Contents

2C Intro to real analysis 2020/21

HANDOUT

Introduction

This course is our first introduction to *mathematical analysis*. Our focus in this course, and the 3H course "Analysis of differentiation and integration" that follows it, is the rigorous study of limits and limiting processes.

You have seen limits before – for example in first year, when the derivative of a function f at x was defined by

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

However, back then not too much attention was paid to what the symbol $\lim_{h\to 0}$ means. Essentially, the first year courses take an intuitive view of limits: the expression $\lim_{h\to 0} g(h)$ means "the value that $g(h)$ gets close to when h gets close to 0". Then one computes the derivative of functions like $f(x) = x^2$ via calculations like

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

In this course we take a more sophisticated point of view: we will try to really understand how our intuitive notions can be turned into precise definitions, and then examine the basics of calculus again. Our aim in mathematical analysis is no longer to be able to find explicit formulae for the derivatives and integrals of various functions, but instead to understand rigorously what the processes of integration and differentiation are, and to use this to obtain qualitative information about functions.

Why do we bother?

It is easy to believe that a number of the theorems we will discuss (such as the intermediate value theorem at the end of the course) are "obviously true". Moreover, the proofs of these statements can be involved and tricky, leading to people asking "why do we spend so much time carefully proving the obvious?".

Let us take a look at an example of how an innocent looking "obviously true" argument can lead to strange conclusions.

Example Taking $x = 1$ in the McLaurin expansion

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots,$$

(which is valid for $-1 < x \leq 1$), we have

$$\log(2) = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \frac{1}{7} - \frac{1}{8} + \cdots.$$

Now let us rearrange the series on the right hand side above, by alternating one positive term followed by two negative terms. This rearrangement is

$$\begin{aligned}
&1-\frac{1}{2}-\frac{1}{4}+\frac{1}{3}-\frac{1}{6}-\frac{1}{8}+\frac{1}{5}-\frac{1}{10}-\frac{1}{12}+\cdots \\
&=\left(1-\frac{1}{2}\right)-\frac{1}{4}+\left(\frac{1}{3}-\frac{1}{6}\right)-\frac{1}{8}+\left(\frac{1}{5}-\frac{1}{10}\right)-\frac{1}{12}+\cdots \\
&=\frac{1}{2}-\frac{1}{4}+\frac{1}{6}-\frac{1}{8}+\frac{1}{10}-\frac{1}{12}+\cdots \\
&=\frac{1}{2}\log(2).
\end{aligned}$$

Since $\log(2) \neq \frac{1}{2}\log(2)$, we conclude that the two rearranged series are not equal.

When I first saw this, it certainly surprised me: at the time I'd assumed that it was obvious that the addition could be performed in any order we liked. I had never really thought about what it meant to add up infinitely many real numbers, or whether this made sense, and instead just assumed everything would be fine.

In this course, we'll put mathematical analysis on a sound footing, making precise definitions of what is meant by an expression of the form

$$\log(2) = 1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\frac{1}{5}-\frac{1}{6}+\frac{1}{7}-\frac{1}{8}+\cdots.$$

By making all these notions precise we can carefully explore how they behave: we'll do this by stating and proving theorems describing what operations can be legitimately performed. This way we can be sure that the manipulations we perform are correct; moreover in doing this, we discover places where our intuition isn't quite right, which for me is the most interesting part of the study of real analysis (well explained examples of this sort can be found in the first and last section of each chapter of [Ab])

An equally important reason for this course is to continue the themes of 2B and 2F in developing skills in handling rigourous mathematical argument. A key aim is for you to be able to write your own proofs of statements in analysis, write up solutions to exercises in the form of an argument[1], and be able to decide whether a given mathematical argument is correct or not (there will be questions of this type on the webassign — various answers to the exercise questions will be given, and you will be asked to identify the correct arguments). These are not easy skills and take a lot of practice to master[2], but once mastered these skills will be vital in a range of future courses. Moreover, rigorous and precise thinking is one of the key skills sought after by employers that successful mathematics graduates have.

[1] not in the form of some notes that the person reading your solution can use to piece together an argument from.

[2] I'd recommend reading the sections of [A] and [H] on mathematical writing.

References

There are a wide range of books covering the material on this course (try searching for words like "introduction to mathematical analysis").

Analysis Books

[Ab] Stephen Abbot, Understanding Analysis, Undergraduate texts in mathematics, 2000.

This is a relatively short and nice treatment of the material in both this, and the 3H course "Analysis of differentiation and integration", which complements the course well. [Ab] has relatively few worked examples, but more explanation of why definitions are made the way they are. We'll cover material from chapters 1, 2 and 4 of [Ab].

[Al] Lara Alcock, How to think about analysis, Oxford University Press, 2014.

This book complements nicely the material covered in this course and the 3H course "Analysis of differentiation and integration". The first 7 chapters are particularly relevant. As the title suggests, [Al] provides a lot of useful tips and tricks how to organise your thinking about analysis.

[ERA] Colin McGregor, Elementary Real Analysis, 2002. Available via Moodle.

This is a complete set of lecture notes for two 25 hour lecture courses (2U and 2V) which preceded the current course. We will roughly cover the first half of [ERA] in this course. [ERA] provides a number of methods and worked examples, but is sometimes short of explanations.

[Ru] Walter Rudin, Principles of Mathematical Analysis (3rd Edition), McGraw-Hill, 1976

I'd recommend this classic book to students who enjoy the material in this course, and would like to see a more advanced treatment. The book [Ru] starts with a detailed construction of the real numbers showing one can construct a set satisfying the axioms described in Chapter 2 of our course. The book continues in this style, with many excellent, and often quite challenging exercises.

[Sp] Michael Spivak, Calculus (6th Edition), Cambridge University Press, 2006.

This is a classic text covering the material of this course, and the following 3H course "Analysis of differentiation and integration." However, it does so in a different order from these courses, and also in my view it's a bit over long. Having said that, it's a nice book which provides a large amount of extra exercises.

[K] Tom Körner, A Companion to Analysis, Grad. Studies in Math. 62, AMS, 2004.

Everything you wanted to know about analysis but were afraid to ask.

General study guides

[A] Lara Alcock, How to study for a mathematics degree, Oxford University Press, 2013.

I strongly recommend that everyone planning a maths degree buy and read[3] this book. It's a guide for how to study mathematics which should stand you in good stead for this course, and all the future courses you plan to take; I certainly learnt quite a lot about the implicit assumptions made in teaching mathematics at University.

[3] in my experience too many mathematics books are bought and not read (including far too many of the books on my bookshelf)! In this case, there's very little technical maths in [A], so while it's thought provoking, it's also relatively easy reading.

[H] Kevin Houston, How to think like a mathematician, Cambridge University Press, 2009.

This is another useful book, which is well worth reading. It's more mathematical than the previous book, and I'd particular recommend parts 2 and 3 on logic, definitions, theorems and proofs.

Chapter 1: Logic and Inequalities

This chapter is a brief background on two key topics which will appear throughout the course: mathematical logic and inequalities.[4]

[4] Both of these topics are major reasons why some students find mathematical analysis difficult — and regularly the source of lost marks in the exam. To succeed in this course, it's important to be able to confidently manipulate inequalities, and logical expressions, and to be able to write clearly in a logical fashion.

Logic and Writing

Our aim is to understand the structure of written mathematics, so that you can write your own mathematical arguments in a logically correct fashion. I recommend reading [H] Chapters 6-13 alongside this section. Some material can also be found in Sections 0.1 and 0.2 of [ERA].

Statements

The building blocks of written mathematics are statements, by which we mean the following.

Definition 1.1. *A* statement *is a sentence which is either true or false, but not both. Statements may contain* free variables, *for example x, where the truth of the statement depends on the value of the variable.*

Examples

- "$2 + 7 = 9$" is a statement, which is true.
- "Dogs can fly" is a statement, this one happens to be false.[5]
- "Turn on the lights" is not a statement from our point of view (it is a request or command).
- "There are infinitely many prime numbers" is a true statement.
- "There are infinitely many pairs of prime numbers with difference 2" is also a statement — it is either true or false, though at the time of writing we do not know which.[6]
- "$x^2 + 2x + 1 = 0$" is a statement, with *free variable* x; the truth of the statement depends on the value of the variable (it's true if and only if $x = -1$).
- "There is some $x \in \mathbb{R}$ with $x^2 + 2x + 1 = 0$" is also a statement. This time there are no free variables, and the statement is true (take $x = -1$). This last example of constructing a new statement by *quantifying* over a free variable of another statement is an important concept.

[5] Context is always important. In one sense I am no more able to fly than a dog but you *know what I mean* when I say that I "fly down to Stansted". Wittgenstein says: "A language is a way of life."

[6] This is the twin prime conjecture, which having been an open question for centuries has recently seen much progress. In 2013 it was proved that there exists a constant $N \in \mathbb{N}$ such that there are infinitely many pairs of primes whose difference is at most N. The first bound on this constant was $N \approx 70$ million, which subsequently could be reduced significantly.

We can combine and negate statements using the *logical connectives* "and", "or" and "not". In what follows we often use symbols like P and Q to denote abstract statements, and $P(x)$ to denote an abstract statement depending on the free variable x.

Definition 1.2. *Let P and Q be statements.*

- *The negation of P is the statement that is true when P is false and false when P is true. We write* not(P) *for the negation of P.*
- *The statements $(P$ and $Q)$ and $(P$ or $Q)$*[7] *are defined by the truth tables*

P	Q	$(P \text{ and } Q)$	$(P \text{ or } Q)$
T	T	T	T
T	F	F	T
F	T	F	T
F	F	F	F

Note how to negate joined statements:

- "not(P or Q)" is equivalent to "not(P) and not(Q)"[8].
- "not(P and Q)" is equivalent to "not(P) or not (Q)".

[7] In mathematics we *always* use the inclusive or: so that "P or Q" is defined to be true when both P and Q are true (as well as when P is true and Q is false, and when Q is true and P is false). This isn't always the case in every day English — if you're asked do you want a baked potato or chips with your steak, it's understood that both is not an allowed option. Be careful when trying to check logic with examples from every day speech.

[8] that is, these two statements have the same truth table (which you should check). We'll discuss equivalence of statements again later.

Implications

Implications are a key logical connective. Again, we combine two statements to produce a new statement.

Definition 1.3. *Let P and Q be statements. The statement $P \implies Q$*[9] *has truth table*

P	Q	$P \implies Q$
T	T	T
T	F	F
F	T	T
F	F	T

The statement $P \implies Q$ can be written in English in a number of ways: "P implies Q", "If P is true, then Q is true", "if P, then Q"[10], "P only if Q" (i.e. "P is true only if Q is true.")[11], "Q whenever P" (i.e. "Q is true whenever P is true"), amongst others.

> Note that "$P \implies Q$" is true whenever P is false.

For example "If pigs can fly, then there is a man on the moon" is a true statement as is "If pigs can fly, then there is not a man on the moon".

The crucial point about the truth table for $\implies$ is simply that it disallows us to deduce a falsehood from a truth.

In the implication $P \implies Q$ we call P the *hypothesis* of the implication and Q the *conclusion*. Note that pretty much every mathematical theorem can be stated in the form of an implication, though sometimes the hypotheses are not explicitly stated. For example "There are no real solutions to $x^2 + 2x + 1 = -1$" can be written as "$x \in \mathbb{R} \implies x^2 + 2x + 1 \neq -1$".

[9] Note that the symbol $\implies$ is reserved for this meaning. It should only ever be used to connect two mathematical statements, the first of which implies the second. Any other use of $\implies$ is at best a serious mathematical grammar error. In particular, I've noted in homeworks and exams the symbols $=, \rightarrow, \implies$ being used interchangeably: each of these has it's own meaning, and needs to be used correctly.

[10] This is essentially shorthand for "If P is true, then Q is true"

[11] It is harder to convince yourself that this is another logically equivalent way of stating the implication. See the discussion on page 67 of [H].

Definition 1.4. *Let P and Q be statements and consider the implication $P \implies Q$.*

- *The* converse *of $P \implies Q$ is the implication $Q \implies P$.*
- *The* contrapositive *of $P \implies Q$ is the implication (not Q) $\implies$ (not P).*

The implication $P \implies Q$ is equivalent to its contrapositive (not Q) $\implies$ (not P)[12]. If we want to prove these equivalent statements, we can prove which ever is easiest. It's critical not to get an implication $P \implies Q$ confused with the converse $Q \implies P$. These two statements are not in general equivalent[13]. This is another point where precise mathematical argument often differs from every day speech. Consider the implication "If you don't tidy your room, then you can't go out." From our precise mathematical view point, this statement says nothing about what happens if you do tidy your room, whereas the unspoken assumption is that you can go out if you do tidy your room. In mathematics we must write down what we mean, not leave it to be inferred: the correct statement here is "you can go out if and only if you tidy your room". Always check to make you use an implication in the direction you state it!

[12] To justify this claim, work out the truth table for the contrapositive.

[13] Compare the two truth tables

Negating implications

Looking at the truth table

P	Q	$P \implies Q$
T	T	T
T	F	F
F	T	T
F	F	T

we can see that the implication $P \implies Q$ is false in just the second row of the table, that is when P is true and Q is false. Therefore

"not($P \implies Q$)" is equivalent to "P and not(Q)".

This is another source of errors. In more involved questions I often see the negation of "$P \implies Q$" written as "not (P) $\implies Q$" or "$P \implies$ not (Q). Be very careful when negating an implication to make sure what you've written is genuinely the negation. In particular, if you've ended up with another implication then you've probably made a logical mistake — remember there is only one way $P \implies Q$ can be false: when P is true **and** Q is not.

Equivalent statements

Definition 1.5. *Let P and Q be statements (with the same free variables). Then P and Q are* equivalent *precisely when they have the same truth values. We write $P \iff Q$ in this case.*

Note that $P \iff Q$ is equivalent to $((P \implies Q)$ and $(Q \implies P))$. This can be seen using the truth table, which shows that they have the same truth values.

P	Q	$P \Rightarrow Q$	$Q \Rightarrow P$	$(P \Rightarrow Q)$ and $(Q \Rightarrow P)$	$P \Leftrightarrow Q$
T	T	T	T	T	T
T	F	F	T	F	F
F	T	T	F	F	F
F	F	T	T	T	T

When you want to prove an equivalence $P \iff Q$, you can either provide separate arguments for $P \implies Q$ and $Q \implies P$ or look for an argument proving the equivalence $P \iff Q$ directly.

We have already seen a number of examples of equivalences:

- The implication $P \implies Q$ is equivalent to its contrapositive "not $(Q) \implies$ not (P)".
- The implication $P \implies Q$ is equivalent to the statement "((P and Q) or not(P))"[14].

[14] Check this claim by writing down a truth table.

How many logical connectives on two statements are there? Since we can choose the value of the connective in only two ways but there are four possible arguments, there must be $2^4 = 16$ such functions. We can write out all the possibilities and assign them to the known connectives. Thus (using a bar to denote negation)

P	Q	T	or	$\Longleftarrow$	$\Longrightarrow$	$\overline{and}$	P	Q	$\Longleftrightarrow$	$\overline{\Longleftrightarrow}$	$\bar{Q}$	$\bar{P}$	$\bar{or}$	$\overline{\Longleftarrow}$	$\overline{\Longrightarrow}$	and	F
T	T	T	T	T	T	F	T	T	T	F	F	F	F	F	F	T	F
T	F	T	T	T	F	T	T	F	F	T	T	F	F	F	T	F	F
F	T	T	T	F	T	T	F	T	F	T	F	T	F	T	F	F	F
F	F	T	F	T	T	T	F	F	T	F	T	T	T	F	F	F	F

Note that the list includes a "terminal" function which always returns the value T (and the negation of this function) and projections P and Q onto the first and second arguments (and their negations).

Quantification

We often consider statements containing free variable(s), such as "$x^2 + 2x + 1 = 0$" whose truth depends on the value of the variable x. There are two fundamental ways of *quantifying* the free variable x so as to obtain a statement with no free variables: we can ask for $P(x)$ to be true for every possible value of x, or we can ask for $P(x)$ to be true for at least one value of x. For example $x^2 + 2x + 1 = 0$ is not true for every real value of x, but there is a solution to the equation $x^2 + 2x + 1 = 0$, that is, there exists at least one value of x for which this statement is true.

