain briefly why $f(x) \to \infty$ as $x \to \infty$, why f is continuous and why f is strictly increasing.

(b) We know from the Intermediate Value Theorem that the range (image) $f([0, \infty))$ of f is an interval. What is this interval?

(c) What conclusions can we draw from the Inverse Function Theorem (for continuous functions)?

★ 9.10. Suppose $x, y \in \mathbb{R}$ with $x \neq y$ and $y \neq 0$ and suppose $n \in \mathbb{N}$. Use the standard formula for a finite geometric sum to evaluate

$$\sum_{k=0}^{n-1} \left(\frac{x}{y}\right)^k$$

and deduce that

$$\frac{x^n - y^n}{x - y} = \sum_{k=0}^{n-1} x^k y^{n-1-k}$$

Verify that this is also true if $y = 0$ (and $x \neq y$). Use this to evaluate

$$\lim_{y \to x} \frac{x^n - y^n}{x - y}$$

As presented, this is just a calculation involving limits; but it can also be interpreted as finding the derivative of $x \mapsto x^n$. Compare with Example 21.2.

★ Hand in.
⚠ Harder.

Exercises 9: Suggested Solutions

9.1. We can do this using ε and δ. Suppose f is odd and continuous at x_0.Then for any $\varepsilon > 0$ there exists $\delta_\varepsilon > 0$ such that if $|x - x_0| < \delta_\varepsilon$ then $|f(x) - f(x_0)| < \varepsilon$. Now, if $|x - (-x_0)| < \delta_\varepsilon$ then $|(-x) - x_0| < \delta_\varepsilon$ so $|f(-x) - f(x_0)| < \varepsilon$ (*). Using the oddness of f, this gives $|-f(x) + f(x_0)| < \varepsilon$, and hence $|f(x) - f(x_0)| < \varepsilon$. This shows that f is continuous at $-x$.

Now suppose f is even. The argument up to the point marked (*) is identical. From there, using the evenness of f, this gives $|f(x) - f(x_0)| < \varepsilon$, showing that f is continuous at $-x$.

Alternatively, we could use a sequential argument. If f is odd and continuous at x_0, consider a sequence $x_n \to -x_0$. (†) Then $f(x_n) = -f(-x_n)$ and $-x_n \to x_0$, so continuity at x_0 implies that $f(x_n) \to -f(x_0) = f(-x_0)$. This shows that f is continuous at $-x_0$.

Now suppose f is even. The argument up to the point marked (†) is identical. From there, using the evenness of f, $f(x_n) = f(-x_n)$ and $-x_n \to x_0$, so continuity at x_0 implies that $f(x_n) \to f(x_0) = f(-x_0)$. This shows that f is continuous at $-x_0$.

9.2. We can use sequences: if $x_0 \in I$ and $x_n \to x_0$ as $n \to \infty$ then $|f(x_n) - f(x_0)| \leq k|x_n - x_0|$; the RHS tends to zero, so the LHS must also tend to zero, showing that f is continuous at x_0 and hence on I.

We can also argue in ε-δ style. The core argument is: given $x_0 \in I$ and $\varepsilon > 0$, let $\delta_\varepsilon = \varepsilon/k$; then, if $x \in I$ and $|x - x_0| < \delta_\varepsilon$, we have $|f(x) - f(x_0)| < k\varepsilon/k = \varepsilon$, showing that f is continuous at any x_0 and hence on I.

We have to be slightly careful, though: this ε-δ argument does not work if $k = 0$. If $k = 0$, we can observe that $|f(x) - f(y)| \leq k|x-y| \implies f(x) = f(y)$, so f is constant and hence continuous. Or we could note that if k is a Lipschitz constant for f then so is any number greater than k, so the previous argument works with any fixed $k > 0$.

Note for later in the course: the formula $\delta_\varepsilon = \varepsilon/k$ contains no reference to x_0. This shows that f is in fact uniformly continuous on I.

9.3. If we had a Lipschitz constant k for g, then we would have

$$\left|\frac{x^2 - y^2}{x - y}\right| \leq k$$

for all $x, y \in \mathbb{R}$ with $x \neq y$; equivalently,

$$|x + y| \leq k$$

for all $x, y\mathbb{R}$ with $x \neq y$; but the LHS is plainly unbounded, so no such constant can exist.

9.4. This is the reverse triangle inequality (Theorem 6.1):

$$\big||x| - |y|\big| \leq |x - y|$$

so $f(x) = |x|$ defines a Lipschitz function with Lipschitz constant 1.

9.5. (a) There are lots of ways to do this. Possibly the simplest is to use Theorem 3.21 to choose rational numbers a_0 and b_0 with $a < a_0 < b_0 < b$, then choose an irrational $h \in (0,1)$ (e.g. $h = \sqrt{2}/2$). Having done this, it is straightforward to check that $(1-h)a_0 + hb_0$ is irrational and that $a < a_0 < (1-h)a_0 + hb_0 < b_0 < b$.

(b) Suppose x_0 is an irrational number, so $f(x_0) = 0$. Then (Theorem 3.21) for any $\delta > 0$, there exists a rational number $x \in (x_0 - \delta, x_0 + \delta)$; $f(x) = 1$ so $|f(x) - f(x_0)| = 1$.

Now suppose x_0 is a rational number, so $f(x_0) = 1$. Then (previous part of this exercise) for any $\delta > 0$, there exists an irrational number $x \in (x_0 - \delta, x_0 + \delta)$; $f(x) = 0$ so $|f(x) - f(x_0)| = 1$.

In either case, the ε definition of continuity at x_0 fails at $\varepsilon = 1$.

9.6. Fix $x_0 \in I$ and suppose $\varepsilon > 0$. There exists $\delta_0 > 0$ such that if $y \in I$ and $0 < |y - x_0| < \delta_0$ then $|f(y) - g(x_0)| < \varepsilon/2$ (A). Suppose $|x - x_0| < \delta_0$. There exists $\delta > 0$ such that if $y \in I$ and $0 < |y - x| < \delta$ then $|f(y) - g(x)| < \varepsilon/2$ (B). Combining (A) and (B) using the triangle inequality, if $y \in I$ satisfies both $0 < |y - x_0| < \delta_0$ and $0 < |y - x| < \delta$ then

$$|g(x) - g(x_0)| \le |g(x) - f(y)| + |f(y) - g(x_0)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

If any such y (such that $0 < |y - x_0| < \delta_0$ and $0 < |y - x| < \delta$) exists, this is enough to show that $g(x) \to g(x_0)$ as $x \to x_0$, which establishes continuity of g at an arbitrary $x_0 \in I$ and hence on I.

But does such a y exist? Yes: because $|x - x_0| < \delta_0$, any y sufficiently close to x (within distance $\delta_0 - |x - x_0|$) satisfies $|y - x_0| < \delta_0$; so, any $y \in I$ with $0 < |y - x| < \min\{\delta, |x - x_0|, \delta_0 - |x - x_0|\}$ will do (the $|x - x_0|$ term in the minimum is to ensure $0 < |y - x_0|$).

9.7. g as defined in the question is certainly continuous. Also, since $f : [a,b] \to [a,b]$, $f(a) \ge a$ and $f(b) \le b$, from which we see that $g(a) \le 0$ and $g(b) \ge 0$. By the IVT, we have $g(x_0) = 0$ for some $x_0 \in [a,b]$; that is, $f(x_0) = x_0$ for some $x_0 \in [a,b]$.

9.8. As suggested, consider three separate cases. Start with the case $x_0 > 0$. We can (localisation) consider only x such that $|x - x_0| < x_0$; in particular, $x > 0$. For such x we can write

$$|x^{1/3} - x_0^{1/3}| = \frac{|x - x_0|}{x^{2/3} + (xx_0)^{1/3} + x_0^{2/3}} < \frac{|x - x_0|}{x_0^{2/3}} \qquad (*)$$

Now, given $\varepsilon > 0$ we let $\delta_\varepsilon = \min\{x_0, x_0^{2/3}\varepsilon\}$. If $|x - x_0| < \delta_\varepsilon$ then $|x - x_0| < x_0$ for $x > 0$ and inequality (*) holds; because $|x - x_0| < x_0^{2/3}\varepsilon$, we have $|x^{1/3} - x_0^{1/3}| < \varepsilon$, i.e. $|f(x) - f(x_0)| < \varepsilon$. This shows that f is continuous at x_0.

The result for $x_0 < 0$ follows similarly or (because f is odd) from this and Exercise 9.1.

We are left with the case $x_0 = 0$ (cf. Exercise 5.1). We want to reach the conclusion $|x^{1/3} - 0| < \varepsilon$, from the premise $|x - 0| < \delta_\varepsilon$; $\delta = \varepsilon^3$ therefore seems to be an appropriate choice.

Give $\varepsilon > 0$, let $\delta = \varepsilon^3$. If $|x - 0| < \delta_\varepsilon$ then $|x| < \varepsilon^3$; taking cube roots, we have $|x^|1/3 < \varepsilon$, which we can write as $|x^{1/3} - 0| < \varepsilon$, i.e. $|f(x) - f(0)| < \varepsilon$, showing that f is continuous at 0.

Note for later in the course: the formula $\delta_\varepsilon = \min\{x_0, \varepsilon/x_0^{2/3}$ refers to x_0. This suggests that f is not uniformly continuous, and indeed it is not (because it has unbounded derivative).

9.9.

(a) The function f is continuous and strictly increasing because it is the sum of two continuous and strictly increasing functions (see example after the Inverse Function Theorem for Continuous Functions in the notes). Also, these two functions tend to ∞ at ∞, so the same is true for their sum.

(b) For any $x \geq 0$, we have $f(x) \geq 0$, so the range of f is contained in $[0, \infty)$. Also, $f(0) = 0$, f is unbounded and the range is an interval; the range must therefore contain $[0, \infty)$. Combining these, we see that the range is exactly $[0, \infty)$

(c) The Inverse Function Theorem for Continuous Functions now shows us that f is invertible between $[0, \infty)$ and itself and that the inverse function f^{-1} is continuous and strictly increasing,

9.10. The geometric series formula gives us

$$\sum_{k=0}^{n-1} (x/y)^k = \frac{(x/y)^n - 1}{x/y - 1}$$

which can be written as

$$\sum_{k=0}^{n-1} x^k y^{-k} = \frac{(x^n y^{-n} - 1}{xy^{-1} - 1}$$

Multiplying both sides by y^{n-1} (to recreate the sum given in the question) gives

$$\sum_{k=0}^{n-1} x^k y^{n-1-k} = \frac{(x^n y^{-1} - y^{n-1}}{xy^{-1} - 1}$$

and multiplying numerator and denominator on the RHS by y gives

$$\sum_{k=0}^{n-1} x^k y^{n-1-k} = \frac{(x^n - y^n}{x - y}$$

as required. If $y = 0$ (so $x \neq 0$, because $y \neq x$ throughout), we can verify the identity directly: the RHS (as written above, or the LHS as stated in the question) is x^{n-1} and the LHS is also x^{n-1} because the first term is x^{n-1} and all the subsequent terms are zero. Letting $y \to x$ in the summand $x^k y^{n-1-k}$ gives $x^k x^{n-1-k} = x^{n-1}$; there are n terms in the sum, so the sum converges to nx^{n-1} as $y \to x$. By the above identity,

$$\lim_{y \to x} \frac{x^n - y^n}{x - y} = nx^{n-1}$$

Exercises 10

Questions (or question parts) marked with a star ★ constitute Assignment 10; please submit on Moodle by the due date. Assignments make up 5% of the module mark.

10.1. Suppose $I \subseteq \mathbb{R}$ is an interval and $f : I \to \mathbb{R}$ is differentiable at $x \in I$. Use the ε-δ definition of limit to show that

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

This emphasises that the h is the definition of derivative can be positive or negative.

★ 10.2. Show from the definition of the derivative (i.e. not using the chain rule) that if $f : \mathbb{R} \to \mathbb{R}$ is differentiable (at every point) and $g : \mathbb{R} \to \mathbb{R}$ is defined by $g(x) = f(ax)$ for some fixed $a \in \mathbb{R}$ then g is differentiable and, for all $x \in \mathbb{R}$, $g'(x) = af'(ax)$. *Consider the cases $a = 0$ and $a \neq 0$ separately; note that if $h \to 0$ then $ah \to 0$.*

10.3. Define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} x^2 & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

Show that f is differentiable on $\mathbb{R}$. At what points is f' differentiable? *Most of this can be done by differentiating constants or powers of x, in the usual way; you only need explicit limit calculations at one point. See the counterexample in Theorem 21.4 for a similar sort of calculation.*

10.4. A basic form of the l'Hôpital rule: suppose $I \subseteq \mathbb{R}$ is an interval, $f, g : I \to \mathbb{R}$, $x_0 \in I$, $f(x_0) = g(x_0) = 0$, f and g are both differentiable at x_0 and $g'(x_0) \neq 0$. We can (Theorem 21.3) write

$$f(x_0+h) = f(x_0) + hp(h)$$

where p is continuous at 0 and $p(0) = f'(x_0)$ and a similar expression for g. Deduce from these that, as $h \to 0$,

$$\frac{f(x_0+h)}{g(x_0+h)} \to \frac{f'(x_0)}{g'(x_0)}$$

⚠ 10.5. Suppose $f, g, \phi : (a, b) \to \mathbb{R}$ are functions with $f(x) \leq \phi(x) \leq g(x)$ for all $x \in (a, b)$ and that there is a point $x_0 \in (a, b)$ such that f and g are both differentiable at x_0 and $f(x_0) = g(x_0)$. Show that:

(a) $f'(x_0) = g'(x_0)$

(b) ϕ is differentiable at x_0 and $\phi'(x_0) = f'(x_0) = g'(x_0)$

This is an analogue of the Sandwich Theorem, for derivatives. The notation and layout reflects the sequence of events in the proof: first prove (a), then give a different argument to prove (b). Note that differentiability at x_0 is a hypothesis for f and g, but a conclusion for ϕ.

★ 10.6. Suppose f and g are real functions, continuous on an interval $[a,b]$ with $a < b$ and differentiable on (a,b).

(a) Under the assumption that $g(b) \neq g(a)$, find $m \in \mathbb{R}$ such that $f(a) + mg(a) = f(b) + mg(b)$.

(b) Apply Rolle's Theorem to $h(x) = f(x) + mg(x)$ on $[a,b]$ to show that there exists $x_0 \in (a,b)$ with

$$f'(x_0)(g(b) - g(a)) = g'(x_0)(f(b) - f(a))$$

(c) If $g(a) = g(b)$, is the conclusion from the previous part still true?

10.7. Suppose we have a function $\log : (0,\infty) \to \mathbb{R}$ with the properties that $\log(1) = 0$ and $\log'(x) = 1/x$ (such a function must indeed be the (natural) logarithm, but we haven't proved that and you don't need to know that). Given $x > 0$, apply the Mean Value Theorem to the interval with endpoints 1 and x to show that $\log(x) \leq x - 1$ for all $x \in (0,\infty)$, with $\log(x) = x - 1$ if and only if $x = 1$. *Handle $x < 1$, $x = 1$ and $x > 1$ as separate cases.*

10.8. Suppose $a < b$, f is differentiable on $[a,b]$ and f' is strictly increasing on $[a,b]$. Given $t \in (0,1)$, let $c = (1-t)a + tb$, so $c \in (a,b)$. Apply the Mean Value Theorem to f on $[a,c]$ and on $[c,b]$ to deduce that

$$\frac{f(c) - f(a)}{c - a} < \frac{f(b) - f(c)}{b - c}$$

Rearrange this to show that

$$f((1-t)a + tb) < (1-t)f(a) + tf(b)$$

This property is called convexity*: the chord between $(a,f(a))$ and $(b,f(b)$ lies above the graph of f.*

▲ 10.9. Leading to a theorem of Darboux, surprisingly little-known.

(a) Suppose $a < b$ and $g : [a,b] \to \mathbb{R}$ is differentiable, with $g'(a) < 0$ and $g'(b) > 0$. Show that there exist points a', b' in (a,b) such that $g(a') < g(a)$ and $g(b') < g(b)$. Mimic the roof of Rolle's Theorem to conclude that $g'(x_0) = 0$ for some $x_0 \in (a,b)$.

(b) Deduce that derivatives have the intermediate value property: Suppose I is an interval, $f : I \to \mathbb{R}$ is differentiable and $a,b \in I$ with $a < b$. If y is any number such that $f'(a) \leq y \leq f'(b)$ or $f'(b) \leq y \leq f'(a)$ then there exists $x_0 \in [a,b]$ such that $f'(x_0) = y$.

★ Hand in.
▲ Harder.

Exercises 10: Suggested Solutions

10.1. Start with the ε-δ definition of limit, as applied to the difference quotient and derivative. "For any $\varepsilon > 0$ there exists $\delta_\varepsilon > 0$ such that if $0 < |h| < \delta_\varepsilon$ and $x + h \in I$ then

$$\left|\frac{f(x+h)-f(x)}{h} - f'(x)\right| < \varepsilon\text{"}$$

Now, $|-h| < \delta_\varepsilon \iff |h| < \delta_\varepsilon$, so this can equally be written as "For any $\varepsilon > 0$ there exists $\delta_\varepsilon > 0$ such that if $0 < |h| < \delta_\varepsilon$ and $x - h \in I$ then

$$\left|\frac{f(x-h)-f(x)}{-h} - f'(x)\right| < \varepsilon\text{"}$$

10.2. In the case $a = 0$, we have $g(x) = f(0)$ for all x, so g is constant and $g'(x) = 0 = af'(ax)$ for all x, trivially.

In the more substantive case $a \neq 0$, the simplest, and essentially correct, approach is to fix $x \in \mathbb{R}$ and write

$$\frac{g(x+h)-g(x)}{h} = \frac{f(ax+ah)-f(ax)}{h} = a\frac{f(ax+ah)-f(ax)}{ah} \to af'(ax)$$

as $h \to 0$ (because $ah \to 0$ as $h \to 0$), concluding that $g'(x) = af'(ax)$.

This does, however, contain an awkward point: why exactly can we exchange $h \to 0$ with $ah \to 0$? One way to nail this down is to go right down to the definition: if

$$0 < |h| < \delta \implies \left|\frac{f(ax+h)-f(ax)}{h} - f'(ax)\right| < \varepsilon$$

then

$$0 < |ah| < \delta \implies \left|\frac{f(ax+ah)-f(ax)}{ah} - f'(ax)\right| < \varepsilon$$

or equivalently

$$0 < |h| < \delta/|a| \implies \left|\frac{f(ax+ah)-f(ax)}{ah} - f'(ax)\right| < \varepsilon$$

which (surrounded by the usual "for any $\varepsilon > 0$ there exists $\delta_\varepsilon \ldots$" framework) shows that $(g(x+h)-g(x))/(ah) \to f'(ax)$ as $h \to 0$. We can then use the algebra of limits to multiply by a and draw the final conclusion.

Another way to be completely clear is to write

$$f(ax+h) = f(ax) + hp(h)$$

where $p(0) = f'(ax)$, so p is continuous on $\mathbb{R}$ (p is continuous at 0 because f is differentiable at ax; p is continuous away from 0 because $p(h) = (f(ax+h) - f(ax))/h$ and f is differentiable, hence continuous, on $\mathbb{R}$). We then have

$$g(x+h) = f(ax+ah) = f(ax) + ahp(ah) = g(x) + ahp(ah)$$

Now, as $h \to 0$, $ah \to 0$ so $p(ah) \to f'(ax)$ by Theorem 19.5 and we conclude that $g'(x) = af'(ax)$. In this form, the fact that we have a composition ("function of a function") is made very explicit.