Definition 1.6. *Let $P(x)$ be a statement conditional on the variable x. We form two quantified statements:*

- *"For all x, $P(x)$ is true", written symbolically "$\forall x, P(x)$". We can express this in a number of ways, for example "For every $x, P(x)$ (is true)", "For each $x, P(x)$".*
- *"There exists x such that $P(x)$ is true", written symbolically "$\exists x$ s.t. $P(x)$".*

$\forall$ is a called *universal quantifier* and $\exists$ is called *existential quantifier*.

When working with universal or existential quantifiers, we should be clear what the range of possible x we are considering is. For example "for all x, $x^2 \neq -1$" is true when we consider the variable x to be ranging over the real numbers and false if we consider x to be ranging over the complex numbers. We make this more precise, by writing quantified statements like "$\forall x \in X, P(x)$" for the statement "for all $x \in X, P(x)$", and similarly for existential quantifiers[15]. Here X is the set of possible values of the variable x. The statement "$\exists x \in \mathbb{Q}$ s.t. $x^2 = 2$" is unambiguous (and false), whereas in $\exists x$ s.t. $x^2 = 2$ it is not clear from the context what are the allowed values of x. We should always specify the range of a quantification unless it is completely clear from the context.

[15] Recall the membership symbol $\in$. Make sure you only use it as allowed in that course; $a \in A$ means that a is a member of the set A.

> In this course, we shall take the convention that $x > 0$ means that x is a real number and $x > 0$, and write quantified statements like "$\forall x > 0, P(x)$" to mean "For all strictly positive real numbers $x, P(x)$ is true".

You're free to relabel quantified variables in statements: "$\forall x, P(x)$" is equivalent to "$\forall y, P(y)$" as both claim that the statement $P(\cdot)$ is true no matter what the value of the variable is.

This is often particularly important when working with multiple statements. Suppose we know that $\exists N \in \mathbb{N}$ s.t. $P(N)$ and $\exists N \in \mathbb{N}$ s.t. $Q(N)$ are both true. This does not mean that there is an individual $N \in \mathbb{N}$ such that $P(N)$ and $Q(N)$ are both true[16]. If we want to use our hypotheses we must introduce two different symbols, and take $N_1 \in \mathbb{N}$ such that $P(N_1)$ is true and $N_2 \in \mathbb{N}$ such that $Q(N_2)$ is true, we can't just use N for both[17].

[16] For example, let $P(n)$ be "n is even" and $Q(n)$ be "n is odd."

[17] I'll give concrete examples of errors caused by this later.

We can combine quantifiers to form more complicated statements. For example if $S \subseteq \mathbb{R}$, the statement that S is bounded above[18] is

$$\exists M \in \mathbb{R} \text{ s.t. } \forall x \in S, x \leq M. \tag{1}$$

[18] We'll make this the definition that S is bounded above in the next chapter.

The order of the quantifiers matters: the statement above has been built up as follows:

- starting with the statement "$x \leq M$" in two free (real) variables x and M;
- we quantify over $x \in S$ to form "for all $x \in S$, $x \leq M$" a statement in one free variable M;
- finally we quantify over $M \in \mathbb{R}$ to obtain $\exists M \in \mathbb{R}$ s.t. $\forall x \in S, x \leq M$.

Decomposing statements in this fashion can help in understanding what is going on. For instance, "for all $x \in S, x \leq M$" is true when M is greater than or equal to every element in S. Thus the full statement "$\exists M \in \mathbb{R}$ s.t. $\forall x \in S, x \leq M$" is true when there is a real number M which is greater than or equal to every element in S. When $S = [1, 3]$

the statement is true: we can take $M = 3$, or $M = \pi$, $M = 4, \dots$, whereas if $S = \mathbb{N}$ the statement is false — there is no real number greater than or equal to every natural number[19]. If you find yourself confused by a long quantified statement, try to understand it piece by piece, *starting at the right*.

[19] This rather obvious looking statement is actually not as obvious as it looks. It is known as Archimedes axiom, and we'll come back to this in the next chapter.

Note that the order we perform quantification matters[20]. Continuing with the previous example, the statement

[20] You can interchange the order of quantifiers of the same type: $\forall x, \forall y, P(x,y)$ and $\forall y, \forall x P(x,y)$ are equivalent; similarly $\exists x$ s.t. $\exists y$ s.t. $P(x,y)$ and $\exists y$ s.t. $\exists x$ s.t. $P(x,y)$ are equivalent. We write these statements as $\forall x, y, P(x,y)$ and $\exists x, y$ s.t. $P(x,y)$ respectively.

$$\forall x \in S, \exists M \in \mathbb{R}, x \leq M \tag{2}$$

is not always equivalent to "$\exists M \in \mathbb{R}$ s.t. $\forall x \in S, x \leq M$." To see this think about what the statement (2) means. For any real value x, the statement "$\exists M \in \mathbb{R}, x \leq M$" is true — the point is that M is allowed to depend on x, so we can take $M = x + 1$. Thus we have proved that "$\forall x \in \mathbb{R}, \exists M \in \mathbb{R}$ s.t. $x \leq M$". In particular (2) will be true no matter what the subset $S \subseteq \mathbb{R}$ is. The key difference between (1) and (2) is that in (2) the value of M is allowed to depend on x as M is specified after x in the statement; whereas in (1) the value of M must be specified first, and for the statement to be true, this value of M must be greater than or equal to every element of S. You may find it helpful to record the possible dependence of M on x in (2) explicitly and write this statement as

$$\forall x \in S, \exists M_x \in \mathbb{R}, x \leq M_x.$$

Negating Quantified Statements

The statement $\forall x \in X, P(x)$ is defined to be true precisely when $P(x)$ is true for every $x \in X$. Therefore the statement $\forall x \in X, P(x)$ is false precisely when there is at least one $x \in X$ such that $P(x)$ is false[21]. This gives

[21] Note that we do not change the range of quantification. An error I've seen many times is to write "$\exists \varepsilon \leq 0$ s.t. not $P(\varepsilon)$" for the negation of "$\forall \varepsilon > 0, P(\varepsilon)$". The range of quantification of $\forall \varepsilon > 0, P(\varepsilon)$ is all strictly positive real numbers. This range does not change: if it is not true that for every strictly positive ε, $P(\varepsilon)$ is true, then there must be at least one strictly positive ε such that $P(\varepsilon)$ is not true.

$$\boxed{\text{not}(\forall x \in X, P(x)) \text{ is equivalent to } \exists x \in X \text{ s.t. not}(P(x)).} \tag{3}$$

Similarly,

$$\boxed{\text{not}(\exists x \in X \text{ s.t. } P(x)) \text{ is equivalent to } \forall x \in X, \text{not}(P(x)).} \tag{4}$$

We negate nested quantified statements by applying these two procedures in turn.

Example To find the negation of (1), we proceed by negating the quantifiers in turn. The statement (1) is "$\exists M \in \mathbb{R}$ s.t. $\forall x \in S, x \leq M$" and so can be written "$\exists M \in \mathbb{R}$ s.t. $P(M)$", where $P(M)$ is the statement "$\forall x \in S, x \leq M$". The first equivalence below obtained by applying (4) to "not($\exists M \in \mathbb{R}$ s.t. $P(M)$", and the second by applying (3) to find an expression for "not $P(M)$". We have[22]

$$\begin{aligned}
&\text{not}(\exists M \in \mathbb{R} \text{ s.t. } \forall x \in S, x \leq M) \\
\Leftrightarrow &\forall M \in \mathbb{R}, \text{not}(\forall x \in S, x \leq M) \\
\Leftrightarrow &\forall M \in \mathbb{R},\ \exists x \in S \text{ s.t. } \text{not}(x \leq M) \\
\Leftrightarrow &\forall M \in \mathbb{R},\ \exists x \in S \text{ s.t. } x > M.
\end{aligned}$$

[22] It is not necessary to include all these steps if you are asked to write down a negated form of a quantified statement. What matters is that you can do so accurately, as you'll often need to do this in order to work out what needs to be proved.

Take particular care when negating complicated quantified statements which involve implications.

Direct proofs of quantified statements

Suppose we are given a quantified statement

$$\forall a \in A, \exists b \in B \text{ s.t. } \forall c \in C, P(a,b,c). \tag{5}$$

What structure should a direct proof of this statement take[23]? The statement (5) makes a claim about all $a \in A$; we must prove "$\exists b \in B$ s.t. $\forall c \in C, P(a,b,c)$" for every value of a. We start our proof by fixing the symbol a, to indicate to the reader that it represents an arbitrary element of A (but won't change in the rest of the argument). I would open the formal proof with "Let $a \in A$ be arbitrary".

With our fixed value of a, we now have to prove "$\exists b \in B$ s.t. $\forall c \in C, P(a,b,c)$", so we should show the reader that there is some $b \in B$ which has the property "$\forall c \in C, P(a,b,c)$." Note that the b is allowed to depend on a, but the same b must work for all values of $c \in C$, i.e. we can think of b as a function of a. One way to do this, is to state a formula, or expression for b, in terms of a[24]. You'll probably need to find such a condition by means of a load of rough computations, which don't need to appear in your final proof; instead the next step of the formal proof, would be to take an arbitrary $c \in C$ and then carefully check that $P(a,b,c)$ is true. Thus the proof has the following structure.

- "Let $a \in A$ be arbitrary,"
- "Define $b \in B$ to be (some expression in terms of a", or perhaps "Take $b \in B$ to satisfy (some condition in terms of a)" where you can see the condition has at least one solution $b \in B$. For example, if B is the set $\mathbb{N}$ and a is a real number, a typical condition might be $b > a^2$, which we know has a solution[25].
- "Let $c \in C$ be arbitrary."
- Proof that $P(a,b,c)$ is true.

Let's see this in practice, with an example[26] we'll come back to in Chapter 2.

Example 1.7. *Prove*

$$\forall \varepsilon > 0, \exists n \in \mathbb{N} \textit{ s.t. } \frac{3n}{n+1} > 3 - \varepsilon.$$

We start our proof with "Let $\varepsilon > 0$ be arbitrary." Now we need to work out what value of $n \in \mathbb{N}$ we should take in terms of ε. Let's look at the condition that we need to satisfy and work out when it's true. We have

$$\frac{3n}{n+1} > 3 - \varepsilon \Leftrightarrow \varepsilon(n+1) > 3 \Leftrightarrow n > \frac{3}{\varepsilon} - 1.$$

Thus we should take any $n \in \mathbb{N}$ satisfying this last inequality. The point is that we know this last inequality has solutions in the natural numbers; this is not so obvious for the first form of the inequality.[27]

[23] We will see that the statement $\lim_{n\to\infty} x_n = L$ is exactly of this form, so it's a very relevant example for this course.

[24] This isn't the only way to show things exist; there are some beautiful non-constructive existence proofs in mathematics which demonstrate that certain objects exist without exhibiting them. You'll see some of these in future courses, but in this course, most of the time we need to show some thing exists, we'll be able to find some inequality, and then take anything satisfying this inequality.

[25] This looks very obvious; it will actually be a theorem, known as Archimedes axiom, in the next chapter, that it is always possible to find a natural number greater than a specified real number.

[26] albeit one with only two quantifiers.

[27] If you end up with an inequality of the form "$n < K$" it could be that something has gone wrong. In particular, if K is negative, this will not in general have solutions. If something like this happens in one of your answers, look carefully to see if you have multiplied an inequality by a negative number without reversing the direction of the inequality.

I might write "Take $n \in \mathbb{N}$ with $n > 3/\varepsilon - 1$" as the next line of my proof, and then say "for such an n, we have $(n+1)\varepsilon > 3$ and hence $\frac{3n}{n+1} > 3 - \varepsilon$, as required" to complete my proof.

Proof. Let $\varepsilon > 0$ be arbitrary. Take $n \in \mathbb{N}$ with $n > 3/\varepsilon - 1$. For such an n, we have $(n+1)\varepsilon > 3$ and hence $\frac{3n}{n+1} > 3 - \varepsilon$, as required. □

This isn't the only way such a proof can be written. You might think it is better to keep the indication of how the n was found, and write[28]:

Proof. Let $\varepsilon > 0$ be arbitrary. For $n \in \mathbb{N}$, we have

$$\frac{3n}{n+1} > 3 - \varepsilon \Leftrightarrow \varepsilon(n+1) > 3 \Leftrightarrow n > \frac{3}{\varepsilon} - 1.$$

Therefore take $n \in \mathbb{N}$ with $n > \frac{3}{\varepsilon} - 1$, so that this n satisfies $\frac{3n}{n+1} > 3 - \varepsilon$. □

[28] If you use the structure below, make sure you get the logic the 'right way round'. A standard error is to 'solve the inequality' and write

$$\begin{aligned} &\frac{3n}{n+1} > 3 - \varepsilon \\ \Longrightarrow &\varepsilon(n+1) > 3 \\ \Longrightarrow &n > \frac{3}{\varepsilon} - 1. \end{aligned}$$

and then say "So take $n > \frac{3}{\varepsilon} - 1$". Of course the implications above are correct; they're just not what is being used. What matters in this proof is that the implication $n > \frac{3}{\varepsilon} - 1 \Longrightarrow \frac{3n}{n+1} > 3 - \varepsilon$ is true — we can see that there is an n with $n > \frac{3}{\varepsilon} - 1$ and hence there is an n with $\frac{3n}{n+1} > 3 - \varepsilon$. If you are using the implication $n > \frac{3}{\varepsilon} - 1 \Longrightarrow \frac{3n}{n+1} > 3 - \varepsilon$ make sure you state this implication, not it's converse.

Inequalities and the modulus function

We use inequalities regularly in mathematical analysis in our arguments. Most often we do not need to get the best possible estimates to make our arguments work, and we can get away with correct but non-optimal inequalities in order to correctly prove statements. For example, in the two proofs given of Example 1.7 we took the inequality $\frac{3n}{n+1} > 3 - \varepsilon$ and rearranged it to make n the subject, showing that

$$\frac{3n}{n+1} > 3 - \varepsilon \Longleftrightarrow n > \frac{3}{\varepsilon} - 1.$$

However we didn't need to do this, and in some cases it may be essentially impossible to come up with a condition of the form $n > K$ which is *equivalent* to the condition we started with. Here's another proof of Example 1.7.

Proof. Let $\varepsilon > 0$ be arbitrary. For $n \in \mathbb{N}$ we have

$$\frac{3n}{n+1} > 3 - \varepsilon \Leftrightarrow \varepsilon(n+1) > 3 \Leftarrow \varepsilon n > 3 \Leftrightarrow n > \frac{3}{\varepsilon}.$$

Therefore take $n \in \mathbb{N}$ with $n > \frac{3}{\varepsilon}$, so that $\frac{3n}{n+1} > 3 - \varepsilon$ holds. □

Basically we've ended up with a "simpler condition" on n which implies that $\frac{3n}{n+1} > 3 - \varepsilon$. Of course, here it was straightforward to rearrange the inequality to make n the subject, so on this occasion we haven't made our life any easier with this third proof. In later examples, however, making the right sort of simplifying estimate will help a lot. The moral of the story is when you're doing this sort of "n-ε argument" we don't need to find the best n which works; any n which works will do.

The modulus function

The modulus function on the real line is defined by

$$|x| = \begin{cases} x, & x \geq 0; \\ -x, & x < 0. \end{cases}$$

We will repeatedly use the following four key properties of the modulus function.

a) For all $x, y \in \mathbb{R}$, $|xy| = |x||y|$;[29]

b) (The triangle inequality). For all $x, y \in \mathbb{R}$, $|x + y| \leq |x| + |y|$;[30]

c) For all $y \in \mathbb{R}$ and $r > 0$, $|y| < r \iff -r < y < r$;[31]

d) For all $x, a \in \mathbb{R}$ and $r > 0$,

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

[29] This can be proved directly from the definition above by checking all four cases; when x and y are both positive, when x is positive and y is negative, ...

[30] Again this can be proved by checking all four cases.

[31] This follows directly from the definition, and then one obtains (d) by taking $y = x - a$ in (c).

A key piece of intuition is that for $x, a \in \mathbb{R}$, the quantity $|x - a|$ represents **the distance from x to a**. We expect distances to obey the inequality[32]

$$\text{distance}(a, c) \leq \text{distance}(a, b) + \text{distance}(b, c).$$

[32] Think of a, b and c being the vertices on a triangle; it's shorter to go from a to c directly compared to traveling from a to c via b.