10.3. On $(0,\infty)$, $f(x) = x^2$; by localisation and standard results of differentiation, $f'(x) = 2x$. On $(-\infty,0)$, $f(x) = 0$; by localisation and standard results of differentiation, $f'(x) = 0$. At 0, we write down the difference quotient:

$$\frac{f(0+h)-f(0)}{h} = \begin{cases} \frac{h^2-0}{h} = h & \text{if } h > 0 \\ 0 & \text{if } h < 0 \end{cases}$$

Now, we can see that the left and right limits as $h \to 0$ are both 0, so f is differentiable at 0 with $f'(0) = 0$. In summary:

$$f'(x) = \begin{cases} 2x & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

Again by localisation, f' is differentiable everywhere except possibly at 0. At 0, we write down another difference quotient:

$$\frac{f'(0+h)-f'(0)}{h} = \begin{cases} \frac{2h-0}{h} = 2 & \text{if } h > 0 \\ 0 & \text{if } h < 0 \end{cases}$$

Now, the left and right limits as $h \to 0$ are respectively 0 and 2, so there is no limit as $h \to 0$ and f' is not differentiable at 0.

10.4. As instructed, write

$$\begin{aligned} f(x_0+h) &= f(x_0) + hp(h) \\ g(x_0+h) &= g(x_0) + hq(h) \end{aligned}$$

where p and q are continuous at 0, $p(0) = f'(x_0)$, $q(0) = g'(x_0)$. Substituting $f(x_0) = g(x_0) = 0$, we have

$$\frac{f(x_0+h)}{g(x_0+h)} = \frac{hp(h)}{hq(h)} = \frac{p(h)}{q(h)}$$

Now, p and q have limits $f'(x_0)$ and $g'(x_0)$ at 0 and we are given $g'(x_0) \neq 0$ so (Algebra of Limits) $p(h)/q(h) \to f'(x_0)/g'(x_0)$ as $h \to 0$.

10.5. (a) We can write

$$\begin{aligned} f(x_0+h) &= f(x_0) + hp(h) \\ g(x_0+h) &= g(x_0) + hq(h) \end{aligned}$$

where p and q are continuous at 0, $p(0) = f'(x_0)$, $q(0) = g'(x_0)$. Writing down the inequality $f \leq g$ and cancelling the equal terms $f(x_0) = g(x_0)$ gives

$$hp(h) \leq hq(h)$$

If we divide by h, we conclude that

$$\begin{aligned} p(h) &\leq q(h) & (h > 0) \\ p(h) &\geq q(h) & (h < 0) \end{aligned}$$

Now take left and right limits: as $h \to 0+$ we get $f'(x_0) \leq g'(x_0)$ and as $h \to 0-$ we get $f'(x_0) \geq g'(x_0)$; between them, we get $f'(x_0) = g'(x_0)$.

(b) Now, we have from the sandwiching inequality $f(x) \leq \phi(x) \leq g(x)$ that $f(x_0) = \phi(x_0) = g(x_0)$. We therefore have

$$f(x_0+h) - f(x_0) \leq \phi(x_0+h) - \phi(x_0) \leq g(x_0+h) - g(x_0)$$

For $h > 0$, we can divide by h to give

$$\frac{f(x_0+h) - f(x_0)}{h} \leq \frac{\phi(x_0+h) - \phi(x_0)}{h} \leq \frac{g(x_0+h) - g(x_0)}{h}$$

and let $h \to 0+$ to give $f'(x_0) \leq \phi'(x_0+) \leq g'(x_0)$; but $f'(x_0) = g'(x_0)$ by hypothesis, so $f'(x_0) = \phi'(x_0+) = g'(x_0)$ Similarly, for $h < 0$ we can divide by h to give

$$\frac{g(x_0+h) - g(x_0)}{h} \leq \frac{\phi(x_0+h) - \phi(x_0)}{h} \leq \frac{f(x_0+h) - f(x_0)}{h}$$

and let $h \to 0-$ to give $g'(x_0) \leq \phi'(x_0-) \leq f'(x_0)$; again, $f'(x_0) = g'(x_0)$ by hypothesis, so $f'(x_0) = \phi'(x_0-) = g'(x_0)$.

Now, $\phi'(x_0-) = \phi'(x_0+)$ so $\phi'(x_0)$ exists and is equal to this common value; that is, $\phi'(x_0) = f'(x_0) = g'(x_0)$.

10.6. First case: $x > 1$. MVT gives us

$$\frac{\log(x) - \log(1)}{x-1} = \frac{1}{\xi}$$

for some $\xi \in (1, x)$; in particular, $\xi > 1$ so $1/\xi < 1$. Multiply by $x - 1 > 0$ and substitute $\log(1) = 0$ to give $\log(x) < x - 1$.

Second case: $x < 1$. MVT gives us

$$\frac{\log(1) - \log(x)}{1-x} = \frac{1}{\xi}$$

for some $\xi \in (x, 1)$; in particular, $\xi < 1$ so $1/\xi > 1$. Multiply by $1 - x > 0$ and substitute $\log(1) = 0$ to give $-\log(x) > 1 - x$; now multiply by -1 to give $\log(x) < x - 1$

Final case: $x = 1$. This is does not require further use of the MVT. Firstly, $\log(1) = 1 - 1$. Secondly $\log(x) = x - 1$ is not possible for $x \neq 1$ because (by the previous results) we then have $\log(x) < x - 1$.

10.7. (a) Solving $f(a) + mg(a) = f(b) + mg(b)$ gives $m = -(f(b) - f(a))/(g(b) - g(a))$

(b) We now have $h(b) = h(a)$ (and h continuous on $[a, b]$ and differentiable on (a, b)) and can use Rolle's Theorem to find x_0 such that $h'(x_0) = 0$, i.e. $f'(x_0) + mg'(x_0) = 0$. Substituting the definition of m,

$$f'(x_0) - \frac{f(b) - f(a)}{g(b) - g(a)} g'(x_0) = 0$$

which rearranges to the equation in the question.

(c) If $g(b) = g(a)$, then Rolle's Theorem applied to g gives x_0 with $g'(x_0) = 0$. With this value of x_0, the equation in the question reads $0 = 0$, which is certainly true, so the same conclusion can indeed be reached.

10.8. Applying the MVT on $[a,c]$ gives

$$\frac{f(c)-f(a)}{c-a} = f'(x_1)$$

for some $x_1 \in (a,c)$. Applying the MVT on $[c,b]$ gives

$$\frac{f(b)-f(c)}{b-c} = f'(x_2)$$

for some $x_2 \in (c,b)$. Now, $x_1 < c < x_2$ so, by the hypothesis that f' is strictly increasing, $f'(x_1) < f'(x_2)$. This gives

$$\frac{f(c)-f(a)}{c-a} < \frac{f(b)-f(c)}{b-c}$$

Now, note that $c-a = (1-t)a + tb - a = t(b-a)$ and similarly $b-c = (1-t)(b-a)$. Making this substitution in the denominators of the previous inequality and cancelling all the $(b-a)$ terms gives

$$\frac{f(c)-f(a)}{t} < \frac{f(b)-f(c)}{1-t}$$

Multiply by $t(1-t)$ to give

$$(1-t)f(c) - (1-t)f(a) < tf(b) - tf(c)$$

which rearranges to

$$f(c) < (1-t)f(a) + tf(b)$$

Substituting $c = (1-1)a + tb$ gives the inequality stated in the question.

10.9. (a) The picture to have in mind here is that if g is decreasing near a and increasing near b then it must have a minimum somewhere in between. To make this precise:

We are given $g'(a) < 0$, and

$$\frac{g(a+h)-g(a)}{h} \to g'(a)$$

as $h \to 0$ (implicitly, $h \to 0+$ as we're at the left endpoint). The LHS is therefore negative for small h such that $a+h \in [a,b]$, i.e. for small positive h; it follows that $g(a+h) - g(a) < 0$ for small positive h and we can let $a' = a+h$ for some such h.

Similarly, we are given $g'(b) > 0$, and

$$\frac{g(b+h)-g(b)}{h} \to g'(b)$$

as $h \to 0$ (implicitly, as $h \to 0-$). The LHS is therefore positive for small h such that $b+h \in [a,b]$, i.e. for small negative h; it follows that $g(b+h)-g(b) < 0$ for small negative h and we can let $b' = b+h$ for some such h.

Now, the Extreme Value Theorem (Theorem 19.13) tells us that g takes on its minimum value at some $x_0 \in [a,b]$. We cannot have $x_0 = a$ or $x_0 = b$, because $g(a') < g(a)$ and $g(b') < g(b)$, so we must have $x_0 \in (a,b)$. At an interior local extremum, the derivative is zero, so $g'(x_0) = 0$.

(b) Trivial cases: if $y = f'(a)$ or $y = f'(b)$, we can let $x_0 = a$ or $x_0 = b$. There are two substantive cases: $f'(a) < y < f'(b)$ and $f'(b) < y < f'(a)$. We consider only the first case, $f'(a) < y < f'(b)$; applying this to $-f$ yields the other case. For $x \in [a,b]$, let $g(x) = f(x)-xy$, so $g'(x) = f'(x)-y$ and hence $g'(a) < 0 < g'(b)$. We can therefore apply the previous part of the question to g on $[a,b]$ to find $x_0 \in (a,b)$ with $g'(x_0) = 0$, i.e. $f'(x_0) = y$.

Exercises 11

Questions (or question parts) marked with a star ★ constitute Assignment 11; please submit on Moodle by the due date. Assignments make up 5% of the module mark.