Taking $x = a - b$ and $y = b - c$ in the triangle inequality, we have

$$|a - c| = |(a - b) + (b - c)| \leq |a - b| + |b - c| \tag{6}$$

so verifying this intuition. It's from this view point that (b) should be thought of as the triangle inequality. The "trick" of adding and subtracting b as in (6) is used a lot; this trick and the inequality (6) needs to be part of your tool kit.

The condition $|x - a| < r$ in (d) says that "the distance between x and a is less than r", i.e. that x is trapped within a band of radius r around a (as described by the inequality $a - r < x < a + r$. We'll move between these two equivalent statements a lot from chapter 2 onwards.

Example 1.8. *Show that*[33] *for all $x \in \mathbb{R}$,*

$$|x - 2| \leq 1 \implies \frac{5}{6} \leq \left|\frac{x + 4}{x + 3}\right| \leq \frac{7}{4}. \tag{7}$$

Hence, find $K \in \mathbb{R}$ such that for all $x \in \mathbb{R}$,

$$|x - 2| \leq 1 \implies \left|\frac{x^2 + 7x + 7}{x + 3} - 5\right| \leq K|x - 2|. \tag{8}$$

[33] At this point it is far from clear why we would want to do this. By the end of the course, you will see that the estimates here form the meat of the proof that the function $f(x) = \frac{x^2+7x+7}{x+3}$ is continuous at $x = 2$.

Let's look at the implication (7) first. We're told that x is within distance 1 of the point 2, and have to obtain an estimate on the size of $\left|\frac{x+4}{x+3}\right|$. Using item (c) above[34], we have

$$\begin{aligned}|x-2| \leq 1 &\Leftrightarrow -1 \leq x-2 \leq 1\\ &\Leftrightarrow 4 \leq x+3 \leq 6\\ &\Leftrightarrow 5 \leq x+4 \leq 7.\end{aligned}$$

In particular, [35]

$$|x-2| \leq 1 \Rightarrow 4 \leq |x+3| \leq 6 \text{ and } 5 \leq |x+4| \leq 7. \tag{9}$$

Now

$$4 \leq |x+3| \leq 6 \Rightarrow \frac{1}{6} \leq \frac{1}{|x+3|} \leq \frac{1}{4},$$

so that[36]

$$|x-2| \leq 1 \Rightarrow \frac{5}{6} \leq \frac{|x+4|}{|x+3|} \leq \frac{7}{4}, \tag{10}$$

as required.

To make progress with the inequality (8) I'd start by simplifying the expression $\left|\frac{x^2+7x+7}{x+3} - 5\right|$, looking for factors of $|x-2|$. We have

$$\begin{aligned}\left|\frac{x^2+7x+7}{x+3} - 5\right| &= \left|\frac{x^2+7x+7-5x-15}{x+3}\right|\\ &= \left|\frac{(x-2)(x+4)}{x+3}\right| = \frac{|x+4|}{|x+3|}|x-2|.\end{aligned}$$

Thus we can take $K = \frac{7}{4}$, and then using (10), we have

$$|x-2| \leq 1 \Rightarrow \left|\frac{x^2+7x+7}{x+3} - 5\right| \leq K|x-2|.$$

A useful estimation lemma

We will often need to estimate polynomials $p(x)$ for large values of x. Intuitively we know that the dominating term is the term of highest degree, and the next lemma gives a method of controlling a polynomial by this term.[37]

Lemma 1.9 (Polynomial estimation lemma). *Let $n \in \mathbb{N}$, and suppose we are given real numbers $a_0, a_1, \cdots, a_n$ with $a_n > 0$. Write*

$$p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0.$$

Then there exists $N > 0$ such that

$$x \geq N \implies \frac{1}{2}a_n x^n \leq p(x) \leq \frac{3}{2}a_n x^n.$$

Proof. Let $n, a_n, \cdots, a_0$ be as in the statement of the lemma. Using the triangle inequality, and property (a) of the modulus function, we have

$$\begin{aligned}|p(x) - a_n x^n| &= |a_{n-1}x^{n-1} + \cdots + a_1 x + a_0|\\ &\leq |a_{n-1}||x|^{n-1} + \cdots + |a_1||x| + |a_0|. \end{aligned}\tag{11}$$

[34] we could also use item (d). Note too that while both of (c) and (d) are stated for strict inequalities, the corresponding versions also hold when we work with $\leq$ and $\geq$ as here.

[35] Would it be correct to use $\Leftrightarrow$ in place of $\Rightarrow$ in (9)?

[36] A typical mistake is to incorrectly conclude that

$$|x-2| \leq 1 \Rightarrow \frac{5}{4} \leq \frac{|x+4|}{|x+3|} \leq \frac{7}{6}$$

from (9) by not thinking about what happens when you "cross multiply" inequalities. Whenever you are multiplying and dividing inequalities makes sure what you've done is correct; errors like the one above happen quite frequently.

[37] When you're reading the proof of the lemma, there's a number of things you should try and think about in order to understand what is going on. Where is the hypothesis that $a_n > 0$ used — it's not explicitly mentioned? How would you state (and prove) a version of the lemma for a polynomial with $a_n < 0$? How would you prove a version of the lemma which replaced the conclusion (1.9) with

$$x \geq N \implies \frac{3}{4}a_n x^n \leq p(x) \leq \frac{5}{4}a_n x^n?$$

(This is perfectly possible, and the choice to use $\frac{1}{2}$ and $\frac{3}{2}$ in the lemma is just that — a choice). Could you prove the stronger statement that: $\forall \varepsilon > 0, \exists N > 0$ such that

$$x \geq N \implies (1-\varepsilon)\,a_n x^n \leq p(x) \leq (1+\varepsilon)\,a_n x^n?$$

Note that for $x \geq 1$, we have $|x|^r \leq x^{n-1}$ for $r = 0, 1, \ldots, n-1$. This gives

$$x \geq 1 \implies$$
$$|a_{n-1}||x|^{n-1} + \cdots + |a_1||x| + |a_0| \leq x^{n-1}\left(|a_{n-1}| + \cdots + |a_1| + |a_0|\right).$$

Setting $K = |a_{n-1}| + \cdots + |a_1| + |a_0|$ we can rewrite this as

$$x \geq 1 \implies |a_{n-1}||x|^{n-1} + \cdots + |a_1||x| + |a_0| \leq Kx^{n-1}. \tag{12}$$

Then, for $x > 0$, we have

$$Kx^{n-1} \leq \frac{a_n x^n}{2} \iff \frac{2K}{a_n} \leq x. \tag{13}$$

Take[38] $N = \max(1, \frac{2K}{a_n})$. For $x \geq N$, we have $x \geq 1$ and $x \geq \frac{2K}{a_n}$ so by (11), (12) and (13), we have

$$|p(x) - a_n x^n| \leq \frac{1}{2}a_n x^n \Leftrightarrow \frac{1}{2}a_n x^n \leq p(x) \leq \frac{3}{2}a_n x^n,$$

where the last equivalence is property (d) of the modulus function. □

[38] This is a really important idea, which you'll see repeatedly in this course. We want two conditions to hold, namely $x \geq 1$ and $x \geq \frac{2K}{a_n}$, and we need to package this in one condition of the form $x \geq N$. Taking N to be the maximum of 1 and $\frac{2K}{a_n}$ achieves this. Notice that given a finite list of real numbers $b_1, \ldots, b_m$, the number $\max(b_1, \ldots, b_m)$ is defined to be the largest of these numbers.

Note that we can assume that the $N > 0$ given in the lemma is a natural number if we wish (by replacing the N by a larger value which lies in $\mathbb{N}$).

Example 1.10. *Show that there exist $K, N > 0$ such that for $x \in \mathbb{R}$,*

$$x \geq N \implies \frac{3x^2 - 4x + 8}{5x + 6} \geq Kx. \tag{14}$$

Solution. By Lemma 1.9, there exist $N_1, N_2 > 0$ such that[39]

$$x \geq N_1 \implies \frac{1}{2}3x^2 \leq 3x^2 - 4x + 8 \leq \frac{3}{2}3x^2$$
$$x \geq N_2 \implies \frac{1}{2}5x \leq 5x + 6 \leq \frac{3}{2}5x.$$

[39] Note that we should use different symbols N_1 and N_2 for the two applications of the lemma in the first instance.

Then define $N = \max(N_1, N_2)$,[40] so that

$$x \geq N \implies \frac{1}{2}3x^2 \leq 3x^2 - 4x + 8 \leq \frac{3}{2}3x^2 \text{ and } \frac{1}{2}5x \leq 5x + 6 \leq \frac{3}{2}5x.$$

[40] Another application of the "max trick" — see side note 38.

You should note that the last condition $\frac{1}{2}5x \leq 5x + 6 \leq \frac{3}{2}5x$ ensures that $x \geq 0$, so we can deduce that $\frac{2}{15x} \leq \frac{1}{5x+6} \leq \frac{2}{5x}$. Therefore

$$x \geq N \implies \frac{3x^2 - 4x + 8}{5x + 6} \geq \frac{\frac{1}{2}3x^2}{\frac{3}{2}5x} = \frac{1}{5}x.$$

Hence the implication (14) holds with this value of N and $K = 1/5$. □

Using the estimation lemma 1.9 isn't the only approach here. You could also examine the inequality directly, arguing as follows.

Solution. For $x \geq 4$, we have $4x \leq x^2$ so that

$$3x^2 - 4x + 8 \geq 3x^2 - x^2 = 2x^2.$$

Similarly, for $x \geq 1$, we have

$$5x + 6 \leq 11x.$$

Therefore, when $x \geq 4$, we have

$$\frac{3x^2 - 4x + 8}{5x + 6} \geq \frac{2x^2}{11x} = \frac{2}{11}x,$$

and so we can take $N = 4$ and $K = \frac{2}{11}$. $\square$

Order, bounds, and axioms for the real numbers

The ordered field axioms

What are the real numbers anyway? We have been working with them for some time, but without having discussed what it means to be a real number, and exactly what properties they have. In this course we will take an axiomatic approach, describing precisely the properties, or rules, we require our real numbers to obey — in principle all other properties are deducible from these rules[1]. In our work on the foundations of analysis, we will focus on the key differences between the rational numbers and the real numbers.

Let us start with our axiomatic description of $\mathbb{R}$ by discussing the *field axioms*.

Axiom 2.1. The real numbers $\mathbb{R}$ are a set equipped with an addition $+$, and multiplication $\cdot$, and form a *field* under these operations, that is, the following nine axioms are satisfied[2].

a) $\forall a, b \in \mathbb{R}, a + b = b + a$;

b) $\forall a, b, c \in \mathbb{R}, (a + b) + c = a + (b + c)$;

c) $\exists 0 \in \mathbb{R}$ s.t. $\forall a \in \mathbb{R}, 0 + a = a$;

d) $\forall a \in \mathbb{R}, \exists(-a) \in \mathbb{R}$ s.t. $a + (-a) = 0$;

e) $\forall a, b \in \mathbb{R}, a \cdot b = b \cdot a$;

f) $\forall a, b, c \in \mathbb{R}, a \cdot (b \cdot c) = (a \cdot b) \cdot c$;

g) $\exists 1 \in \mathbb{R}$ s.t. $(1 \neq 0$ and $(\forall a \in A, a \cdot 1 = a))$;

h) $\forall a \in \mathbb{R}$ with $a \neq 0, \exists a^{-1} \in \mathbb{R}$ s.t. $a \cdot a^{-1} = 1$;

i) $\forall a, b, c \in \mathbb{R}, a \cdot (b + c) = (a \cdot b) + (a \cdot c)$;

What this essentially means is that we can add, subtract, multiply[3] and divide[4] real numbers in the usual way.

For example, the axioms imply the usual cancellation operations: if $a, b, x \in \mathbb{R}$ satisfy $a + x = b + x$, then $a = b$. To prove this formally, we would argue as follows. For $a, b, x \in \mathbb{R}$,

$$\begin{aligned} & a + x = b + x \\ \implies & (a + x) + (-x) = (b + x) + (-x) \\ \implies & a + (x + (-x)) = b + (x + (-x)) \\ \implies & a + 0 = b + 0 \\ \implies & a = b, \end{aligned}$$

[1] We will not, however, show how to construct the real numbers, i.e. demonstrate that there is a set which obeys these rules. This can be done, though it is relatively time consuming. If you're interested I'd recommend reading Rudin's account of the construction of the reals from the rationals via Dedekind cuts in Chapter 1 of his book. We will also not discuss uniqueness: to what extent is the real number system unique, or are there multiple constructions of systems obeying our axioms which behave differently? The short answer is that the real numbers are unique, but to make this precise would be too big a diversion.

[2] I do not expect you to remember these axioms in this course.

[3] We will usually omit the $\cdot$ sign and write ab instead of $a \cdot b$ for the product of two real numbers.

[4] provided that the denominator is nonzero, of course.

using axiom (d) for the existence of $(-x)$ for the first implication; (b) for the second; (d) for the third; and (c) for the last implication.

It is possible to use the algebraic axioms above to prove a number of simple consequences about the behaviour of the algebraic operations in a field. You should have seen some of these in 2B or 2F, and you can find more examples in Section 1.1 of [ERA].

We have seen a number of examples of fields in the mathematics studied so far: notably the rational numbers $\mathbb{Q}$, the real numbers $\mathbb{R}$, and the complex numbers $\mathbb{C}$. We will need extra axioms to distinguish $\mathbb{R}$ from these other examples.

The next step for this is to introduce the order relation $<$ on $\mathbb{R}$ and specify its behaviour.

Axiom 2.2 (Order axioms for $\mathbb{R}$)**.** There is a relation $>$ on $\mathbb{R}$ satisfying

j) For all $a \in \mathbb{R}$, exactly one of the statements $a = 0, a > 0$ and $0 > a$ is true;

k) For all $a, b \in \mathbb{R}$, $b > a \iff b - a > 0$;

l) For all $a, b \in \mathbb{R}$ with $a > 0$ and $b > 0$, we have $a + b > 0$ and $ab > 0$.

As usual, we also write $a < b$ instead of $b > a$, and $a \geq b$ for $a > b$ or $a = b$. Similarly, we also write $a \leq b$ for $b \geq a$. Notice that the order axiom only introduces the relation $>$, so that in our axiomatic approach we have to explain these other symbols before we can use them [5].

The thing to take away so far is that the usual arithmetic rules for manipulating addition, subtraction, multiplication and division, and the order structure can be axiomatised[6].

The rational numbers $\mathbb{Q}$ also have an order $<$, and they also obey the order axioms above — in particular, the 12 axioms so far do not suffice[7] to distinguish $\mathbb{R}$ from $\mathbb{Q}$. Among other things, this means it is not possible to prove there exists a real number x with $x^2 = 2$ from the ordered field axioms, since we know (from 1R and 1X) that there does not exist a rational number x with $x^2 = 2$. Therefore we should not talk of the square root function $\sqrt{\cdot}$ at present — we have not defined it yet[8]. Likewise, many other familiar mathematical functions, such as $\exp, \log, \sin, \cos$, have not yet been defined from our point of view — we will formally define these using power series in the 3H course Analysis of Differentiation and Integration. At the moment, the only operations that we have defined are the basic arithmetic operations of addition, subtraction, multiplication[9] and division. Nevertheless, functions like exp and log may occasionally appear in the exercises — when this happens you should feel free to use any familiar properties of these functions unless otherwise stated.

Bounds

In order to state the final axiom that is needed to describe the real numbers, we will have to investigate bounded sets. Let us make some precise definitions.

[5] Try to rewrite the conditions in the order axiom in terms of $<, \geq, \leq$.

[6] As already mentioned above, I do not expect you to know these axioms by heart, though of course I do expect you to be able to manipulate inequalities accurately.

[7] They suffice to distinguish $\mathbb{R}$ from $\mathbb{C}$: there is no order $<$ on $\mathbb{C}$ satisfying axioms (i), (j) and (k). Note that I really do mean that no possible order $<$ can be constructed, and I invite you to prove this as an extra exercise on sheet 2.

[8] By the end of this chapter, we will see one possible way of defining $\sqrt{\cdot}$.

[9] and hence taking powers of the form a^n where a is real and $n \in \mathbb{N}$.

Definition 2.3. Let $A \subseteq \mathbb{R}$. We say that $M \in \mathbb{R}$ is *an upper bound for* A if and only if for all $a \in A$, we have $a \leq M$. Define A to be *bounded above* if and only if there exists an upper bound for A.

Similarly, $m \in \mathbb{R}$ is said to be a *lower bound* for $A \subseteq \mathbb{R}$ if and only if for all $a \in A$, we have $m \leq a$. Say that A is *bounded below* if and only if there exists a lower bound for A.