11.1. Suppose f is twice differentiable on an interval I and $x \in I$. Taylor's Theorem tells us that if $h \neq 0$ and $x+h \in I$ then there exists c between x and $x+h$ with

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2}f''(c) \implies f'(x) = \frac{f(x+h)-f(x)}{h} + \frac{h}{2}f''(c) \qquad (1)$$

which we can think of as a formula for the error in approximating $f'(x)$ by the difference quotient $(f(x+h)-f(x))/h$. Now suppose f is three times differentiable on I. By expanding $f(x+h)$ and $f(x-h)$, show that for some c between $x-h$ and $x+h$ we have

$$f'(x) = \frac{f(x+h)-f(x-h)}{2h} - \frac{h^2}{6}f'''(c) \qquad (2)$$

Can you see how to modify this to give a similarly accurate formula for the second derivative $f''(x)$? *You can get close to (2) using just Taylor's Theorem, but you'll need Darboux's Theorem, the final conclusion of Exercise 10.9, to finish. This is slightly tricky; think about turning a sum into an average. These are important formulae in Numerical Analysis; for small h, (2) gives a much more accurate approximation that the more obvious (1).*

11.2. Suppose f is defined by a power series

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

with non-zero radius of convergence. Show that the Nth order Taylor polynomial of f, as defined in Definition 25.3, is

$$P_N(x) = \sum_{n=0}^{N} a_n(x-x_0)^n$$

This ties together two ideas: that of Taylor's Theorem with remainder, applied to a given function, and that of using a convergent Taylor (=power) series to define a function. The Taylor polynomials are exactly the partial sums of the Taylor series. Don't try to prove this using Taylor's Theorem with remainder: it's all about differentiating power series.

★ 11.3. Define $f : (-1,1) \to \mathbb{R}$ by

$$f(x) = \sum_{k=0}^{\infty} \frac{x^{2k+1}}{2k+1}$$

(a) Verify that the power series has radius of convergence 1, so the definition of f on $(-1,1)$ makes sense.

(b) Differentiate the power series to see that

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

(c) Given that the inverse hyperbolic tangent function (strictly speaking, the branch that of the inverse function that is real on $(-1,1)$) has derivative $(\tanh^{-1})'(x) = 1/(1-x^2)$, deduce that $f(x) = \tanh^{-1}(x)$ for $x \in (-1,1)$.

11.4. Suppose $a, u_0 \in \mathbb{R}$ and that $u : \mathbb{R} \to \mathbb{R}$ satisfies the IVP $u'(t) = au(t)$ (for all $t \in \mathbb{R}$) and $u(0) = u_0$. Modify the opening argument in Section 27 to show that $u(t) = u_0 \exp(at)$.

11.5. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x\mathrm{e}^x$.

(a) Show that $f(x) \to \infty$ as $x \to \infty$.

(b) Find $f'(x)$. Find (the unique) $x_0 \in \mathbb{R}$ such that $f'(x_0) = 0$ and show that $f'(x) > 0$ for $x \in (x_0, \infty)$. Let $y_0 = f(x_0)$.

(c) It follows from the Inverse Function Theorem for Differentiable Functions (Theorem 24.1) that f has an inverse function $W = f^{-1} : (y_0, \infty) \to (x_0, \infty)$. Find the derivative W', expressing $W'(y)$ in terms of y and $W(y)$ (note that, because $f(W(y)) = y$, $\exp(W(y))$ can be expressed in a simpler way).

W is called Lambert's function*; solutions to many equations involving both polynomial and exponential terms can be expressed in terms of W.*

★ 11.6. Fix some $b \in \mathbb{R}$, with $b > 0$ and $b \neq 1$.

(a) From the definition of b^x ($x \in \mathbb{R}$), find the derivative of b^x with respect to x.

(b) For $y > 0$, take (natural) logarithms to solve the equation $b^x = y$, expressing x in terms of y and b. Explain briefly why your formula shows that the equation always has exactly one solution.

(c) The logarithm to base b of $y > 0$ is defined by $\log_b(y) = x$ where x is the unique real number such that $b^x = y$. Find the derivative of the function $\log_b : (0, \infty) \to \mathbb{R}$.

⚠ 11.7. Show that if $x \in \mathbb{R}$ then

$$\left(1 + \frac{x}{n}\right)^n \to \exp(x)$$

as $n \to \infty$. More generally, if $k > 0$, what happens to

$$\left(1 + \frac{x}{n^k}\right)^n$$

as $n \to \infty$?

- ★ Hand in.
- ⚠ Harder.

Exercises 11: Suggested Solutions

11.1. We can expand $f(x \pm h)$ as follows:

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \frac{h^3}{6}f'''(c^+) \tag{*}$$
$$f(x-h) = f(x) - hf'(x) + \frac{h^2}{2}f''(x) - \frac{h^3}{6}f'''(c^-)$$

where c^+ lies between x and $x+h$ and c^- lies between x and $x = h$. If we subtract these, half the terms cancel:

$$f(x+h) - f(x-h) = 2hf'(x) + \frac{h^3}{6}(f'''(c^+) + f'''(c^-))$$

and we can divide by $2h$, almost reaching the final answer:

$$\frac{f(x+h) - f(x-h)}{2h} = f'(x) + \frac{h^3}{12}(f'''(c^+) + f'''(c^-))$$

The last step is a little non-obvious: we can write the error term as

$$\frac{h^3}{6}\frac{f'''(c^-) + f'''(c^+)}{2}$$

The second factor is the average of $f'''(c^-)$ and $f'''(c^+)$, so it lies between $f'''(c^-)$ and $f'''(c^+)$. By Darboux's Theorem (Exercise 10.9)

$$\frac{f'''(c^-) + f'''(c^+)}{2} = f'''(c)$$

for some c between c^- and c^+, hence between $x-h$ and $x+h$. If we make this substitution, we obtain

$$\frac{f(x+h) - f(x-h)}{2h} = f'(x) + \frac{h^3}{6}f'''(c)$$

and a minor rearrangement gives (2)

To obtain a similar formula for the second derivative, look back at (*) and the following equation. In essence, by subtracting, we cancelled all the terms involving an even-order derivative, leaving only the first derivative, which we're approximating, and the third, which contains the error. If, instead, we add the equations, the odd-order derivatives will cancel, leaving only the function itself (order zero) and the second derivative. To make this work properly, we need one more term in the Taylor expansion, so the third-order terms completely cancel; we will then have the function, the second derivative and a fourth-derivative term which will contain the error. Assume f is four times differentiable and write

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \frac{h^3}{6}f^{(3)}(x) + \frac{h^4}{24}f^{(4)}(c^+)$$
$$f(x-h) = f(x) - hf'(x) + \frac{h^2}{2}f''(x) - \frac{h^3}{6}f^{(3)}(x) + \frac{h^4}{24}f^{(4)}(c^-)$$

Adding these gives, anticipating the need for the same averaging manoeuvre as in the previous part,

$$f(x+h)+f(x-h)=2f(x)+h^2f''(x)+\frac{h^4}{12}\frac{f^{(4)}(c^-)+f^{(4)}(c^+)}{2}$$

As above, the average of $f^{(4)}(c^-)$ and $f^{(4)}(c^+)$ lies between $f^{(4)}(c^-)$ and $f^{(4)}(c^+)$, so by Darboux's Theorem we can write $(f^{(4)}(c^-)+f^{(4)}(c^+))/2 = f^{(4)}(c)$ for some c between c^- and c^+, hence between $x-h$ and $x+h$. Making this substitution, moving the $f(x)$ term from right to left and dividing by h^2 gives

$$\frac{f(x+h)-2f(x)+f(x-h)}{h^2}=f''(x)+\frac{h^4}{12}f^{(4)}(c)$$

which is indeed similar to (2); the important point in applications is that both errors are multiple of h^2 so, as long as the $f^{(n)}(c)$ term is under control (which it will be if e.g. $f^{(n)}$) is continuous), both errors fall away at a similar speed when h is reduced.

11.2.

11.3.

(a) This follows from the ratio test, as usual: the absolute ratio of two consecutive terms is

$$\frac{2k+3}{2k+1}|x^2|\to x^2$$

as $k\to\infty$; we thus have convergence if $x^2<1$, i.e. if $|x|<1$, and divergence if $x^2>1$, i.e. $|x|>1$. That is, the radius of convergence is 1.

(b) Differentiating the series term-by-term, we see that

$$f'(x)=\sum_{k=0}^{\infty}x^{2k}=\sum_{k=0}^{\infty}(x^2)^k$$

This is a geometric series with common ratio x^2 and first term 1; the usual formula gives us

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

Now, we are given that $(\tanh^{-1})'=f'$; it follows that $(\tanh^{-1}-f)'=0$, so $(\tanh^{-1}-f)$ is constant. At 0. we have $\tanh^{-1}(0)=f(0)=0$, so we must have $f(x)=\tanh^{-1}(x)$ for all $x\in(-1,1)$.

11.4. Following the corresponding argument in the notes, differentiate $u(t)\exp(-at)$ w.r.t. t:

$$\frac{\mathrm{d}}{\mathrm{d}t}u(t)\exp(-at)=u(t)(-a\exp(at))+u'(t)\exp(-at)=0$$

because u satisfies the DE $u'=au$. It follows that $u(t)\exp(at)$ is constant as t varies. We can find the constant value by looking at $t=0$, because of the initial value $u(0)=u_0$; since $\exp(-a0)=1$, $u(t)\exp(-at)$ is constant at u_0, i.e. $u(t)\exp(-at)=u_0$ for all t. Rearranging, we get $u(t)=u_0\exp(at)$ for all $t\in\mathbb{R}$.

11.5. (a) As $x \to \infty$, $\mathrm{e}^x \to \infty$ so $x\mathrm{e}^x \to \infty$.