Symbolically, the set $A \subseteq \mathbb{R}$ is bounded above if and only if

$$\exists M \in \mathbb{R} \text{ s.t. } \forall a \in A, a \leq M,$$

and A is bounded below if and only if

$$\exists m \in \mathbb{R} \text{ s.t. } \forall a \in A, m \leq a.$$

Definition 2.4. Let $A \subseteq \mathbb{R}$. We say that A is *bounded* if and only if A is bounded above and bounded below.

It is often useful to reformulate boundedness using the modulus function, as in the following lemma[10].

[10] I'm quite happy for you to use this lemma without comment in your work, so if you want to prove a set A is bounded, feel free to either show it is both bounded above, and bounded below, or show there exists $K > 0$ such that for all $x \in A$, we have $|x| \leq K$.

Lemma 2.5. *Let $A \subseteq \mathbb{R}$. Then A is bounded if and only if,*

$$\exists K > 0 \textit{ s.t. } \forall a \in A, |a| \leq K.$$

Proof. Suppose first that A is bounded. Then there exists $M \in \mathbb{R}$ and $m \in \mathbb{R}$ such that for all $x \in A$, we have $m \leq x \leq M$. Let us define $K = \max(|M|, |m|) + 1$, so that $K > 0$ and for $x \in A$, we have $|x| \leq K$ as desired[11].

[11] The reason I added 1 to $\max(|M|, |m|)$ was to ensure that $K > 0$ — if I'd just taken $K = \max(|M|, |m|)$, then when $M = m = 0$ (which could happen when $A = \{0\}$) we would have $K = 0$, and this doesn't satisfy all the conditions. There are other ways of solving this problem, for example, by taking $K = \max(|M|, |m|, 1)$.

Conversely, assume there exists $K > 0$ such that for all $x \in A$, we have $|x| \leq K$. Then $-K \leq x \leq K$, so A is bounded above by K and below by $-K$, that is, A is bounded. □

Let us look at some examples.

Example Show that the set $P = \{4\sin(x) - \cos(3y) \mid x, y \in \mathbb{R}\}$ is bounded and $Q = \{\frac{1}{x-1} \mid x > 1\}$ is not bounded.

Proof that P is bounded. For $x, y \in \mathbb{R}$, we have[12]

[12] Be careful in performing the inequalities below: in particular note how we use the triangle inequality to estimate the modulus $|4\sin(x) - \cos(3y)|$.

$$\begin{aligned}|4\sin(x) - \cos(3y)| &\leq |4\sin(x)| + |-\cos(3y)| \\ &\leq 4|\sin(x)| + |\cos(3y)| \leq 5.\end{aligned}$$

Therefore P is bounded by Lemma 2.5. □

For Q it is worth getting a feel for what is going on. Firstly, we can see that all the elements of Q are positive, so Q is certainly bounded below. Therefore, we should be trying to show that Q is not bounded above. If you're not sure how to do this, write down the formal negation of the statement that Q is bounded above, so you know what to do. That is, Q is not bounded above if and only if

$$\forall M \in \mathbb{R}, \exists q \in Q \text{ s.t. } q > M.$$

This tells us we should start our proof by taking an arbitrary value of $M \in \mathbb{R}$, and showing that this is not an upper bound[13] for Q.

[13] In particular, an answer consisting of a waffly explanation that Q contains very large values, so isn't bounded above, won't get any credit in this course.

Proof that Q is not bounded. Let $M \in \mathbb{R}$ be arbitrary. If $M \leq 0$, then as $1 = \frac{1}{2-1} \in Q$, we see that M is not an upper bound for Q. Suppose then that $M > 0$. Take $x = \frac{1}{2M} + 1$, so that[14] $x > 1$. Then $\frac{1}{x-1} = 2M > M$. Since $2M \in Q$, it follows that M is not an upper bound for Q. Since $M \in \mathbb{R}$ was arbitrary, Q is not bounded above, and so not bounded. □

[14] How did I decide to make this choice? I knew that I wanted to have $x > 1$ with $\frac{1}{x-1} > M$, so I rearranged this last inequality: $\frac{1}{x-1} > M \Leftrightarrow \frac{1}{M} + 1 > x$ (provided $x - 1 > 0$). Therefore I need to choose some x satisfying $1 < x < 1 + \frac{1}{M}$, leading me to take the average of the two end points of this inequality, namely $x = 1 + \frac{1}{2M}$.

We apply all these boundedness definitions also to *functions*. Given a set X and a function $f : X \to \mathbb{R}$, we say that f is *bounded* (bounded above or bounded below) if and only if the set $f(X)$ is bounded (bounded above or bounded below). Here

$$f(X) = \{f(x) \mid x \in X\}$$

is the range of f — the set of all values taken by f, compare 2F.

Example Let $f : \mathbb{N} \to \mathbb{R}$ be given by

$$f(n) = \frac{2n^2 - 7n + 1}{3n^2 - 2n - 5}.$$

Show that f is bounded above.

We can approach this example in a number of different ways. Firstly, let us use the polynomial estimation lemma (Lemma 1.9) to control the values of $f(n)$ in terms of the leading terms of the denominator and numerator when n is large.

Solution. By Lemma 1.9, there exist[15] $N_1, N_2 \in \mathbb{N}$ such that

$$\begin{aligned} n \geq N_1 &\implies \frac{1}{2}2n^2 \leq 2n^2 - 7n + 1 \leq \frac{3}{2}2n^2; \\ n \geq N_2 &\implies \frac{1}{2}3n^2 \leq 3n^2 - 2n - 5 \leq \frac{3}{2}3n^2. \end{aligned}$$

[15] Recall the remark following lemma 1.9, that says that we can take the real number N in that lemma to be a natural number.

Define $N = \max(N_1, N_2)$ so that for $n \geq N$, we have

$$f(n) \leq \frac{\frac{3}{2}2n^2}{\frac{1}{2}3n^2} = 2.$$

Define[16] $M = \max(f(1), f(2), \ldots, f(N-1), 2)$. Then, for any $n \in \mathbb{N}$, we have $f(n) \leq M$, so that f is bounded above[17]. □

[16] The point is that since the domain of f is $\mathbb{N}$, there are only finitely many values $n \in \mathbb{N}$ with $n \leq N$, so we can list $f(1), \ldots, f(N-1)$ and take the largest number in this list.

[17] To see this, note that if $n \leq N - 1$, we have $f(n) \leq \max(f(1), \ldots, f(N-1)) \leq M$, while if $n \geq N$, then $f(n) \leq 2 \leq M$.

We could also proceed directly, by thinking about how to get upper bounds for the numerator and denominator.

Another Solution. For $n \in \mathbb{N}$, we have $2n^2 - 7n + 1 \leq 2n^2$ (as $-7n + 1 \leq 0$). We also have[18] $2n + 5 \leq n^2$ when $n \geq 4$. Therefore, for $n \geq 4$, we have $3n^2 - 2n - 5 \geq 3n^2 - n^2 = 2n^2$. This gives

[18] Make sure that you check this.

$$f(n) \leq \max(f(1), f(2), f(3), 1),$$

and so f is bounded above[19]. □

[19] The advantage of this method, is that we could compute $f(1), f(2), f(3)$ to find an explicit upper bound, unlike the previous solution, which only shows that an upper bound exists, but doesn't give a method for finding one.

Least upper bounds and greatest lower bounds

Let $A \subseteq \mathbb{R}$ be a set and let M be an upper bound for A. Then any real number K with $K \geq M$ is also an upper bound for A; in particular, if A has an upper bound then it will have infinitely many upper bounds. So if you are asked to prove a set A is bounded above, exhibiting any upper bound M will do, you do not need to try and find the smallest M which works[20].

Our next focus is on looking for the best, or more accurately *least* upper bound; this concept will enable us to describe the final axiom needed to distinguish between $\mathbb{Q}$ and $\mathbb{R}$.

Definition 2.6. Let $A \subseteq \mathbb{R}$ and $M \in \mathbb{R}$. Define M to be a[21] *least upper bound for A* if and only if the following two conditions are satisfied:

a) M is an upper bound for A; and

b) for all upper bounds M' for A, we have $M \leq M'$.

Note that we do not talk of the "maximum element" of A. For example, when $A = (0, 1)$, it is straightforward to see that 1 is a least upper bound for A, but $1 \notin A$, so it is not a maximal element of A. Note also that when $B = (0, 1]$ we again have that 1 is a least upper bound for B, so it is possible for a least upper bound to be in the set, and also for a least upper bound not to be in the set. When we work with least upper bounds, the following reformulation of the second condition is very helpful[22].

Lemma 2.7. *Let $A \subseteq \mathbb{R}$ and let $M \in \mathbb{R}$ be an upper bound for A. Then M is a least upper bound for A if and only if*

$$\forall \varepsilon > 0, \exists a \in A \text{ s.t. } a > M - \varepsilon. \tag{1}$$

Proof. $\Rightarrow$ Let M be a least upper bound for A, and let $\varepsilon > 0$ be arbitrary. Then $M - \varepsilon < M$, so $M - \varepsilon$ is not an upper bound for A. Therefore[23] there exists $a \in A$ with $a > M - \varepsilon$.

$\Leftarrow$ Let M be an upper bound for A and suppose that condition (1) holds. Let $M' < M$, and write $\varepsilon = M - M'$, so that $\epsilon > 0$. Then $M' = M - \varepsilon$, and so the condition shows that there exists $a \in A$ with $a > M'$. Thus M' is not an upper bound for A, and hence M is a least upper bound for A. □

Throughout this course I will be happy for you to use this lemma to prove that M is a least upper bound for a set A without mentioning it[24]. Let's do a quick example.

Example Show that 3 is a least upper bound of $A = \left\{ \frac{3n}{n+1} \mid n \in \mathbb{N} \right\}$.

Solution. We first show that 3 is an upper bound. Given $n \in \mathbb{N}$, we have

$$\frac{3n}{n+1} \leq \frac{3n}{n+1} + \frac{3}{n+1} = 3,$$

[20] For example, consider $A = \{\cos(x) + \sin(x+3) \mid x \in \mathbb{R}\}$. We can see that for any $x \in \mathbb{R}$, $\cos(x) + \sin(x+3) \leq 1 + 1 = 2$, so the set A is certainly bounded above by 2. We could work harder to try and find a smaller upper bound for A, but we don't need to in order to show that A is bounded above.

[21] The use of "a" here is deliberate. We write "a" when it is possible that there could be more than one object satisfying a certain definition. In Theorem 2.9 below, we will show that a set has at most one least upper bound. From that point on we can refer to "the" least upper bound (when it exists).

[22] Putting it in the form of a quantified statement, in a style we've discussed how to prove.

[23] Recall that, by negating the definition of M' is an upper bound for A, M' is not an upper bound for A if and only if there exists $a \in A$ with $a > M'$.

[24] So if you want to prove that M is a least upper bound for A, first show M is an upper bound for A, then show $\forall \varepsilon > 0, \exists a \in A$ s.t. $a > M - \varepsilon$.

so 3 is an upper bound for A. Now, let $\varepsilon > 0$ be arbitrary. We need to show that there exists $n \in \mathbb{N}$ such that

$$\frac{3n}{n+1} > 3 - \varepsilon,$$

as then $a = \frac{3n}{n+1} \in A$, and $a > 3 - \varepsilon$, showing that 3 is a least upper bound for A by Lemma 2.7. Fortunately we've shown that such an n exists in example 1.7! $\square$

We also introduce the analogous concept of greatest lower bounds.

Definition 2.8. Let $A \subseteq \mathbb{R}$ and $m \in \mathbb{R}$. Define m to be a *greatest lower bound for A* if and only if:

a) m is an lower bound for A; and

b) for all lower bounds m' for A, we have $m' \leq m$.

In the same way as Lemma 2.7, given a lower bound m for A, m is a greatest lower bound for A if and only if

$$\forall \varepsilon > 0, \exists a \in A \text{ s.t. } a < m + \varepsilon.$$

We also note that least upper bounds, and greatest lower bounds are unique (when they exist).

Theorem 2.9. *Let $A \subseteq \mathbb{R}$. Then A has at most one least upper bound and at most one greatest lower bound*[25].

[25] This is a uniqueness theorem, and we prove it using a standard strategy for theorems of this sort: we assume there are two least upper bounds, and show that they are equal. Compare this with the proof that the identity element in a group is unique from 2F.

Proof. Let us consider uniqueness of least upper bounds. Suppose that M and M' are least upper bounds for A. Then, as M is a least upper bound for A, and M' is an upper bound for A, we have $M \leq M'$. Similarly, as M' is a least upper bound for A, and M is an upper bound for A, we have $M' \leq M$. Therefore $M = M'$.

The proof of the uniqueness of greatest lower bounds (when they exist) is similar, it is left as an exercise. $\square$

Now we know that least upper bounds and greatest lower bounds are unique (when they exist), it makes sense to refer to "the" least upper bound and "the" greatest lower bound, respectively. We also introduce some terminology which we'll use henceforth[26]. When M is the least upper bound for A, we call M the *supremum* of A and write $M = \sup(A)$. When m is the greatest lower bound for A we call m the infimum of A and write $M = \inf(A)$.

[26] The terminology "least upper bound" is useful as it describes the definition precisely, but its a bit cumbersome: supremum and infimum are more standard.

Example Show that

$$\inf \left\{ \frac{n+1}{n} \mid n \in \mathbb{N} \right\} = 1.$$

Solution. Let $A = \left\{ \frac{n+1}{n} \mid n \in \mathbb{N} \right\}$. We have to prove two statements, namely that 1 is a lower bound for A, and secondly that it is the greatest lower bound for A[27]. Firstly, for $n \in \mathbb{N}$, we have

[27] We expect the first statement to be easier to prove than the second statement.

$$\frac{n+1}{n} \geq 1,$$

so that 1 is a lower bound[28] for A. Now let $\varepsilon > 0$ be arbitrary[29]. For $n \in \mathbb{N}$, we have

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

Now take [30] $n \in \mathbb{N}$ with $n > \frac{1}{\varepsilon}$, so that $\frac{n+1}{n} \in A$ and $\frac{n+1}{n} < 1+\varepsilon$. Therefore we have $\inf(A) = 1$. $\square$

[28] In this case, this was quite straightforward, but nevertheless we should still write enough so that someone reading our work can see that we have checked that 1 is a lower bound.

[29] We are going to prove the quantified statement $\forall \varepsilon > 0, \exists a \in A$ such that $a < 1+\varepsilon$, so we start by letting $\varepsilon > 0$ be arbitrary.

[30] The point is that we can see that such an n exists, as the natural numbers are not bounded above (see the discussion of Archimedes axiom later), in contrast it wasn't so immediate that there is an $n \in \mathbb{N}$ such that $\frac{n+1}{n} < 1+\varepsilon$.

Let's do one more example. This time the set we will establish the supremum of is specified in terms of two other sets.

Example Let A and B be two subsets of $\mathbb{R}$ such that $\sup(A)$ and $\sup(B)$ exists. Define

$$C = \{3a + b \mid a \in A, b \in B\}.$$

Show that $\sup(C)$ exists and $\sup(C) = 3\sup(A) + \sup(B)$.

The strategy will be to show that the number $M = 3\sup(A) + \sup(B)$ satisfies the definition to be a least upper bound[31], i.e. that M is an upper bound and that it satisfies the statement $\forall \varepsilon > 0, \exists c \in C$ with $c > M - \varepsilon$. To do this we will need to use the defining properties of $\sup(A)$ and $\sup(B)$.

[31] As you'll see, to prove that M is an upper bound for C, we will use that $\sup(A)$ is an upper bound for A, and that $\sup(B)$ is an upper bound for B. Then to prove the second statement, we use the corresponding statements for $\sup(A)$ and $\sup(B)$: i.e. that $\forall \varepsilon > 0, \exists a \in A$ s.t. $a > \sup(A) - \varepsilon$ and $\forall \varepsilon > 0, \exists b \in B$ s.t. $b > \sup(B) - \varepsilon$.

Solution. Let $M = 3\sup(A) + \sup(B)$. For $a \in A$ and $b \in B$, we have $a \leq \sup(A)$ and $b \leq \sup(B)$ (as $\sup(A)$ and $\sup(B)$ are upper bounds of A and B, respectively). Therefore $3a + b \leq 3\sup(A) + \sup(B) = M$, i.e. M is an upper bound for C.

Now let $\varepsilon > 0$ be arbitrary. Then[32] $\frac{\varepsilon}{4} > 0$. As $\sup(A)$ is a least upper bound for A, there exists $a \in A$ with $a > \sup(A) - \frac{\varepsilon}{4}$, and as $\sup(B)$ is a least upper bound for B, there exists $b \in B$ with $b > \sup(B) - \frac{\varepsilon}{4}$. Therefore[33]

$$3a + b > 3\left(\sup(A) - \frac{\varepsilon}{4}\right) + \left(\sup(B) - \frac{\varepsilon}{4}\right)$$
$$= 3\sup(A) + \sup(B) - \varepsilon = M - \varepsilon.$$

Since $3a + b \in C$, this shows that M is the least upper bound for C, so that $\sup(C)$ exists and is equal to $M = 3\sup(A) + \sup(B)$. $\square$

[32] We will use $\frac{\varepsilon}{4}$ in the definition of least upper bound of A and B respectively, so here I am emphasising that we can do this as $\frac{\varepsilon}{4} > 0$.

[33] In this calculation, we use $\frac{3\varepsilon}{4} + \frac{\varepsilon}{4} = \varepsilon$, and it is because of this I made the choice to use $\frac{\varepsilon}{4}$ in the first place. I'd seen this coming, when I made the first choice.

The completeness axiom

We now turn to the key additional axiom for the real numbers which distinguishes $\mathbb{R}$ from $\mathbb{Q}$, namely, we insist that those subsets of $\mathbb{R}$ which could possibly have suprema do have suprema.

Axiom 2.10 (The completeness axiom)**.** Every non-empty subset of $\mathbb{R}$ which is bounded above has a supremum.

Note that there are two conditions needed before we learn that $A \subseteq \mathbb{R}$ has a supremum. Firstly, the set A needs to be non-empty[34]. Secondly, if A is not bounded above, it certainly can not have a least upper bound [35].

We will see why the completeness axiom is required at the end of the chapter, when we will use it to show that there exists $x \in \mathbb{R}$ with $x^2 = 2$. Since we know that such an x is necessarily irrational, we can not use the earlier axioms to prove the completeness axiom (as the rational numbers do not satisfy the completeness axiom).

We start with an immediate consequence of the axiom, namely the corresponding statement for greatest lower bounds[36].

[34] The empty set is bounded above, as any $M \in \mathbb{R}$ is an upper bound for the empty set — this also shows that the empty set does not have a least upper bound.

[35] Don't forget these two conditions, as I expect you to be able to state the completeness axiom precisely.

[36] It's reasonable to ask why Theorem 2.11 is a theorem and not an axiom. The reason is that mathematicians want to use the smallest possible set of axioms to describe $\mathbb{R}$, and prove other properties from these axioms. Since we can prove Theorem 2.11 from the completeness axiom (together with the other axioms for $\mathbb{R}$), we should do so.
An equally valid approach is to make Theorem 2.11 the axiom, and prove from this axiom that every non-empty subset of $\mathbb{R}$ which is bounded above has a supremum.

Theorem 2.11. *Every non-empty subset of $\mathbb{R}$ which is bounded below has an infimum.*

The proof I give here is rather similar to the example of $C = \{3a + b \mid a \in A, b \in B\}$ above. For a different proof see Theorem 1.4.7 in ERA.

Proof. Let A be a non-empty subset of $\mathbb{R}$ which is bounded below. Define $B = \{-a \mid a \in A\}$. Since A is non-empty, there exists $a \in A$, so that $-a \in B$ and hence B is non-empty. Since A is bounded below, take $m \in \mathbb{R}$ such that for all $a \in A, m \leq a$. Then $-a \leq -m$ for all $a \in A$, so that B is bounded above. Therefore $\sup(B)$ exists by the completeness axiom[37].

We now claim that $-\sup(B)$ is the greatest lower bound of A, and so $\inf(A)$ exists. For $a \in A$, we have $-a \in B$, so that $-a \leq \sup(B)$, and hence $-\sup(B) \leq a$. Therefore $-\sup(B)$ is a lower bound for A. Now let $\varepsilon > 0$ be arbitrary. There exists $a \in A$ such that $-a > \sup(B) - \varepsilon$, so that $a < -\sup(B) + \varepsilon$, and so $-\sup(B)$ is the greatest lower bound for A, as required. □

[37] Note that I check that B satisfies the hypotheses of the completeness axiom before I assert that $\sup(B)$ exists.

Two consequences of completeness

We end this chapter with two consequences of completeness. The first is that the natural numbers are not bounded above. Archimedes observed that an additional ingredient beyond the usual properties of addition, multiplication (which today we encode in the field axioms) is needed to describe the real numbers, and so introduced this fact as an axiom[38].

Let us see how Archimedes axiom follows from our completeness axiom.

Theorem 2.12 (Archimedes axiom)**.** *The natural numbers $\mathbb{N} \subseteq \mathbb{R}$ are not bounded above. In particular, given any $x \in \mathbb{R}$, there exists $n \in \mathbb{N}$ such that $n > x$.*

Proof. We argue by contradiction, so suppose that $\mathbb{N}$ is bounded above in $\mathbb{R}$. Then as $\mathbb{N}$ is certainly non-empty, $M = \sup(\mathbb{N})$ exists by the completeness axiom. As M is an upper bound for $\mathbb{N}$, we have $n \leq M$ for all $n \in \mathbb{N}$. Given $n \in \mathbb{N}$, we have $n + 1 \in \mathbb{N}$. Therefore

[38] We can prove directly that $\mathbb{N}$ is not bounded above in $\mathbb{Q}$ as follows: let $q \in \mathbb{Q}$ be arbitrary. If $q \leq 0$, then $1 \in \mathbb{N}$ has $1 > q$, so q is not an upper bound for $\mathbb{N}$. If $q > 0$, write $q = \frac{m}{n}$ for $m, n \in \mathbb{N}$. Then $q = \frac{m}{n} < m + 1 \in \mathbb{N}$, so q is not an upper bound for $\mathbb{N}$. The reason this argument works here is that we have an explicit form for every element of $\mathbb{Q}$; this is no longer the case for $\mathbb{R}$.
We do need an extra axiom beyond the ordered field axioms to see that the natural numbers are not bounded above in $\mathbb{R}$. In his book 'A companion to analysis: A Second First and First Second Course in Analysis', Tom Körner constructs an example of an ordered field in which the natural numbers are bounded above. In particular, this field has an infinitesimal element $\alpha > 0$ with the property that $\alpha < 1/n$ for all $n \in \mathbb{N}$.

$n+1 \leq M$, and hence $n \leq M-1$. This proves that $M-1$ is also an upper bound for $\mathbb{N}$, but since $M-1 < M$, this contradicts the fact that M is the least upper bound for $\mathbb{N}$. Therefore $\mathbb{N}$ is not bounded above. $\square$

We have already used Archimedes axiom repeatedly in examples. In example 1.7, we proved that

$$\forall \varepsilon > 0, \exists n \in \mathbb{N} \text{ s.t. } \frac{3n}{n+1} < 3 - \varepsilon.$$

To do this we showed that for a fixed value of $\varepsilon > 0$, we had

$$\frac{3n}{n+1} < 3 - \varepsilon \Leftrightarrow n > \frac{3}{\varepsilon} - 1.$$

Now we can see that it is possible to take $n \in \mathbb{N}$ with $n > \frac{3}{\varepsilon} - 1$, as $\frac{3}{\varepsilon} - 1$ is not an upper bound for $\mathbb{N}$.

Our second example shows that the axiom we have introduced is enough to show that the square root of 2 exists.

Theorem 2.13 (Square root 2 exists)**.** *There exists $x \in \mathbb{R}$ with $x^2 = 2$.*

How do we proceed to use the completeness axiom to prove this theorem? To use the completeness axiom to produce certain real numbers, we write down non-empty subsets A which are bounded above, and then get a real number $\sup(A)$. So we should look for a set $A \subset \mathbb{R}$ which we think will have supremum $\sqrt{2}$. Of course, the set $[0, \sqrt{2}]$ will have this property, for instance, but since the point of this theorem is to use the axioms for $\mathbb{R}$ to prove that $\sqrt{2}$ exists, we need to define a set A entirely using the basic algebraic operations and the order relation given to us by the algebraic and order axioms. We should then carefully check that our set satisfies the hypotheses of the completeness axiom, namely that A is non-empty and bounded above. Then $\sup(A)$ will exist, and the final and most difficult step of the proof is to show that $\sup(A)^2 = 2$.

Proof. Define

$$A = \{y \in \mathbb{R} \mid y^2 \leq 2\}.$$

As $0^2 \leq 2$, we have $0 \in A$, so A is non-empty. Now suppose $y \in \mathbb{R}$ has $y \geq 2$, then $y^2 \geq 4$, so $y \notin A$. Thus 2 is an upper bound for A and hence A is bounded above. Therefore, by the completeness axiom, $M = \sup(A)$ exists. Note that $M \geq 1$ as $1 \in A$ and $M \leq 2$ as 2 is an upper bound for A.

Suppose[39] $M^2 < 2$. Choose $\delta > 0$ with $\delta < \min(\frac{2-M^2}{5}, 1)$, and then define[40] $y = M + \delta$. As $\delta < 1$, we have $\delta^2 < \delta$ and so[41]

$$y^2 = (M+\delta)^2 = M^2 + 2M\delta + \delta^2 \leq M^2 + 4\delta + \delta = M^2 + 5\delta < 2.$$

Thus $y \in A$, but this is a contradiction as $y > M$. Therefore $M^2 \geq 2$.

Suppose[42] $M^2 > 2$. Choose $\delta > 0$ with $\delta < \min(M, \frac{M^2-2}{2M})$, so that $M^2 - 2M\delta > 2$ and $M - \delta > 0$. Then,

$$(M-\delta)^2 = M^2 - 2M\delta + \delta^2 \geq M^2 - 2M\delta > 2.$$

[39] We aim to reach a contradiction, and we will do this by showing that for a small enough value of δ we have $M + \delta \in A$, as then M will not be an upper bound for A.

[40] This choice of δ was made by doing the calculation first to see what would work. Note also, that unlike some calculations in lectures it's not an equivalent form of $(M+\delta)^2 < 2$, but instead an inequality which implies that $(M+\delta)^2 < 2$.

[41] In this estimate I use that $M \leq 2$.

[42] This time the contradiction will come from showing that $M - \delta$ is also an upper bound for A, provided δ is small enough.

Now if $y \geq (M-\delta)$, then $y^2 \geq (M-\delta)^2 > 2$ so $y \notin A$. Thus $M-\delta$ is an upper bound for A, contradicting the fact that M is the supremum of A.

Since $M^2 < 2$ and $M^2 > 2$ both lead to contradictions, we conclude that $M^2 = 2$, as required. $\square$

We could use the same process to show that for every $x \geq 0$, there exists $y \geq 0$ with $y^2 = x$, and use this to define the square root function[43]. In a similar fashion you should be able to show that for each $x \geq 0$ and each $n \in \mathbb{N}$, there exists $y \geq 0$ with $y^n = x$, so we can define powers of the form $x^{\frac{m}{n}}$ when $m, n \in \mathbb{N}$. Defining what we mean by irrational powers will take far more work, and will be done in the third year course.

[43] Note that the uniqueness of y doesn't require completeness: it can be deduced from the algebraic axioms, as if $y^2 = z^2$, then $(y-z)(y+z) = 0$, so $y = z$ or $y = -z$. Hence for each $x \geq 0$, there is a unique $y \geq 0$ with $y^2 = x$, and so we really do define a function when we do this.

2C Intro to real analysis 2020/21

Handout

Sequences

This chapter is devoted to *sequences*. We will start our study of the most central concept in analysis in this context, namely that of *limits*.

A real sequence is an infinite list of real numbers, like

$$1, 1/2, 1/3, 1/4, \ldots .$$

To make this precise we give the following formal definition.

Definition 3.1. A *real sequence* is a function $\mathbb{N} \to \mathbb{R}$.

According to this definition[1], we may denote the elements of a real sequence by $a(1), a(2), \ldots,$ corresponding to the values of the function $a : \mathbb{N} \to \mathbb{R}$. However, usually we prefer a slightly different notation. We will write a_n for the n-th term of a sequence a, rather than $a(n)$, and write[2] "let $(a_n)_{n=1}^{\infty}$ be a real sequence". It is equally possible to have sequences which start at other integers. For example, a sequence starting at $n = 0$ will be denoted $(a_n)_{n=0}^{\infty}$, and such a sequence[3] corresponds to a function $\{0, 1, 2, \ldots\} \to \mathbb{R}$. Most often in what follows sequences will be specified by giving a formula describing the n-th term, for example: "consider the sequence $(a_n)_{n=1}^{\infty}$ given by $a_n = (-1)^n$ for $n \in \mathbb{N}$." However, it would be a mistake to think that the n-th term of a sequence is always given by a nice formula: a sequence is just a list of real numbers, there is no reason why these terms must fit into a nice pattern[4].

[1] In this course I am defining the natural numbers to be the set $\{1, 2, \ldots, \}$, i.e. the smallest natural number is 1. Some authors consider 0 as the smallest natural number, but we will not do this.

[2] Note that [ERA], and some other books, use the notation $\{a_n\}_{n=1}^{\infty}$ for a sequence. The downside of this notation is that it leads to potential confusion between sequences and sets. For this reason, I will stick to using round brackets for sequences.

[3] Strictly speaking, this doesn't satisfy the definition of a sequence we've just given - but we'll follow common practice and allow for a wider use of the term *sequence*, including this variant.

[4] For instance, the sequence defined by

$$a_n = \begin{cases} e, & n = 1 \\ \pi, & n = 2 \\ 0, & n = 3, \\ -1, & n \geq 4 \end{cases}$$

is a perfectly valid sequence.

Note that the definitions of boundedness (bounded above, and bounded below) from the previous chapter apply to sequences. More precisely, a real sequence $(a_n)_{n=1}^{\infty}$ is *bounded above* if and only if there exists $M \in R$ such that for all $n \in \mathbb{N}$, we have $a_n \leq M$. The sequence is *bounded below* if and only if there exists $m \in \mathbb{R}$ such that for all $n \in \mathbb{N}$, we have $m \leq a_n$.

Limits of sequences

The definition of convergence below is perhaps the most important definition in this course. I expect you to know this definition and to be able to work with it.

Definition 3.2. Let $(a_n)_{n=1}^{\infty}$ be a real sequence and let $L \in \mathbb{R}$. We say that a_n *converges to* L as n tends to infinity if and only if for each $\varepsilon > 0$, there exists $n_0 \in \mathbb{N}$, such that for $n \in \mathbb{N}$ with $n \geq n_0$, we have $|a_n - L| < \varepsilon$. When a_n converges to L as n tends to infinity, we write[5] "$a_n \to L$ as $n \to \infty$" or $\lim_{n\to\infty} a_n = L$.

We say that the sequence $(a_n)_{n=1}^{\infty}$ *diverges* if it does not converge to any limit.

[5] The symbol $\to$ is reserved for this use, and also in specifying the domain and codomain of a function (for example, "$f : X \to Y$"). Please don't use it for any other purposes.

As we will discuss in more detail below, this definition encapsulates our intuitive notion that a sequence $(a_n)_{n=1}^{\infty}$ converges to the limit L if and only if the values of the sequence get closer and closer to L as n gets large. The important point here is that the above definition formalises this is in a precise mathematical way that doesn't use phrases like "closer and closer to L", which we couldn't use in a proof.

The definition can be written concisely as a quantified statement: the sequence $(a_n)_{n=1}^{\infty}$ converges to $L \in \mathbb{R}$ if and only if

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N} \text{ s.t. } \forall n \in \mathbb{N}, (n \geq n_0 \implies |a_n - L| < \varepsilon).$$

An alternative form is

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N} \text{ s.t. } \forall (n \in \mathbb{N} \text{ with } n \geq n_0), |a_n - L| < \varepsilon.$$

You will also often see the $n \in \mathbb{N}$ suppressed in the literature, and in lectures. In these statements, n must be a natural number as it is used as a subscript on a_n to indicate which term of the sequence we are taking, so in using this notation we are implicitly saying that n must be a natural number. Therefore you will see the definition of convergence written as

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N} \text{ s.t. } \forall n \geq n_0, |a_n - L| < \varepsilon,$$

or as[6]

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N} \text{ s.t. } n \geq n_0 \implies |a_n - L| < \varepsilon.$$

How to interpret the definition of convergence?