(b) We can differentiate f using the product rule:

$$f'(x) = x\exp(x) + \exp(x) = x\exp(x)(x+1)$$

We always have $\mathrm{e}^x > 0$, so $f'(x) = 0$ if and only if $x+1 = 0$; this gives $x_0 = -1$. If $x > x_0$ then $x+1 > 0$ and $\mathrm{e}^x > 0$, so $f'(x) > 0$. We have $y_0 = f(x_0) = -\mathrm{e}^{-1}$

(c) By the Inverse Function Theorem,

$$W'(y) = \frac{1}{f'(W(y)} = \frac{1}{(1+W(y))\exp(W(y)}$$

Now, by definition, $f(W(y)) = y$, i.e. $W(y)\exp(W(y)) = y$. We can therefore write $\exp(W(y)) = y/W(y)$ and simplify to

$$W'(y) = \frac{1}{y/W(y) + y} = \frac{W(y)}{y(1+W(y))}$$

(a) We have by definition $b^x = \exp(x\log(b))$. Differentiating using $\exp' = \exp$ and the chain rule gives

$$\frac{\mathrm{d}}{\mathrm{d}x} b^x = \exp'(x\log(b))\log(b) = \exp(x\log(b))\log(b) = b^x\log(b)$$

(b) Taking logarithms as instructed:

$$b^x = y \iff \log(b^x) = \log(y) \iff x\log(b) = \log(y) \iff x = \log(y)/\log(b)$$

This all makes sense because $y > 0$, so $\log(y)$ is well defined, and $b \neq 1$, so $\log(b) \neq 0$ and we can divide by it. Beyond this, the reason we always have a unique solution is the "if and only if" nature of the calculation: following the logic from left to right (i.e. the $\implies$ arrows) says that if there is a solution x, then we must have $x = \log(y)/\log(b)$, so in particular we have uniqueness; following the logic in the opposite direction says that $x = \log(y)/\log(b)$ describes a solution, so in particular we have existence.

(c) There are two equally natural approaches to this, based on the earlier parts of the question. One is to start from

$$\log_b(y) = \frac{\log(x)}{\log(b)}$$

from the previous part, which we can differentiate to give

$$\log_b'(y) = \frac{1}{x\log(b)}$$

Alternatively, we could use the inverse function theorem and the first part of the question:

$$\log_b'(y) = \frac{1}{b^{\log_b(y)}\log(b)} = \frac{1}{x\log(b)}$$

11.6. The key here is to take logarithms:

$$\log\left[\left(1+\frac{x}{n}\right)^n\right] = n\log\left(1+\frac{x}{n}\right)$$

Now, we expand $\log(1+x/n)$. We could use Taylor's Theorem, but it is simpler to go back to the definition of differentiability and write

$$\log(1+h) = \log(1) + hp(h) = hp(h)$$

where $p(h) \to \log'(1) = 1$ as $h \to 0$. Substituting this into the previous formula with $h = x/n$ gives

$$\log\left[\left(1+\frac{x}{n}\right)^n\right] = n(x/n)p(x/n) = xp(x/n) \to x$$

as $n \to \infty$. Now, taking exponentials if both sides and using the fact that exp is continuous,

$$\left(1+\frac{x}{n}\right)^n \to \exp(x)$$

as $n \to \infty$. We can try the same maneouvres with the more general formula:

$$\log\left[\left(1+\frac{x}{n^k}\right)^n\right] = n\log\left(1+\frac{x}{n^k}\right) = n(x/n^k)p(x/n^k) = xn^{1-k}p(x/n^k)$$

Now, as $n \to \infty$, $x/n^k \to 0$ (we are told $k > 0$) so $p(x/n^k) \to 1$. We also have $n^{1-k} \to 0$ if $k > 1$ and $n^{1-k} \to \infty$ if $k < 1$. So, except in the special case $x = 0$,

$$\left(1+\frac{x}{n^k}\right)^n \to \begin{cases} \infty & \text{if } 0 < k < 1 \\ \exp(x) & \text{if } k = 1 \\ 1 & \text{if } k > 1 \end{cases}$$

If $x = 0$, the sequence is constant at 1, independent of k.

Exercises 12

There is nothing to hand in from these exercises.

12.1. Suppose $c \in [a,b]$ and define $f : [a,b] \to \mathbb{R}$ by

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

Show that:

(a) For any partition P of $[a,b]$, $L(f,P) = 0$

(b) For any partition P of $[a,b]$, $U(f,P) > 0$

(c) If $0 < h \leq b-a$ then there is a partition P_h of $[a,b]$ with $U(f,P_h) = h$

Deduce that f is integrable and that $\int_a^b f = 0$. *For (c), don't try anything too complicated: there are lots of possible examples, but you can describe a suitable partition P_h with at most three subintervals, i.e. four points including the endpoints, such that $L(f,P_h) = h$. Writing down an explicit formula for such a partition is a bit fiddly, and isn't necessary if you describe it clearly enough. This exercise can be solved without knowing in advance that $L_a^b f \leq U_a^b f$.*

12.2. Suppose $a < b$ and $f : [a,b] \to \mathbb{R}$ is bounded. Define $\tilde{f} : [-b,-a] \to \mathbb{R}$ by $\tilde{f}(x) = f(-x)$. Show that if $P = (x_0, x_1, \dots, x_N)$ is a partition of $[a,b]$ then

$$\tilde{P} = (-x_N, -x_{N-1}, \dots, -x_0) = (\tilde{x}_0, \dots, \tilde{x}_N)$$

where $\tilde{x}_n = -x_{N-n}$ defines a partition of $[-b,-a]$ with

$$L(f,P) = L(\tilde{f},\tilde{Q}); \qquad U(f,P) = U(\tilde{f},\tilde{Q})$$

Deduce that f is integrable if and only if $\tilde{f}$ is integrable, in which case $\int_a^b f = \int_{-b}^{-a} \tilde{f}$. *This is a very basic form of integration by substitution. Linear changes of variables can all be handled like this, but require a little more work. Nonlinear changes of variables require quite a lot more work!*

12.3. This is a continuation of Example 29.10. It can be shown (see Exercise 12.4) that there is a polynomial P_k of degree $k+1$ with leading coefficient $1/(k+1)$ such that

$$\sum_{n=1}^{N} n^k = P_k(N)$$

(e.g. $P_1(N) = N(N+1)/2$ was used in Example 29.10 and you might have come across $P_2(N) = N(N+1)(2N+1)/6$ and $P_3(N) = N^2(N+1)^2/4$, which are often used as examples when teaching proof by induction). Use this fact to show using upper and lower sums and integrals that

$$\int_0^1 x^k \, \mathrm{d}x = \frac{1}{k+1}$$

⚠ 12.4. Prove that the polynomials P_k described in the previous exercise exist and have the stated leading term. These are called *Faulhaber polynomials*, if you want to look them up, but you might not find online accounts very accessible! One practical way to proceed is first to define

$$P_k(N) = \sum_{n=1}^{N} n^k$$

for $k \in \mathbb{N}_0$ (the $k = 0$ case is trivial but it is convenient to have it defined). The task is then to prove that P_k is a polynomial of degree $k+1$ with leading coefficient $k+1$. Consider the sum

$$\sum_{n=1}^{N} [(n+1)^{k+1} - n^{k+1}]$$

On the one hand, this is a telescopic series so there is a simple formula for its sum. On the other hand, the summand can be expanded using the binomial theorem; the n^{k+1} term cancels one term from the binomial expansion and the order of summation of the double sum can be changed to give an identity about a weighted sum of P_k terms. Everything required can then be deduced (by induction) from this identity.

12.5. This is essentially a generalisation of Example 29.10, which you might find useful in formulating your solution. Suppose $f : [a,b] \to \mathbb{R}$ is increasing. Show that f is bounded. For $N \in \mathbb{N}$, consider the partition created by dividing $[a,b]$ into N subintervals of of equal width $w = (b-a)/N$:

$$P_N = (a, a+w, a+2w, \ldots, b-w, b)$$

Explain why

$$L(f, P_N) = \sum_{n=1}^{N} wf(a+(n-1)w); \qquad U(f, P_N) = \sum_{n=1}^{N} wf(a+nw)$$

Show from these that

$$U(P,f) - L(P,f) = \frac{(b-a)(f(b)-f(a))}{N}$$

and deduce that f is integrable.

Deduce that every decreasing function on $[a,b]$ is integrable (think about $-f$).

12.6. Recall that the *Dirichlet function* (see Exercise 9.5) $f : \mathbb{R} \to \mathbb{R}$ is defined by

$$f(x) = \begin{cases} 1 & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is irrational} \end{cases}$$

Show that f is not integrable over any interval $[a,b]$ with $a < b$.

⚠ Harder.

Exercises 12: Suggested Solutions

12.1. Note that f takes on only the values 0 and 1, and it takes on the value 1 at only one point. Any interval $[x_{n-1}, x_n]$ of a partition $P = (x_0, x_1, \ldots, x_N)$ contains infinitely many points (the definition of a partition requires $x_{n-1} < x_n$), so the set of values taken on by f over a partition interval is

$$\{f(x) : x \in [x_{n-1}, x_n]\} = \begin{cases} \{0\} & \text{if } c \notin [x_{n-1}, x_n] \\ \{0,1\} & \text{if } c \in [x_{n-1}, x_n] \end{cases}$$

(a) For $P = (x_0, x_1, \ldots, x_N)$, we have

$$L(f,P) = \sum_{n=1}^{N} (x_n - x_{n-1}) \inf_{x \in [x_{n-1}, x_n]} f(x)$$

By the above remarks, we have the infimum of either $\{0\}$ or of $\{0,1\}$, both of which are zero. All of the terms in the sum are therefore zero, so $L(f,P) = 0$.