Let's try and understand how Definition 3.2 works[7]. In general, when I'm trying to understand a quantified statement, I start at the end and work backwards, so here I initially focus on the condition $|a_n - L| < \varepsilon$, and make sure I understand this. Remember that we view the quantity $|a_n - L|$ as the distance from a_n to L. Thus the last part "$|a_n - L| < \varepsilon$" of the definition says that "a_n is within distance ε of L"; in this statement n and ε are variables. The role ε plays in this definition is that of an allowed error tolerance, we are asking when is it true that the distance between a_n and L lies within this error tolerance.

Working back with the quantified statement, we next look at the statement "$\forall (n \in \mathbb{N} \text{ with } n \geq n_0), |a_n - L| < \varepsilon$" (or equivalently "$\forall n \in \mathbb{N}, n \geq n_0 \implies |a_n - L| < \varepsilon$") which is a statement with variables ε and n_0. This statement asks that for all x_n for $n \geq n_0$ lie within the error tolerance ε of L, and finally that $\exists n_0 \in \mathbb{N}$ s.t. $\forall n \in \mathbb{N}, n \geq n_0 \implies |a_n - L| < \varepsilon$, which asks that there is some index n_0, such that past this point in the sequence, the a_n lies within the error tolerance of L. We should think of $\exists n_0 \in \mathbb{N}$ s.t. $\forall n \in \mathbb{N}, n \geq n_0 \implies |a_n - L| < \varepsilon$ as meaning that "$|a_n - L| < \varepsilon$ holds *eventually*": for all sufficiently large n it's true that $|a_n - L| < \varepsilon$.

With $\varepsilon > 0$ fixed, another way to think about the condition $\exists n_0 \in \mathbb{N}$ s.t. $\forall n \in \mathbb{N}, n \geq n_0 \implies |a_n - L| < \varepsilon$ is to imagine that we

[6] I also want to discuss the effects of using $<$ and $\leq$ in this definition, as this is a regular source of confusion. It's possible to change the condition $|a_n - L| < \varepsilon$ to $|a_n - L| \leq \varepsilon$, and also change $n \geq n_0$ to $n > n_0$, without changing the meaning of the definition; making these changes doesn't affect the intuitive interpretation of the definition, and we can in fact give formal proofs that making these changes leads to equivalent statements.

For example, the statement that $a_n \to L$ as $n \to \infty$ is equivalent to

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N} \text{ s.t. } \forall n \geq n_0, |a_n - L| \leq \varepsilon. \tag{1}$$

It is immediate that if $a_n \to L$, then (1) holds (given $\varepsilon > 0$, take n_0 such that $n \geq n_0 \implies |a_n - L| < \varepsilon$, so that $n \geq n_0 \implies |a_n - L| \leq \varepsilon$). Conversely, suppose (1) holds, then given $\varepsilon > 0$, note that $0 < \frac{\varepsilon}{2} < \varepsilon$. Taking $\frac{\varepsilon}{2}$ in (1), gives $n_0 \in \mathbb{N}$ such that for $n \geq n_0$, we have $|a_n - L| \leq \frac{\varepsilon}{2} < \varepsilon$, proving that $a_n \to L$ as $n \to \infty$.

In a similar vein, $a_n \to L$ as $n \to \infty$ is equivalent to

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N} \text{ s.t. } \forall n > n_0, |a_n - L| < \varepsilon.$$

Can you see how to give a formal proof of this equivalence?

[7] It is really important to get to grips with this definition, and to be able to state it without having to look it up.

display the sequence on a computer or graphical calculator, with ε representing the size of one pixel on the screen, so that if two numbers differ by at most ε, then they are indistinguishable on the screen. From this view point, the statement $\exists n_0 \in \mathbb{N}$ s.t. $\forall n \geq n_0, |a_n - L| < \varepsilon$ says that there is some value n_0 such that from this point onwards, you can't distinguish the values a_n and L on the screen. Of course, one could change the resolution and zoom in: this corresponds to decreasing the value of ε, and convergence of the sequence means that no matter what the pixel size, if you look far enough along the sequence it is impossible to distinguish between values of the sequence and the limit.

Proving convergence *directly from the definition*

You will sometimes be asked to prove that certain sequences converge *directly from the definition*. This means that you must verify the condition in definition 3.2 directly, without using subsequent results. Results from earlier in the course, like the polynomial estimation lemma 1.9, can be used in such questions. As we have already seen, the definition can be written in the form of a quantified statement, with three quantifiers, so the general methods we discussed for proving such quantified statements in Chapter 1 apply.

When you are asked to prove $a_n \to L$ directly from the definition, you should start by introducing an arbitrary value of $\varepsilon > 0$ to work with by saying "Let $\varepsilon > 0$ be arbitrary." To complete the proof you need to give a value of n_0 for which $n \geq n_0 \implies |a_n - L| < \varepsilon$. For this it is best to start by examining the expression $|x_n - L| < \varepsilon$, and to try to simplify it. You may be able to rearrange the inequality $|a_n - L| < \varepsilon$ in the form $n > \dots$, giving you a value of n_0. However, it may not be that straightforward to rearrange $|a_n - L| < \varepsilon$ exactly for n. Fortunately we can get away with less: it suffices to find *some* value of n_0 such that $|a_n - L| < \varepsilon$ when $n \geq n_0$ — you do not need to find the least value of n_0 with this property. Note that when verifying the definition of convergence, the value of n_0 is allowed to depend on ε, and examples below will show that one should expect it to do so. When ε gets made smaller, we will typically have to use a larger value of n_0, corresponding to looking further down the sequence, in order to verify the condition $n \geq n_0 \implies |a_n - L| < \varepsilon$.

There are many ways you can write a solution, but it is usually best to do some rough work first to examine $|a_n - L|$ and find how large n needs to be to ensure that $|a_n - L| < \varepsilon$, and then start to write down the answer.

Let's start with two simple examples. The first one is really a warm-up exercise.

Example 3.3. Let $K \in \mathbb{R}$, and define $x_n = K$ for all $n \in \mathbb{N}$. Show that $x_n \to K$ as $n \to \infty$, directly from the definition.

Solution. As explained above, the first step in proving convergence directly from the definition is to give ourselves an arbitrary $\varepsilon > 0$. Next we have to find $n_0 \in \mathbb{N}$ such that $|x_n - K| \leq \varepsilon$ for all $n \geq n_0$.

In this example this is very easy: Taking $n_0 = 1$, for $n \in \mathbb{N}$ with $n \geq n_0$, we have $|x_n - K| = |K - K| = 0 < \varepsilon$. This means $x_n \to K$ as $n \to \infty$. $\square$

Sequences of the form $x_n = K$ for all $n \in \mathbb{N}$ as in the previous example are also called *constant sequences*. Thus, we have proved that constant sequences converge.

Example 3.4. Let $(x_n)_{n=1}^{\infty}$ be given by

$$x_n = \frac{1}{n}.$$

Show that $x_n \to 0$ as $n \to \infty$, directly from the definition.

Solution. Again, the first step is to let $\varepsilon > 0$ be arbitrary. Our task it to find $n_0 \in \mathbb{N}$ such that $|x_n - 0| = \frac{1}{n} < \epsilon$ for all $n \geq n_0$. Now we have

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

for $n \in \mathbb{N}$, so if we take[8] $n_0 \in \mathbb{N}$ such that $n_0 > \frac{1}{\varepsilon}$, then we get $|x_n - 0| < \epsilon$ for all $n \geq n_0$. That is, we have shown $x_n \to 0$ as $n \to \infty$. $\square$

[8] It's worth being clear that here, and elsewhere where we will use expressions like "Take $n_0 > \frac{14}{\varepsilon}$", this is justified by Archimedes' axiom, which we proved in Section 2 as a consequence of the completeness axiom. Indeed, we can find n_0 as claimed from the fact that the natural numbers are not bounded above. More precisely, for $\varepsilon > 0$, the expression $\frac{1}{\varepsilon}$ is not an upper bound for $\mathbb{N}$, and so there exists $n_0 \in \mathbb{N}$ with $n_0 > \frac{1}{\varepsilon}$. I don't expect you to justify this step in this way in your answers: you may assume that such an n_0 can be found without further comment, but it's worth being aware of what's going on here.

This example is good to keep in mind, it displays very clearly the idea behind the notion of convergence. Intuitively, the numbers $x_n = \frac{1}{n}$ get closer and closer to 0 as n gets large, so that we should expect the sequence $(x_n)_{n=1}^{\infty}$ to converge to 0. Indeed, that's exactly what we have just shown.

Let's have a look at more examples.

Example 3.5. Let $(x_n)_{n=1}^{\infty}$ be given by

$$x_n = \frac{3n-2}{n+4}.$$

Show that $x_n \to 3$ as $n \to \infty$, directly from the definition.

Solution. Let $\varepsilon > 0$. For $n \in \mathbb{N}$, we have[9]

$$\left| \frac{3n-2}{n+4} - 3 \right| = \left| \frac{3n-2-3(n+4)}{n+4} \right| = \frac{14}{n+4} < \frac{14}{n} < \varepsilon, \qquad (2)$$

provided $n > \frac{14}{\varepsilon}$. Take $n_0 \in \mathbb{N}$ with $n_0 \geq \frac{14}{\varepsilon}$. Then for $n \in \mathbb{N}$ with $n \geq n_0$, we have $|x_n - 3| < \varepsilon$, so that $x_n \to 3$ as $n \to \infty$. $\square$

[9] Note that in equation (2), the last inequalities $\frac{14}{n} < \varepsilon$ is not always true: this only holds under the condition that $n > \frac{14}{\varepsilon}$, hence the sentence continues with the condition "provided $n > \frac{14}{\varepsilon}$".

The step $\frac{14}{n+4} < \frac{14}{n}$ is not necessary in this answer; it's included as it is often a good idea to estimate complicated expressions in n, by simpler ones which you can still make small. Instead you may want to give a solution starting: "Let $\varepsilon > 0$ be arbitrary. Then for $n \in \mathbb{N}$, we have

$$\left| \frac{3n-2}{n+4} - 3 \right| = \left| \frac{3n-2-3(n+4)}{n+4} \right| = \frac{14}{n+4} < \varepsilon,$$

provided $n > \frac{14}{\varepsilon} - 4$. Take $n_0 \in \mathbb{N}$ with $n_0 > \frac{14}{\varepsilon} - 4 \ldots$".

As another alternative, you could have done the calculation $|x_n - L| < \varepsilon \iff \frac{14}{\varepsilon} - 4 < n$ as rough work, and then started your solution with: "Let $\varepsilon > 0$ be arbitrary. Take $n_0 \in \mathbb{N}$ with $n_0 > \frac{14}{\varepsilon} - 4$, and $n \in \mathbb{N}$ with $n \geq n_0$. Then ...".

For more complicated expressions, it may be convenient to use Lemma 1.9 to get the required estimate on n.

Example 3.6. Let $(y_n)_{n=1}^{\infty}$ be the sequence given by

$$y_n = \frac{3n^3 + 4n - 5}{5n^3 - 3n^2 + 4}$$

Show that $y_n \to \frac{3}{5}$ as $n \to \infty$.

Solution. Let $\varepsilon > 0$ be arbitrary. For $n \in \mathbb{N}$, we have

$$\left|\frac{3n^3 + 4n - 5}{5n^3 - 3n^2 + 4} - \frac{3}{5}\right| = \left|\frac{9n^2 + 20n - 37}{5(5n^3 - 3n^2 + 4)}\right|.$$

By[10] Lemma 1.9 there exist $n_1, n_2 \in \mathbb{N}$ such that

$$n \geq n_1 \implies \frac{1}{2}9n^2 \leq 9n^2 + 20n - 37 \leq \frac{3}{2}9n^2$$
$$n \geq n_2 \implies \frac{1}{2}25n^3 \leq 5(5n^3 - 3n^2 + 4) \leq \frac{3}{2}25n^3.$$

In particular, when $n \geq \max(n_1, n_2)$, we have

$$\left|\frac{9n^2 + 20n - 37}{5(5n^3 - 3n^2 + 4)}\right| = \frac{9n^2 + 20n - 37}{5(5n^3 - 3n^2 + 4)} \leq \frac{\frac{3}{2}9n^2}{\frac{1}{2}25n^3} = \frac{27}{25n} < \varepsilon,$$

provided n also satisfies $n > \frac{27}{25\varepsilon}$. Therefore[11] take $n_0 \in \mathbb{N}$ with $n_0 > \max(n_1, n_2, \frac{27}{25\varepsilon})$. For $n \in \mathbb{N}$ with $n \geq n_0$, we have $\left|y_n - \frac{3}{5}\right| < \varepsilon$, so $y_n \to \frac{3}{5}$ as $n \to \infty$. □

You don't need to use Lemma 1.9 to answer this question, another approach would be to estimate

$$\left|\frac{9n^2 + 20n - 37}{5(5n^3 - 3n^2 + 4)}\right|$$

directly[12].

One may ask "why doesn't this work when you change $\frac{3}{5}$ to some other number?", i.e. why can't we use Lemma 1.9 to prove that y_n converges to any real number we like? This is a very good question, and exactly the sort of question you should be asking yourself when you read an argument in order to make sure you really understand what's going on. The best way to answer this question is to try and see what happens. Here if we replace $\frac{3}{5}$ by, say, 1, and perform the initial calculation, we get

$$\left|\frac{3n^3 + 4n - 5}{5n^3 - 3n^2 + 4} - 1\right| = \left|\frac{2n^3 - 3n^2 - 4n + 9}{5n^3 - 3n^2 + 4}\right|.$$

Now when we use Lemma 1.9, for sufficiently large $n \in \mathbb{N}$ we have

$$\left|\frac{3n^3 + 4n - 5}{5n^3 - 3n^2 + 4} - 1\right| \leq \frac{\frac{3}{2}2n^3}{\frac{1}{2}5n^3} = \frac{6}{5}$$

[10] In your answers to the feedback exercises I expect you to say "By Lemma 1.9" or "By the polynomial estimation lemma" if you use it, however in the exam I do not expect you to know the numbers of results in the course, so when you use a non-named result say something like "by a lemma from lectures". If a result you want to use has a name, such as the monotone convergence theorem below, you should use that name.

[11] This is the "max trick" we saw earlier. We take $n_0 > \max(n_1, n_2, \frac{27}{25\varepsilon})$ as in order to conclude that $|y_n - \frac{3}{5}| < \varepsilon$, we need three conditions to hold: $n \geq n_1, n \geq n_2$ and $n > \frac{27}{25\varepsilon}$. We must package these three conditions in the form $n \geq n_0$ for some suitable n_0, so the easiest option is to take $n_0 > \max(n_1, n_2, \frac{27}{25\varepsilon})$.

[12] For example, I note that for $n \geq 2$, we have $9n^2 + 20n - 37 > 0$ so $|9n^2 + 20n - 37| = 9n^2 + 20n - 37$, and then $9n^2 + 20n - 37 \leq 9n^2 + 20n^2 = 29n^2$. Similarly, we have $5n^3 - 3n^3 + 4 > 0$ for all $n \in \mathbb{N}$, so $|5(5n^3 - 3n^2 + 4)| = 5(5n^3 - 3n^2 + 4)$. Then $5(5n^3 - 3n^2 + 4) \geq 5(5n^3 - 3n^3) = 10n^3$. Therefore, for $n \geq 2$, we have

$$\left|\frac{9n^2 + 20n - 37}{5(5n^3 - 3n^2 + 4)}\right| \leq \frac{29n^2}{10n^3} = \frac{29}{10n}.$$

Then if $\varepsilon > 0$ is specified, we can take $n_0 \in \mathbb{N}$ with $n_0 > \max(2, \frac{29}{10\varepsilon})$ in order to verify that $y_n \to \frac{3}{5}$ as $n \to \infty$.

All the approximations above are made so that I end up with the estimate below, which takes the same form (that is, $|y_n - \frac{3}{5}| \leq \frac{K}{n}$ for some constant $K > 0$) as in the solution above.

which you can't make smaller than an arbitrary value of $\varepsilon > 0$. This also happens with

$$\left| \frac{3n^3 + 4n - 5}{5n^3 - 3n^2 + 4} - L \right|$$

for any other value[13] $L \in \mathbb{R}$ with $L \neq \frac{3}{5}$.

[13] Note that $\frac{3}{5}$ is the only value of L for which when you write

$$\frac{3n^3 + 4n - 5}{5n^3 - 3n^2 + 4} - L = \frac{p(n)}{5n^3 - 3n^2 + 4}$$

for some polynomial $p(n)$, you get cancellation of the n^3 term of $p(n)$, so that $p(n)$ has a lower degree than $5n^3 - 3n^2 + 4$. This was vital to our method.

The key thing that I want you to learn here is that when you come up with a question like this, don't be afraid to try an example and see if you can work out what's going on.

Proving that sequences do *not* converge directly from the definition

We can also use Definition 3.2 to check that a sequence does *not* converge.

Example 3.7. Show that the sequence $(z_n)_{n=1}^{\infty}$ given by $z_n = (-1)^n$, does not converge to any limit, directly from the definition.

Solution. Suppose that there exists $L \in \mathbb{R}$ with $z_n \to L$ as $n \to \infty$. Taking $\varepsilon = 1 > 0$, in the definition of convergence, there exists $n_0 \in \mathbb{N}$, such that for $n \in \mathbb{N}$ with $n \geq n_0$, we have $|z_n - L| < \varepsilon$. In particular, for $n \geq n_0$, note that $|z_n - z_{n+1}| = 2$, as one of z_n and z_{n+1} is 1, while the other is -1. Then[14]

$$\begin{aligned} 2 = |z_n - z_{n+1}| &= |z_n - L + L - z_{n+1}| \\ &\leq |z_n - L| + |L - z_{n+1}| < \varepsilon + \varepsilon = 2. \end{aligned}$$

[14] A key use of the triangle inequality is coming up: we estimate the distance from z_n to z_{n+1} as being at most the distance from z_n to L plus the distance from L to z_{n+1}.

This is a contradiction. Therefore the sequence $(z_n)_{n=1}^{\infty}$ does not converge to L. Since $L \in \mathbb{R}$ was arbitrary, the sequence does not converge to any limit. □

To understand how this answer was constructed, let's have a think about this sequence. We see directly from the formula that the sequence $(z_n)_{n=1}^{\infty}$ alternates between 1 and -1, more precisely it takes the form $-1, +1, -1, +1, \ldots$. If it converges to L, then, fixing a value of $\varepsilon > 0$, for sufficiently large n, the values z_n must lie in the ε-error band around L. In particular both $+1$ and -1 must lie[15] in the interval $(L - \varepsilon, L + \varepsilon)$. This interval has total width 2ε. Since 1 and -1 are distance 2 apart, we will get a contradiction when $2\varepsilon \leq 2$, leading to the choice $\varepsilon = 1$ in the answer above[16].

[15] You might find it useful to draw a picture of this.

[16] In particular, any value of ε with $0 < \varepsilon \leq 1$ would work in the above proof, but values of ε with $\varepsilon > 1$ would not.

Properties of limits

In this section we examine how limits interact with the algebraic operations and the order structure on the real numbers, and obtain rules which demonstrate that our intuition for how limits should behave is valid. These rules provide methods for calculating limits[17] and we use these to establish some standard limits.

[17] which you should feel free to use whenever a question does not require you to work directly from the definition. Still, you should always indicate to the person reading your work which rules you are using.

Intuitively, we don't expect a sequence to be able to converge to two separate values. The next result justifies our intuition.

Theorem 3.8. *Let $(x_n)_{n=1}^{\infty}$ be a convergent sequence. Then its limit is unique.*

Proof. Suppose that[18] $x_n \to L$ and $x_n \to M$ as $n \to \infty$ for some

[18] This is the usual method for starting a uniqueness proof; we introduce two quantities which both satisfy the relevant condition and show that they are the same. This idea was already used to prove that suprema are unique in chapter 2.

$L, M \in \mathbb{R}$. Suppose that $L \neq M$, and take $\varepsilon = \frac{|L-M|}{2}$. Then the definition of convergence gives $n_1, n_2 \in \mathbb{N}$ with

$$n \geq n_1 \implies |x_n - L| < \varepsilon$$
$$n \geq n_2 \implies |x_n - M| < \varepsilon.$$

Take[19] $n = \max(n_1, n_2)$. Then

$$|L-M| = |L - x_n + x_n - M| \leq |L - x_n| + |x_n - M| < \varepsilon + \varepsilon = |L-M|,$$

a contradiction. Therefore $L = M$. □

[19] The "max trick" again.

Note the similarity of the above argument with the proof that $(-1)^n$ does not converge: the key idea is to take ε to be half the distance between L and M, just as we took ε to be half the distance between 1 and -1 in Example 3.7.

Theorem 3.9 (Convergent sequences are bounded). *Let $(x_n)_{n=1}^\infty$ be a convergent sequence. Then $(x_n)_{n=1}^\infty$ is bounded.*

Proof. Suppose that $x_n \to L$ as $n \to \infty$. Taking $\varepsilon = 1$ in the definition of convergence[20] there exists $n_0 \in \mathbb{N}$ such that for $n \geq n_0$, we have $|x_n - L| < \varepsilon$. In particular, for $n \geq n_0$ we have[21]

$$|x_n| = |x_n - L + L| \leq |x_n - L| + |L| < 1 + |L|.$$

Define $M = \max(|x_1|, |x_2|, \ldots, |x_{n_0-1}|, |L| + 1) > 0$. Then for all $n \in \mathbb{N}$, $|x_n| \leq M$, so that $(x_n)_{n=1}^\infty$ is bounded. □

[20] Note that the value 1 was a choice I made, to see whether the exact value matters you could try changing the value of ε. You should find that the proof works no matter what value of $\varepsilon > 0$ you take here.

[21] Note that the estimate does not yet prove that $(x_n)_{n=1}^\infty$ is bounded; in particular we may not have $|x_n| \leq 1 + |L|$ for all $n \in \mathbb{N}$, we only have this for those n with $n \geq n_0$.

The converse to Theorem 3.9 is false: not every bounded sequence converges. For example, taking $z_n = (-1)^n$, gives an example of a bounded sequence which does not converge. The Bolzano–Weierstrass Theorem below encapsulates what can be said about bounded sequences with respect to convergence.

We now come to the key algebraic properties of limits, which show how limits of sequence interact with the algebraic operations of addition, subtraction, multiplication, and division.

Theorem 3.10 (Algebraic properties of limits). *Let $(x_n)_{n=1}^\infty$ and $(y_n)_{n=1}^\infty$ be sequences with limits L and M, respectively.*

a) For $\lambda \in \mathbb{R}$, we have $\lambda x_n \to \lambda L$ as $n \to \infty$;

b) $x_n + y_n \to L + M$ as $n \to \infty$;

c) $x_n y_n \to LM$ as $n \to \infty$;

d) If $M \neq 0$, then $x_n / y_n \to L/M$ as $n \to \infty$.

Note that in the last statement we have to insist that $M \neq 0$, so that we do not try to divide by zero. We don't demand that $y_n \neq 0$ for all n, so the sequence $(\frac{x_n}{y_n})_n$ might not be defined for all $n \in \mathbb{N}$. However, as $y_n \to M \neq 0$, there exists $n_1 \in \mathbb{N}$ such that for $n \geq n_1$, we have $y_n \neq 0$ (take $\varepsilon = |M|$ in the definition of convergence), and so the sequence $(\frac{x_n}{y_n})$ is defined for $n \geq n_1$, and that's all we need to discuss the convergence of $\frac{x_n}{y_n}$ to $\frac{L}{M}$.

I'm going to prove the second and third statement, and encourage you to see if you can prove the other two statements. The strategy we use is similar in spirit to the ideas used to show that if A and B are non-empty subsets of $\mathbb{R}$ which are bounded above, and $C = \{a + b \mid a \in A, b \in B\}$, then $\sup(C) = \sup(A) + \sup(B)$.

Proof of b). Let $(x_n)_{n=1}^{\infty}$ and $(y_n)_{n=1}^{\infty}$ be sequences with limits L and M, respectively. Let $\varepsilon > 0$ be arbitrary. Then[22] $\frac{\varepsilon}{2} > 0$, and by the definition of convergence there exists $n_1, n_2 \in \mathbb{N}$ such that

$$n \geq n_1 \implies |x_n - L| < \frac{\varepsilon}{2}$$
$$n \geq n_2 \implies |y_n - M| < \frac{\varepsilon}{2}.$$

Take $n_0 = \max(n_1, n_2)$. Then, for $n \geq n_0$, we have

$$\begin{aligned}|(x_n + y_n) - (L + M)| &= |(x_n - L) + (y_n - M)| \\ &\leq |x_n - L| + |y_n - M| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, $x_n + y_n \to L + M$ as $n \to \infty$. □

[22] I'm going to use $\frac{\varepsilon}{2}$ in the definition that $x_n \to L$ and $y_n \to M$. Can you see why I want to take $\frac{\varepsilon}{2}$ here?

The key thing that made this proof work was the use of the triangle inequality $|(x_n - y_n) + (L - M)| \leq |x_n - L| + |y_n - M|$ at the end, as this enables me to control the distance from $x_n + y_n$ to $L + M$ in terms to the two distances I know something about, namely $|x_n - L|$ and $|y_n - M|$. I had found this inequality before embarking on writing my proof.

For the third part, we will want to make $|x_n y_n - LM|$ smaller than some quantity ε, based on our ability to control $|x_n - L|$ and $|y_n - M|$. This leads us to add and subtract the term $x_n M$ leading to the critical inequality. More precisely, we consider

$$\begin{aligned}|x_n y_n - LM| &= |x_n y_n - x_n M + x_n M - LM| \\ &\leq |x_n||y_n - M| + |x_n - L||M|.\end{aligned}$$

With this in mind, we can write down a proof[23].

Proof of c). Let $(x_n)_{n=1}^{\infty}$ and $(y_n)_{n=1}^{\infty}$ be sequences with limits L and M respectively. Let $\varepsilon > 0$ be arbitrary. By Theorem 3.9, the sequence $(x_n)_{n=1}^{\infty}$ is bounded, so there exists $K > 0$ such that $|x_n| \leq K$ for all $n \in \mathbb{N}$. As $x_n \to L$ and $y_n \to M$, there exists $n_1, n_2 \in \mathbb{N}$ such that[24]

$$n \geq n_1 \implies |x_n - L| < \frac{\varepsilon}{2(|M| + 1)}$$
$$n \geq n_2 \implies |y_n - M| < \frac{\varepsilon}{2K}.$$

Take $n_0 = \max(n_1, n_2)$. For $n \in \mathbb{N}$ with $n \geq n_0$, we have

$$\begin{aligned}|x_n y_n - LM| &= |x_n y_n - x_n M + x_n M - LM| \\ &\leq |x_n||y_n - M| + |x_n - L||M| \\ &\leq K|y_n - M| + |x_n - L|(|M| + 1) \\ &< K\frac{\varepsilon}{2K} + \frac{\varepsilon}{2(|M| + 1)}(|M| + 1) = \varepsilon,\end{aligned}$$

so that $x_n y_n \to LM$ as $n \to \infty$. □

[23] We do have an additional difficulty to overcome. We must not try and ask for n large enough so that $|y_n - M| < \frac{\varepsilon}{2|x_n|}$, as the quantity we put into the definition that $y_n \to M$ can not depend on n. For this reason we use Theorem 3.9 to find $K > 0$ such that $|x_n| \leq K$ for all n.

[24] I use $2(|M| + 1)$ rather than $2|M|$ below as we don't know that $M \neq 0$; this way we don't have to handle this issue separately.

For part d) you'll need to find a key inequality before you start, relating

$$\left|\frac{x_n}{y_n} - \frac{L}{M}\right|$$

to $|x_n - L|$ and $|y_n - M|$. Try writing the expression above as a single fraction, and then look to add and subtract a term, as in part c). You'll need to take care to insist that y_n is bounded away from 0 for large enough[25] n.

[25] It'll probably help to show first that there exists $n_1 \in \mathbb{N}$, such that for $n \geq n_1$, $|y_n| \geq \frac{|M|}{2}$. If you get stuck, there is a proof in [ERA].

Let us revisit the examples we have seen above using properties of limits. A key strategy when taking limits of fractions, as in the question below, is to identify the dominating terms as n gets large, and divide the numerator and denominator by this quantity.

Example 3.11. Use properties of limits to find

$$\lim_{n\to\infty} \frac{3n^3 + 4n + 5}{5n^3 - 3n^2 + 4}.$$

Solution. The dominating terms in the numerator and denominator are $3n^3$ and $5n^3$, respectively. For $n \in \mathbb{N}$, we have

$$\begin{aligned}
\frac{3n^3 + 4n + 5}{5n^3 - 3n^2 + 4} &= \frac{3 + \frac{4}{n^2} + \frac{5}{n^3}}{5 - \frac{3}{n} + \frac{4}{n^3}} \\
&= \frac{3 + 4\left(\frac{1}{n}\right)^2 + 5\left(\frac{1}{n}\right)^3}{5 - 3\left(\frac{1}{n}\right) + 4\left(\frac{1}{n}\right)^3} \\
&\to \frac{3 + 4 \times 0^2 + 5 \times 0^3}{5 - 3 \times 0 + 4 \times 0^3} = \frac{3}{5},
\end{aligned}$$

as $n \to \infty$, using the standard limit $\frac{1}{n} \to 0$ as $n \to \infty$, and algebraic properties of limits[26]. □

[26] It's always a good idea to indicate what you are using.

In the solution above it's important to use the symbols, $\to$ and $=$ correctly. Remember that when we take limits, we use the symbol $\to$ to indicate this: in particular notice that

$$\frac{3 + 4\left(\frac{1}{n}\right)^2 + 5\left(\frac{1}{n}\right)^3}{5 - 3\left(\frac{1}{n}\right) + 4\left(\frac{1}{n}\right)^3} \neq \frac{3 + 4 \times 0^2 + 5 \times 0^3}{5 - 3 \times 0 + 4 \times 0^3},$$

as the former is an expression in n, the latter a number. Make sure you don't write $=$ unless two quantities are equal[27].

[27] It's also important to make sure you do connect all these mathematical expressions, you certainly can't write them in a list down the side of the page without any indication of how they're related.

The Sandwich Principle and standard limits

Before coming back to more examples of this type, it's useful to have more standard limits at our disposal. To do this, we investigate how limits interact with the order properties of $\mathbb{R}$.

Theorem 3.12 (Limits and order). *Let $(x_n)_{n=1}^\infty$ and $(y_n)_{n=1}^\infty$ be sequences, and suppose $x_n \to L$ and $y_n \to M$ as $n \to \infty$. If there exists $N \in \mathbb{N}$ such that for all $n \geq N$ we have $x_n \leq y_n$, then $L \leq M$.*

Proof. Let $(x_n)_{n=1}^{\infty}$ and $(y_n)_{n=1}^{\infty}$ be sequences, let $N \in \mathbb{N}$ be such that $x_n \leq y_n$ for $n \geq N$, and suppose $x_n \to L$ and $y_n \to M$ as $n \to \infty$. Suppose that $L > M$, and take $\varepsilon = \frac{L-M}{2}$. Then there exists $n_1, n_2 \in \mathbb{N}$ such that

$$n \geq n_1 \implies |x_n - L| < \varepsilon;$$
$$n \geq n_2 \implies |y_n - M| < \varepsilon.$$

Take $n \geq \max(N, n_1, n_2)$. Then $|x_n - L| < \varepsilon$, so $x_n > L - \varepsilon = M + \varepsilon$. Also $|y_n - M| < \varepsilon$, so $y_n < M + \varepsilon$. Combining these inequalities gives $x_n > y_n$, which contradicts the fact that $x_n \leq y_n$ for $n \geq N$. Therefore $L \leq M$. □

Most often we use this when one sequence is constant, for example, taking $x_n = 0$ for all n, the proposition says that if $(y_n)_{n=1}^{\infty}$ is a sequence of eventually positive terms, with $y_n \to M$ as $n \to \infty$, then $M \geq 0$. Note that we must use inequalities of the form $\leq$ and $\geq$ in these results: the information $y_n > 0$ for all n, and $y_n \to M$ as $n \to \infty$, does not[28] imply $M > 0$.

[28] Consider $y_n = \frac{1}{n}$.

Theorem 3.13 (The sandwich principle). *Let $(x_n)_{n=1}^{\infty}, (y_n)_{n=1}^{\infty}$ and $(z_n)_{n=1}^{\infty}$ be sequences and suppose that $x_n \to L$ and $z_n \to L$ as $n \to \infty$. If there exists $N \in \mathbb{N}$ such that*

$$n \geq N \implies x_n \leq y_n \leq z_n,$$

then $\lim_{n\to\infty} y_n = L$.