(b) Most of this is true on very generic grounds: every value taken on by f is non-negative, so all the suprema in the definition of the upper sum are non-negative, so the upper sum is non-negative. More explicitly, for $P = (x_0, x_1, \ldots, x_N)$, we have

$$U(f,P) = \sum_{n=1}^{N} (x_n - x_{n-1}) \sup_{x \in [x_{n-1}, x_n]} f(x)$$

By the above remarks, we have the supremum of either $\{0\}$ or of $\{0,1\}$, which are respectively 0 (for a subinterval not containing c) and 1 (for a subinterval containing c). The lower sum is thus the sum of the lengths of the subintervals of the partition that contain c (there can be either one or two such subintervals; there are two iff $x_n = c$ for some n with $1 \le n \le N-1$). Such a sum of lengths is plainly positive.

(c) Here we need the more explicit understanding of the lower sum from part (b) above. To make $U(f,P) = h$, we need a partition such that the sum of the lengths of the subintervals containing c is h. Any partition in which there is a single subinterval containing c and that subinterval has length h will work; this is possible iff $0 < h \le b - a$.

Now, all the lower sums are zero, so their supremum is zero; that is, $L_a^b f = 0$. All the upper sums are positive, so their infimum satisfies $U_a^b f \ge 0$. Every h with $0 < h \le b-a$ is an upper sum, so their infimum satisfies $U_a^b \le 0$. Combining these inequalities, we see that $L_a^b f = U_a^b f = 0$, so f is integrable over $[a,b]$ with $\int_a^b f = 0$.

12.2. Firstly, $\tilde{x}_0 = -x_N = -b$ and $\tilde{x}_N = -x_0 = -a$, so the first and last entries of $\tilde{P}$ are correct for a partition of $[-b,-a]$. Secondly, $\tilde{x}_n - \tilde{x}_{n-1} = -x_{N-n} + x + N + 1 - n > 0$ (because the (x_n) are in strictly increasing order) so the $(\tilde{x})_n$ are in strictly increasing order, making $\tilde{P}$ a partition of $[-b,-a]$.

Next, start with the definition of $L(f,P)$:

$$L(f,P) = \sum_{n=1}^{N} (x_n - x_{n-1}) \inf_{x \in [x_{n-1}, x_n]} f(x)$$

Change x_n to $-\tilde{x}_{N-n}$ and x_{n-1} to $-\tilde{x}_{N+1-n}$ throughout:

$$L(f,P) = \sum_{n=1}^{N} (-\tilde{x}_{N-n} + \tilde{x}_{N+1-n}) \inf_{x \in [-\tilde{x}_{N+1-n}, -\tilde{x}_{N-n}]} f(x)$$

In the inf, change x to $-x$, so $f(x)$ become $f(-x) = \tilde{f}(x)$:

$$L(f,P) = \sum_{n=1}^{N} (-\tilde{x}_{N-n} + \tilde{x}_{N+1-n}) \inf_{-x \in [-\tilde{x}_{N+1-n}, -\tilde{x}_{N-n}]} \tilde{f}(x)$$

Now, $-x \in [-\tilde{x}_{N+1-n}, -\tilde{x}_{N-n}]$ can be replaced by $x \in [\tilde{x}_{N-n}, \tilde{x}_{N+1-n}]$ and $-\tilde{x}_{N-n} + \tilde{x}_{N+1-n}$ can be reordered to $\tilde{x}_{N+1-n} - \tilde{x}_{N-n}$:

$$L(f,P) = \sum_{n=1}^{N} (\tilde{x}_{N+1-n} - \tilde{x}_{N-n}) \inf_{x \in [\tilde{x}_{N-n}, \tilde{x}_{N+1-n}]} \tilde{f}(x)$$

Finally, we can reverse the order of summation by changing $N+1-n$ to n and $N-n$ to $n-1$

$$L(f,P) = \sum_{n=1}^{N} (\tilde{x}_n - \tilde{x}_{n-1}) \inf_{x \in [\tilde{x}_{n-1}, \tilde{x}_n]} \tilde{f}(x)$$

which is indeed $L(\tilde{f}, \tilde{P})$ as required.

This shows that every lower sum of f is also a lower sum of $\tilde{f}$. Noting that $\tilde{\tilde{f}} = f$ and $\tilde{\tilde{P}} = P$, we can also see that every lower sum of $\tilde{f}$ is also a lower sum of f; the two functions have exactly the same lower sums, so must have the same lower integrals: $L_a^b f = L_{-b}^{-a} \tilde{f}$.

The proof of $U(f,P) = U(f,\tilde{P})$ is effectively identical (just change inf to sup throughout), so we can conclude in the same way that $U_a^b f = U_{-b}^{-a} \tilde{f}$.

By definition, a function is integrable if its lower and upper integrals are equal, in which case the integral is their common value; we conclude that f is integrable if and only if $\tilde{f}$ is integrable, in which case $\int_a^b f = \int_{-b}^{-a} \tilde{f}$.

12.3. Suppose

$$P_k(N) = \frac{N^{K+1}}{k+1} + \sum_{j=0}^{k} a_j N^j$$

Then

$$\begin{aligned}
L(f, P_N) &= \frac{1}{N^{k+1}} \sum_{n=1}^{N} (n-1)^k \\
&= \frac{1}{N^{k+1}} P_k(N-1) \\
&= \frac{(N-1)^{k+1}}{N^{k+1} k+1} + \sum_{j=0}^{k} a_j \frac{(N-1)^j}{N^{k+1}} \\
U(f, P_N) &= \frac{1}{N^{k+1}} \sum_{n=1}^{N} n^k \\
&= \frac{1}{N^{k+1}} P_k(N) \\
&= \frac{1}{k+1} + \sum_{j=0}^{k} a_j N^{j-k-1}
\end{aligned}$$

By standard algebra of limits arguments, both $L(f, P_N)$ and $U(f, P_N)$ tend to $1/(k+1)$ as $N \to \infty$. Now proceed as in Example 29.10: we have

$$L(f, P_n) \leq L_0^1 f \leq U_0^1 f \leq U(f, P_N)$$

and let $N \to \infty$ to give

$$\frac{1}{k+1} \leq L_0^1 f \leq U_0^1 f \leq \frac{1}{k+1}$$

We can now see that

$$L_0^1 f = U_0^1 f = \frac{1}{k+1}$$

so f is integrable on $[0, 1]$ and $\int_0^1 f = 1/(k+1)$.

12.4. Following the line sketched in the question, on the one hand we have the telescopic sum formula

$$\sum_{n=1}^{N} [(n+1)^{k+1} - n^{k+1}] = (N+1)^{k+1} - 1$$

On the other hand,

$$\begin{aligned}
\sum_{n=1}^{N}[(n+1)^{k+1}-n^{k+1}] &= \sum_{n=1}^{N}\left[\sum_{j=0}^{k+1}\binom{k+1}{j}n^j\right]-n^{k+1} \\
&= \sum_{n=1}^{N}\sum_{j=0}^{k}\binom{k+1}{j}n^j \\
&= \sum_{j=0}^{k}\sum_{n=1}^{N}\binom{k+1}{j}n^j \\
&= \sum_{j=0}^{k}\binom{k+1}{j}\sum_{n=1}^{N}n^j \\
&= \sum_{j=0}^{k}\binom{k+1}{j}P_j(N)
\end{aligned}$$

Combining these gives

$$\sum_{j=0}^{k}\binom{k+1}{j}P_j(N) = (N+1)^{k+1}-1 \tag{*}$$

Now, we can use (*) to prove by induction that P_k is a polynomial of degree k. For an induction base, we have $P_0(N) = N$. Assume that for some specific k, $P_0,\dots,P_{k-1}$ are polynomials of degree $0,\dots,k-1$. Breaking off the $j=k$ term on the LHS of (*), we have

$$\binom{k+1}{k}P_k(N)+\sum_{j=0}^{k-1}\binom{k+1}{j}P_j(N) = (N+1)^{k+1}-1$$

which we can think of as

$$\binom{k+1}{k}P_k(N)+(\text{poly in } N \text{ of degree} \le k) = (\text{poly in } N \text{ of degree} = k+1)$$

from which it follows that P_k is a polynomial of degree $k+1$. The result now follows by induction.

We can also find the leading coefficient using this decomposition. On the LHS, the only N^{k+1} term is from $\binom{k+1}{k}P_k(N)$; this has coefficient $(k+1)$ times the coefficient of N^{k+1} in $P_k(N)$. On the RHS, the coefficient of N^{k+1} is 1, from the binomial theorem. Combining these, we see that the leading coefficient of P_k is $1/(k+1)$.

12.5. Because f is increasing on $[a,b]$, we have for any $x \in [a,b]$, $f(a) \le f(x) \le f(b)$, so f is bounded.

In the partition P_N, subinterval n ($1 \le n \le N$) is $[a+(n-1)w, a+nw]$, with width w. The lower sum is thus, by definition,

$$L(f,P_N) = \sum_{n=1}^{\infty} w \inf_{x\in[a+(n-1)w,a+nw]} f(x)$$

But f is increasing, so it attains its minimum (hence its infimum) over each subinterval at its left-hand endpoint; that is,

$$U(f,P_N) = \sum_{n=1}^{N} wf(a+(n-1)w)$$

as stated. For the upper sum, we argue similarly, but the supremum over each subinterval is attained at its right-hand endpoint:

$$U(f,P_N) = \sum_{n=1}^{N} wf(a+nw)$$

Comparing these sums, for $1 < n \le N$, term n in the lower sum is equal to term $n-1$ in the upper sum. In the difference $U(f,P_N) - L(f,P_N)$, all terms cancel except the last term in the upper sum and the first term in the lower sum; that is,

$$U(f,P_N) - L(f,P_N) = wf(b) - wf(a) = \frac{(b-a)(f(b)-f(a))}{N}$$

which tends to zero as $N \to \infty$. If follows that for any $h > 0$ there exists $N \in \mathbb{N}$ such that

$$U(f,P_N) - L(f,P_N) < h$$

so f is integrable by the Cauchy criterion for integrability (Theorem 29.11).

If f is decreasing then $-f$ is increasing; the first part of the question then shows that $-f$ is integrable and it follows from Theorem 30.1 that f is integrable.

12.6. Every interval of positive length contains both rational and irrational numbers (Theorem 3.21, Exercise 9.5). It follows that, on any interval of positive length, f takes on the values 0 and 1 (and, from its definition, no others). In particular, the infimum and supremum of f over any interval of positive length are respectively 0 and 1. Now, if $a < b$ and $P = (x_0, \ldots, x_N)$ is a partition of $[a,b]$ then

$$L(f,P) = \sum_{n=1}^{N} (x_n - x_{n-1}) \inf_{x\in[x_{n-1},x_n]} f(x) = 0$$

$$U(f,P) = \sum_{n=1}^{N} (x_n - x_{n-1}) \sup_{x\in[x_{n-1},x_n]} f(x) = \sum_{n=1}^{N} x_n - x_{n-1} = b-a$$

It follows that

$$L_a^b f = 0; \qquad U_a^b f = b-a$$

so, since $a < b$, the upper and lower integrals are not equal and the Dirichlet function is not integrable.

Exercises 13

There is nothing to hand in from these exercises.

13.1. Show that if $f, g : [a,b] \to \mathbb{R}$ are such that f is integrable over $[a,b]$ and $f(x) = g(x)$ for all but finitely many $x \in [a,b]$, then g is integrable over $[a,b]$ and $\int_a^b g = \int_a^b f$. *Don't try to use partitions! Use Exercise 12.1 and linearity.*

⚠ 13.2. Suppose $f : [a,b] \to \mathbb{R}$ is continuous and that $\int_a^b |f| = 0$. Show that $f(x) = 0$ for all $x \in [a,b]$.

13.3. Define $f : (0,1) \to \mathbb{R}$ by $f(x) = 1/x$. Show that f is not uniformly continuous.

13.4. Suppose f is continuous on $[a,b]$, differentiable on (a,b) and we have $|f'(x)| \leq M$ for all $x \in (a,b)$.

(a) Suppose $a \leq x - h/2 < x + h/2 \leq b$. Use the Mean Value Theorem (for derivatives) to show that if $|t| \leq h/2$ then we have

$$-M|t| \leq f(x+t) - f(x) \leq M|t|$$

(b) Deduce that

$$-\frac{Mh^2}{4} \leq \int_{x-h/2}^{x+h/2} f(t)\,\mathrm{d}t - hf(x) \leq \frac{Mh^2}{4}$$

(c) For $N \in \mathbb{N}$, let $h = (b-a)/N$ and for $1 \leq n \leq N$ let $x_n = a + (n - 1/2)h$, so $x_0 = a + h/2$ and $x_N = b - h/2$. Apply the previous result with $x = x_n$ and sum over $n = 1 \ldots N$ to give

$$-\frac{M(b-a)h}{4} \leq \int_a^b f(t)\,\mathrm{d}t - h\sum_{n=1}^{N} f(x_n) \leq \frac{M(b-a)h}{4}$$

(d) What does the Sandwich Theorem tell us, as $N \to \infty$?

This is the midpoint rule for numerical integration, with an error estimate.

⚠ 13.5. Show that if $a < b$ and $f, g : [a,b] \to \mathbb{R}$ are both integrable, then the product function fg (i.e. the function defined by $(fg)(x) = f(x)g(x)$) is also integrable. *Probably the most direct way to do this is to start by proving that*

$$U(fg,P) - L(fg,P) \leq A[U(g,P) - L(g,P)] + B[U(f,P) - L(f,P)]$$

where P is any partition of $[a,b]$ and

$$A = \sup_{x\in[a,b]} |f(x)|; \qquad B = \sup_{x\in[a,b]} |g(x)|$$

13.6. Suppose $f : [a,b] \to \mathbb{R}$ is continuous and $g : [a,b] \to \mathbb{R}$ is non-negative and integrable, with $\int_a^b g > 0$. Given that fg is integrable on $[a,b]$, show that there exists $c \in [a,b]$ such that

$$f(c)\int_a^b g = \int_a^b fg$$

This is a generalised Mean Value Theorem for Integrals: the version in the notes is a special case, with g the constant function 1. *Proving that fg is integrable on $[a,b]$ is the (tricky) content of the previous exercise.*

13.7. Let s be the sign function:

$$s(x) = \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x > 0 \end{cases}$$

Show that if $a, b \in \mathbb{R}$ with $a \le b$ then

$$\int_a^b s = |b| - |a| \tag{*}$$

You'll need to consider three cases, e.g. $a \le b \le 0$, $a \le 0 \le b$ and $0 \le a \le b$. Exercise 12.1 and Theorem 30.3 are relevant.

We know (example in the proof of Theorem 21.4) that the derivative of the absolute value function is, except at the origin, the sign function. This makes (*) look very like the (second) Fundamental Theorem of Calculus (Theorem 31.7). Why is (*) not, in fact, an example of the FTC in action?

⚠ 13.8. Define $f : [-1,1] \to \mathbb{R}$ by

$$f(x) = \begin{cases} \sin(1/x) & \text{if } x \ne 0 \\ 0 & \text{if } x = 0 \end{cases}$$

Show that f is integrable and find $\int_{-1}^1 f$. *Use any properties of* sin *established anywhere in the course: you need surprisingly few!*

⚠ Harder.

Exercises 13: Suggested Solutions

13.1. Let $x_1, \dots, x_N$ be the points at which $g(x) \neq f(x)$. Define d_j $(1 \leq j \leq N)$ by

$$d_j(x) = \begin{cases} 1 & \text{if } x = x_j \\ 0 & \text{if } x \neq x_j \end{cases}$$

Then we can write

$$g(x) = f(x) + \sum_{j=1}^{N} (g(x_j) - f(x_j)) d_j(x)$$

for all $x \in [a,b]$. We know that f is integrable (by hypothesis) and all the d_j are integrable (by Exercise 12.1), so this linear combination is integrable. Finally, all the d_j have integral zero, so

$$\int_a^b g = \int_a^b f + \sum_{j=1}^{N} (g(x_j) - f(x_j)) \int_a^b d_j = \int_a^b f$$

13.2. We show the contrapositive: that if $f(x_0) \neq 0$ for some $x_0 \in [a,b]$ then $\int_a^b |f| > 0$. To do this, use continuity of f at x_0 with $\varepsilon = |f(x_0)|/2$ to find $\delta > 0$ such that if $x \in [a,b]$ and $|x - x_0| < \delta$ then $|f(x) - f(x_0)| < |f(x_0)|/2$. This last inequality can be written as $-|f(x_0)|/2 < f(x) - f(x_0) < |f(x_0)|/2$, and the LH of these two inequalities tells us that, if $x \in [a,b]$ and $|x - x_0| < \delta$, then $|f(x)| > |f(x_0)|/2$. Now, the constraint "$x \in [a,b]$ and $|x - x_0| < \delta$" can be written as $x \in [a,b] \cap (x_0 - \delta, x_0 + \delta) =: I$, and I is an interval of positive length: its left-hand endpoint is $\max\{a, x_0 - \delta\}$ and its right-hand endpoint is $\min\{b, x_0 + \delta\}$ (think for a moment about why this is true, including why $\max\{a, x_0 - \delta\} < \min\{b, x_0 + \delta\}$; it's not hard, but there are lots of individual cases to consider). If we define a new function $g : [a,b] \to \mathbb{R}$ by

$$g(x) = \begin{cases} |f(x_0)|/2 & \text{if } x \in I \\ 0 & \text{if } x \notin I \end{cases}$$

then $g(x) \leq |f(x)|$ for all $x \in [a,b]$. We can integrate this inequality to give $\int_a^b |f| \geq L|f(x_0)|/2$, where L is the length of the interval I; this is the required inequality.

13.3. For $\delta \in (0,1)$, consider

$$f(\delta/2) - f(\delta) = \frac{1}{\delta} \to \infty$$

as $\delta \to 0$. We have $|\delta/2 - \delta| < \delta$ but there is no bound for $|f(\delta/2) - f(\delta)|$, so no relationship of the form $|f(x) - f(y)| < \varepsilon$ if $|x - y| < \delta$ can hold.

13.4. We are given that f is continuous on $[a,b]$ and differentiable on (a,b) and we have $|f'(x)| \leq M$ for all $x \in (a,b)$.

(a) The MVT tells us that

$$\left|\frac{f(x+t)-f(x)}{t}\right| = |f'(\tau)|$$

for some τ between $x+t$ and x; because of inequality given, $|f'(\tau)| \leq M$ and we can rearrange to give

$$|f(x+t)-f(x)| \leq M|t|$$

which breaks up into two inequalities

$$-M|t| \leq f(x+t)-f(x) \leq M|t|$$

as required.

(b) We can integrate this from $t = x-h/2$ to $t = x+h/2$ using the FTC to give

$$-\frac{Mh^2}{4} \leq \int_{x-h/2}^{x+h/2} f(t)\,\mathrm{d}t - hf(x) \leq \frac{Mh^2}{4};$$

ignoring the constants, the essence of the calculation is

$$\int_{-h/2}^{h/2} |t|\,\mathrm{d}t = 2\int_0^{h/2} t\,\mathrm{d}t = 2\left[\frac{t^2}{2}\right]_{t=0}^{h/2} = \frac{h^2}{4}$$

(c) Replacing x by x_n and summing over n gives

$$-\sum_{n=1}^{N}\frac{Mh^2}{4} \leq \sum_{n=1}^{N}\int_{x_n-h/2}^{x_n+h/2} f(t)\,\mathrm{d}t - \sum_{n=1}^{N} hf(x_n) \leq \sum_{n=1}^{N}\frac{Mh^2}{4}$$

Now, the two outer sums are sums of N copies of the constant $Mh^2/4$, so the sum if $NMh^2/4$. But $h = (b-a)/N$, so this can also be written as $M(b-a)h/4$. We now have

$$-\frac{M(b-a)h}{4} \leq \sum_{n=1}^{N}\int_{x_n-h/2}^{x_n+h/2} f(t)\,\mathrm{d}t - \sum_{n=1}^{N} hf(x_n) \leq \frac{M(b-a)h}{4}$$

The intervals $[x_n-h/2, x_n+h/2]$ are non-overlapping and have union $[a,b]$, so the sum of the integral of f over $[x_n-h/2, x_n+h/2]$ is the integral of f over $[a,b]$. That is,

$$-\frac{M(b-a)h}{4} \leq \int_a^b f(t)\,\mathrm{d}t - \sum_{n=1}^{N} hf(x_n) \leq \frac{M(b-a)h}{4}$$

We now need only pull a factor of h out of the remaining sum to get to the answer claimed.

(d) As $N \to \infty$, $h = (b-a)/N \to 0$; the Sandwich Theorem gives

$$h \sum_{n=1}^{N} f(x_n) \to \int_a^b f(t)\,\mathrm{d}t$$

as $N \to \infty$.

13.5. We have no explicit value for $\int_a^b fg$, so the only device we have is the Cauchy condition for integrability. Like we did for early, bare-handed, proofs about convergent sequences, we start with an inequality. The general idea is that we can make $f(x)-f(y)$ and $g(x)-g(y)$ small for all x, y in each subinterval of a suitable partition; we need to make $f(x)g(x)-g(y)g(y)$ small. The result makes frequent use of the fact that

$$\sup_{x,y\in S} |\phi(x)-\phi(y)| = \sup_{x,y\in S} [\phi(x)-\phi(y)] = \sup_{x\in S} \phi(x) - \inf_{x\in S} \phi(x)$$

which is easily checked. Like in the proof of the algebra of limits, we write

$$f(x)g(x)-f(y)g(y) = f(x)[g(x)-g(y)] + [f(x)-f(y)]g(y) \qquad (*)$$

so by the triangle inequality

$$|f(x)g(x)-f(y)g(y)| \le |f(x)||g(x)-g(y)| + |f(x)-f(y)||g(y)|$$

This gives us some idea of how to make the bounds: we need to control all four terms on the RHS. To develop this idea, suppose $P = (x_0, \dots, x_N)$ is a partition of $[a,b]$ and let

$$h_n = \sup_{x\in[x_{n-1},x_n]} f(x) - \inf_{x\in[x_{n-1},x_n]} f(x) = \sup_{x,y\in[x_{n-1},x_n]} |f(x)-f(y)|$$
$$k_n = \sup_{x\in[x_{n-1},x_n]} g(x) - \inf_{x\in[x_{n-1},x_n]} g(x) = \sup_{x,y\in[x_{n-1},x_n]} |g(x)-g(y)|$$

and, as in the question,

$$A = \sup_{x\in[a,b]} |f(x)|; \qquad B = \sup_{x\in[a,b]} |g(x)|$$

By (*) above, if $x, y \in [x_{n-1}, x_n]$ then we have

$$|f(x)g(x)-f(y)g(y)| \le Ak_n + Bh_n$$

Taking the supremum over all $x, y \in [x_{n-1}, x_n]$, we have

$$\sup_{x,y\in[x_{n-1},x_n]} |(fg)(x)-(fg)(y)| \le Ak_n + Bh_n$$

or equivalently

$$\sup_{x\in[x_{n-1},x_n]} (fg)(x) - \inf_{x\in[x_{n-1},x_n]} (fg)(x) \le Ak_n + Bh_n$$

Now, we can turn this into an inequality about upper and lower sums by multiplying by $x_n - x_{n-1}$ and summing over n:

$$U(fg,P) - L(fg,P) \leq A[U(g,P) - L(g,P)] + B[U(f,P) - L(f,P)] \qquad (\dagger)$$

This was all true for any bounded f and g and any partition P. Now we come to integrability. Assuming f and g are integrable on $[a,b]$ and $\varepsilon > 0$, and provided $A > 0$ and $B > 0$, there exist partitions P_f and P_g of $[a,b]$ such that

$$U(f,P_f) - L(f,P_f) < \frac{\varepsilon}{2B}; \qquad U(g,P_g) - L(g,P_g) < \frac{\varepsilon}{2A}$$

If we let P be a common refinement of P_f and P_g then these inequalities remain true with P_f and P_g replaced by P. If then follows from (†) that $U(fg,P) - L(fg,P) < \varepsilon$, showing that fg is indeed integrable. If $A = 0$ or $B = 0$ then the above doesn't quiet work, but in these case fg is identically zero, so trivially integrable. Or the above argument works with any $A > \sup_{x\in[a,b]} |f(x)|$ and $B > \sup_{x\in[a,b]} |g(x)|$.

13.6. Let m and M be the minimal and maximal values taken on by f over $[a,b]$ (Extreme Value Theorem), so for all $x \in [a,b]$ we have

$$m \leq f(x) \leq M$$

Multiply by $g(x) \geq 0$:

$$mg(x) \leq f(x)g(x) \leq Mg(x)$$

and integrate from a to b:

$$m\int_a^b g \leq \int_a^b fg \leq M\int_a^b g$$

Divide by $\int_a^b g$ (we are told that this is not zero):

$$m \leq \frac{\int_a^b fg}{\int_a^b g} \leq M$$

Now, the Intermediate Value Theorem tells us that f takes on every value between m and M, so we have for some $c \in [a,b]$

$$f(c) = \frac{\int_a^b fg}{\int_a^b g}$$

which is the required result.

13.7. As suggested, we consider three cases. These cases overlap very slightly, which is fine as long as every posibility is considered. Some variation in the exact pattern of $<$ and $\leq$ is possible.

$a \le b \le 0$ In this case, we can proceed as if s is constant at -1: this is literally true if $b < 0$, and if $b = 0$ we can apply Exercise 12.1 to see that the different value at a single point (0) makes no difference. Bearing in mind the sign pattern ($a \le 0$ and $b \le 0$), the integral is

$$\int_a^b s = \int_a^b (-1) = (-1)(b-a) = -b + a = |b| - |a|$$

$a \le 0 \le b$ Here we break up the integral (Theorem 30.3) as

$$\int_a^b s = \int_a^0 s + \int_0^b s$$

Now, again applying Exercise 12.1, we can proceed as if s is constant at -1 on $[a,0]$ and $+1$ on $[0,b]$, giving (using $a \le 0$ and $b \ge 0$)

$$\int_a^b s = (-1)(0-a) + (1)(b-0) = a + b = |b| - |a|$$

$0 \le a \le b$ Exactly as in the first part, we proceed as if s is constant at $+1$ on $[a,b]$, giving

$$\int_a^b s = \int_a^b (+1) = (+1)(b-a) = b - a = |b| - |a|$$

The required formula has now been proved in every possible case.

The reason this is not a consequence of the FTC as stated in Theorem 31.7 is that the derivative of the absolute value function is not everywhere equal to the sign function: at the origin, the absolute value has no derivative. This is enough to stop us applying the theorem.

The fact that the FTC appears to work for an example outside its stated scope is a hint towards the fact that there are much more powerful versions of the FTC available, with the same conclusions but weaker hypotheses.

13.8. We first show that f is integrable, using the Cauchy criterion for integrability (Theorem 29.11). Suppose $h > 0$. We consider three different regions: one containing the origin (containing the discontinuous behaviour), one left of the origin and one right of the origin.

Consider some $\xi \in (0,1)$. Over $[-\xi,\xi]$, the supremum of f is is no more than 1 and the infimum is no more than -1 (in fact they are equal to 1 and -1, but we don't need that much precision). Thinking of $[-\xi,\xi]$ as s subinterval of a partition, the contribution to the upper sum is no more than 2ξ and to the lower sum no less than -2ξ. The contribution to the difference between upper and lower sums is no more than 4ξ. Let ξ be such that $4\xi = h/3$, i.e. $\xi = h/12$.

Now, on $[\xi,1]$, f is continuous, hence integrable. There is therefore a partition P^+ of of $[\xi,1]$ such that $U(f,P^+) - L(f,P^+) < h/3$. Similarly, there is a partition P^- of $[-1,-\xi]$ such that $U(f,P^-) - L(f,P^-) < h/3$.

Let P be the partition of $[-1,1]$ containing all the points of P^- and P^+; because P^- ends at $-\xi$ and P^+ starts at ξ, P has a subinterval $[-\xi,\xi]$. As in the proof of the additivity property (Theorem 30.3), but with three smaller partitions instead of two, we can write $U(f,P)-L(f,P)$ as the sum of three "upper minus lower" terms, one from P^-, one from $[-1,-\xi]$ and the third from P^+. The estimates above give

$$U(f,P) \leq U(f,P^-) - L(f,P^-) + 4\xi + U(f,P^+) - L(f,P^-) < \frac{h}{3} + \frac{h}{3} + \frac{h}{3} = h$$

This shows that f is integrable.

This leaves the question of the value of $\int_{-1}^{1} f$. There is no formula for an antiderivative here, but we can use the fact that f is an odd function, so its integral over the symmetric integrable $[-1,1]$ is zero. To fully justify this, we observe that, on the one hand,

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

using oddness and linearity and, on the other hand,

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

by Exercise 12.2. Combining these gives

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

so we must have $\int_{-1}^{1} f(x)\,\mathrm{d}x = 0$.