Proof. Let $(x_n)_{n=1}^{\infty}, (y_n)_{n=1}^{\infty}, (z_n)_{n=1}^{\infty}, L$ and N be as in the statement of the theorem, and let $\varepsilon > 0$ be arbitrary. Then there exists $n_1, n_2 \in \mathbb{N}$ such that

$$n \geq n_1 \implies |x_n - L| < \varepsilon \implies x_n > L - \varepsilon.$$
$$n \geq n_2 \implies |z_n - L| < \varepsilon \implies z_n < L + \varepsilon.$$

Then take $n_0 = \max(n_1, n_2, N)$. For $n \geq n_0$, we have

$$L - \varepsilon < x_n \leq y_n \leq z_n < L + \varepsilon \implies |y_n - L| < \varepsilon.$$

Therefore $y_n \to L$ as $n \to \infty$. □

We shall often apply the sandwich principle when one of the sequences $(x_n)_{n=1}^{\infty}$ or $(z_n)_{n=1}^{\infty}$ is constant; the challenge is then to find the other sequence which should be easier to work with[29] and relatively straightforward to provide the required inequalities. Don't be afraid to put some large values of n into your calculator to make sure you're aiming for the right limit.

[29] For instance, formed from standard limits.

Example 3.14. Evaluate

$$\lim_{n\to\infty}(\sqrt{n+4} - \sqrt{n}), \quad \lim_{n\to\infty}(3^n + 5^n)^{1/n}.$$

Solution. For the first example, we use a standard trick for working with surds, namely multiplying by $\frac{\sqrt{n+4}+\sqrt{n}}{\sqrt{n+4}+\sqrt{n}}$ in order to use the difference between two squares formula[30]. For $n \in \mathbb{N}$, we have

[30] I'm expecting to try and prove that the limit is 0, using a calculator if necessary to see this. So I would like to write the original expression using terms like $\frac{K}{\sqrt{n}}$, as we know that these converge to 0.

$$0 \leq (\sqrt{n+4} - \sqrt{n}) = \frac{4}{\sqrt{n+4} + \sqrt{n}} \leq \frac{4}{\sqrt{n}} \to 0,$$

as $n \to \infty$.[31] By the sandwich principle,

$$\lim_{n\to\infty} (\sqrt{n+4} - \sqrt{n}) = 0.$$

For the second, again I use a calculator if necessary to decide that the limit of this sequence will be 5, which we get from $(5^n)^{1/n}$. Thus what is happening is that the 5^n term dominates, and the 3^n term becomes insignificant for large n. This way of thinking leads to the following calculation:

$$5 = (0^n + 5^n)^{1/n} \leq (3^n + 5^n)^{1/n} \leq (5^n + 5^n)^{1/n} = 5 \times 2^{1/n} \to 5,$$

as $n \to \infty$, using the standard limit[32] $2^{1/n} \to 1$. Therefore, by the sandwich principle

$$\lim_{n\to\infty} (3^n + 5^n)^{1/n} = 5$$

as claimed. □

[31] Can you prove that $\frac{4}{\sqrt{n}} \to 0$ as $n \to \infty$ directly from the definition?

[32] See the exercises, and the comments on standard limits below.

In order to record another useful limit, we present the following lemma.

Lemma 3.15. *Let $(x_n)_{n=1}^{\infty}$ be a sequence and $L \in \mathbb{R}$. Then*

$$x_n \to L \text{ as } n \to \infty \quad \Leftrightarrow \quad |x_n - L| \to 0 \text{ as } n \to \infty.$$

The lemma is proved by noting that $||x_n - L| - 0| = |x_n - L|$. I leave it to you to write up the details.

Theorem 3.16. *Let $x \in \mathbb{R}$. Then*

$$x^n \to 0 \iff |x| < 1.$$

Proof. Suppose that $|x| \geq 1$. Then $|x|^n \geq 1$ for all n, and hence[33] $|x|^n \not\to 0$. Therefore $x^n \not\to 0$.

Conversely, when $x = 0$, we have $x^n = 0$ for all n, so that $x^n \to 0$, so suppose that $0 < |x| < 1$. Write $\frac{1}{|x|} = 1 + K$ for some $K > 0$. Then, using the binomial expansion, we have

$$\frac{1}{|x|^n} = (1+K)^n = 1 + nK + \cdots + K^n \geq nK.$$

Therefore

$$0 \leq |x|^n \leq \frac{1}{nK} \to 0,$$

as $n \to \infty$. By the sandwich principle, $|x|^n \to 0$ as $n \to \infty$. This implies $x^n \to 0$ due to Lemma 3.15. □

[33] Can you prove this directly from the definition? What value of ε should you take?

Example 3.17. Evaluate

$$\lim_{n\to\infty} \frac{5^n + 3^n}{5^n - 3^n}.$$

This works in a very similar way to previous "algebraic properties" type questions. Identify the dominating term[34] and divide denominator and numerator by this term[35].

[34] In this question that's 5^n.

[35] As with the other algebraic properties questions be careful how you write your solution. In particular, make sure you never claim that some expression in n is equal to it's limit, and use the symbols $\lim_{n\to\infty}$, $=$ and $\to$ correctly.

Solution. Using Theorem 3.16 and algebraic properties of limits, we obtain

$$\frac{5^n + 3^n}{5^n - 3^n} = \frac{1 + \left(\frac{3}{5}\right)^n}{1 - \left(\frac{3}{5}\right)^n} \to \frac{1+0}{1-0} = 1,$$

as $n \to \infty$, since $|\frac{3}{5}| < 1$. □

In exercises using the sandwich principle and properties of limits, you should feel free to use standard limits we've either proved above, or on exercises freely in order to compute other limits[36]. The key standard limits are:

[36] though you should always say that you're doing so.

Theorem 3.18 (Standard limits)**.** *The following results hold:*

a) $\frac{1}{n^\alpha} \to 0$ *as* $n \to \infty$ *for any* $\alpha > 0$.

b) $x^n \to 0$ *as* $n \to \infty$ *if and only if* $|x| < 1$.

c) $x^{1/n} \to 1$ *as* $n \to \infty$ *for all* $x > 0$.

Proof. $a)$ This is an exercise on sheet 5.
$b)$ This is theorem 3.16.
$c)$ We will only prove this for $x \geq 1$ here[37]. Let us write $x = 1 + c$ for $c \geq 0$. Moreover let $\varepsilon > 0$ be arbitrary. Then $|x^{1/n} - 1| < \varepsilon$ if and only if $x < (1+\varepsilon)^n$. Moreover $1 + n\varepsilon \leq (1+\varepsilon)^n$ by the binomial formula. Hence if we chose $n_0 \in \mathbb{N}$ such that $n_0 > \frac{c}{\varepsilon}$, we obtain

[37] Can you prove the claim for $0 < x < 1$?

$$x = 1 + c \leq 1 + n\frac{c}{n_0} < 1 + n\varepsilon \leq (1+\varepsilon)^n$$

for $n \geq n_0$. Therefore $x^{1/n} \to 1$ as $n \to \infty$ in this case. □

Monotonic sequences

We now turn to some important consequences of completeness. Our first objective is the monotone convergence theorem. We start by introducing the definitions of monotonic sequences: those that are either increasing or decreasing.

Definition 3.19. Let $(x_n)_{n=1}^\infty$ be a sequence. We say that $(x_n)_{n=1}^\infty$ is

- *increasing* if and only if, for all $n \in \mathbb{N}, x_n \leq x_{n+1}$;
- *strictly increasing* if and only if, for all $n \in \mathbb{N}, x_n < x_{n+1}$;
- *eventually increasing* if and only if, there exists $N \in \mathbb{N}$, such that for all $n \geq N, x_n \leq x_{n+1}$;
- *eventually strictly increasing* if and only if, there exists $N \in \mathbb{N}$, such that for all $n \geq N, x_n < x_{n+1}$.

In a similar fashion, say that $(x_n)_{n=1}^\infty$ is

- *decreasing* if and only if, for all $n \in \mathbb{N}, x_n \geq x_{n+1}$;
- *strictly decreasing* if and only if, for all $n \in \mathbb{N}, x_n > x_{n+1}$;

- *eventually decreasing* if and only if, there exists $N \in \mathbb{N}$, such that for all $n \geq N, x_n \geq x_{n+1}$;
- *eventually strictly decreasing* if and only if, there exists $N \in \mathbb{N}$, such that for all $n \geq N, x_n > x_{n+1}$;

Note that by this definition a constant sequence is both increasing and decreasing. In fact, a sequence is constant if and only if it is both increasing and decreasing.

Definition 3.20. Let $(x_n)_{n=1}^{\infty}$ be a sequence. We say that $(x_n)_{n=1}^{\infty}$ is *monotonic* if and only if it is either increasing or decreasing. We say that $(x_n)_{n=1}^{\infty}$ is *eventually monotonic* if it is eventually increasing or eventually decreasing.

There are two standard strategies to show that a sequence $(x_n)_{n=1}^{\infty}$ is (eventually) monotonic. Firstly, to decide whether $(x_n)_{n=1}^{\infty}$ is eventually increasing or eventually decreasing you can look at the difference $x_{n+1} - x_n$, simplify this, and see if you can show that $x_{n+1} - x_n$ is eventually positive (when $(x_n)_{n=1}^{\infty}$ will be increasing) or eventually negative (when $(x_n)_{n=1}^{\infty}$ will be decreasing). Let's see an example in action.

Example 3.21. Let $x_n = \frac{n^2-3n}{n^2-5}$ and $y_n = \frac{n!}{n^n}$. Show that $(x_n)_{n=1}^{\infty}$ and $(y_n)_{n=1}^{\infty}$ are eventually monotonic.

Solution. We have[38]

$$x_{n+1} - x_n = \frac{3n^2 - 7n + 10}{(n^2-5)(n^2+2n-4)} > 0,$$

for $n \geq 3$. Therefore $(x_n)_{n=1}^{\infty}$ is eventually increasing.

[38] How did I arrive at the last inequality? Firstly, for $n \geq 3$, both $n^2 - 5 \geq 0$ and $n^2 + 2n - 4 \geq 0$, so the denominator is positive for $n \geq 3$. For the numerator, we have $3n^2 - 7n \geq n(3n-7) \geq 0$ for $n \geq 3$, so the numerator is also positive when $n \geq 3$.

A more systematic approach to this example would be to use Lemma 1.9. If you end up with an expression of the form

$$x_{n+1} - x_n = \frac{p(n)}{q(n)}$$

for polynomials $p(n)$ and $q(n)$ whose leading terms have positive coefficients, then applying Lemma 1.9 suitably will show you[39] that $x_{n+1} - x_n > 0$ for sufficiently large n.

[39] I leave the details to you. Similarly, you can check that if $x_{n+1} - x_n = -\frac{p(n)}{q(n)}$ for polynomials $p(n)$ and $q(n)$ whose leading terms have positive coefficients, Lemma 1.9 will show that $(x_n)_{n=1}^{\infty}$ is eventually decreasing.

Let us now turn to the sequence $(y_n)_{n=1}^{\infty}$. We have $y_n = \frac{n!}{n^n} > 0$ for all n. Moreover

$$\frac{y_{n+1}}{y_n} = \frac{(n+1)!}{(n+1)^{n+1}} \frac{n^n}{n!} = \left(\frac{n}{n+1}\right)^n < 1,$$

so the sequence $(y_n)_{n=1}^{\infty}$ is strictly decreasing.

We see here another key method for deciding whether a sequence is monotonic, which applies when all the terms are positive[40]: when $y_n > 0$ for all n, then we can consider the ratio $\frac{y_{n+1}}{y_n}$ as $y_{n+1} > y_n$ if and only if $\frac{y_{n+1}}{y_n} > 1$.

[40] or at least when they all have the same sign.

So which method should you use? I'd suggest looking at the ratio $\frac{y_{n+1}}{y_n}$ in those cases where the form of y_n is such that there will be lots of cancellation in the expression $\frac{y_{n+1}}{y_n}$. For example, this happens when y_n is defined using exponential functions of n, like $y_n = 7^n$, where we have $\frac{y_{n+1}}{y_n} = 7$. Similarly things like factorials[41] and binomial coefficients (which can be expressed using factorials) naturally bring ratios to mind. □

The key reason for focusing on monotonic sequences is the monotone convergence theorem, which in turn is a consequence of the completeness axiom[42]. This theorem says that **bounded monotonic sequences converge** — a slogan you should know, and which should come to mind whenever you see a monotonic sequence.

Theorem 3.22 (Monotone convergence theorem). *Let $(x_n)_{n=1}^{\infty}$ be an eventually increasing sequence which is bounded above, or an eventually decreasing sequence which is bounded below. Then $(x_n)_{n=1}^{\infty}$ converges.*

Proof. I will prove the theorem in the case that $(x_n)_{n=1}^{\infty}$ is eventually increasing and bounded above. The other case is similar[43].

Suppose then that $N \in \mathbb{N}$ has the property that for $n \geq N$ we have $x_{n+1} \geq x_n$. Define $S = \{x_n \mid n \geq N\}$, and note that S is clearly non-empty. Moreover S is bounded above since the sequence $(x_n)_{n=1}^{\infty}$ is bounded above by assumption[44]. Therefore, by the completeness axiom[45], $\sup(S)$ exists. Write $L = \sup(S)$.

We claim that $x_n \to L$ as $n \to \infty$, so let $\varepsilon > 0$ be arbitrary. By the defining property of the supremum, there exists $n_0 \geq N$ such that $x_{n_0} > L - \varepsilon$. Therefore, for $n \geq n_0$, we have

$$L - \varepsilon < x_{n_0} \leq x_n \leq L < L + \varepsilon \implies |x_n - L| < \varepsilon,$$

as the sequence $(x_m)_{m=1}^{\infty}$ is increasing for $m \geq N$. Hence $x_n \to L$ as $n \to \infty$. □

Let's turn to some applications of the monotone convergence theorem. First note that it gives us a method for showing that sequences converge without having to find the proposed limit first: simply check that the sequence is eventually monotonic by, say, using the methods in Example 3.21) and bounded, then it will converge. After you know the sequence converges, you may be able to find the limit directly using other means, such as properties of limits. We will see an example of this below.

Example 3.23. Let $x \in \mathbb{R}$ and define $a_n = \frac{x^n}{n!}$.

a) For $x > 0$, show that $(a_n)_{n=1}^{\infty}$ is eventually decreasing.

b) For all $x \in \mathbb{R}$, show $a_n \to 0$.

Solution. a) When $x > 0$, we have $a_n > 0$ for all n. Then

$$\frac{a_{n+1}}{a_n} = \frac{x^{n+1}}{(n+1)!}\frac{n!}{x^n} = \frac{x}{n+1}.$$

Thus for $n > x - 1$, we have $a_{n+1} < a_n$. Hence $(a_n)_{n=1}^{\infty}$ is eventually decreasing.

When $x > 0$, we certainly have $a_n > 0$ for all n, so that $(a_n)_{n=1}^{\infty}$ is bounded below, and hence there exists $L \in \mathbb{R}$ with $a_n \to L$ as $n \to \infty$

[41] Note that $\frac{(n+1)!}{n!} = n+1$, but be careful: $\frac{(2(n+1))!}{(2n)!} = (2n+2)(2n+1)$; think about where you are putting brackets when you're using factorials.

[42] It is in fact equivalent to the completeness axiom: an ordered field satisfying the axioms (a)-(l) listed in Axioms 2.1 and 2.2 which satisfies the monotone convergence theorem has the property that suprema exist for every non-empty subset which is bounded above. For more on this theme see Körner's excellent book ''A companion to analysis: a first second course and second first course in analysis".

[43] You can also note that if $(x_n)_{n=1}^{\infty}$ is eventually decreasing and bounded below, then defining $y_n = -x_n$, we obtain an eventually increasing sequence $(y_n)_{n=1}^{\infty}$ which is bounded above. Convergence of $(y_n)_{n=1}^{\infty}$ implies convergence of $(x_n)_{n=1}^{\infty}$ by algebraic properties of limits.

[44] To prove that $(x_n)_{n=1}^{\infty}$ converges we will have to identify the proposed limit, and the natural candidate is the supremum of the set S. Draw a picture to persuade yourself that x_n should converge to the supremum of S.

[45] From a writing style view point, note how I check both conditions in the supremum axiom explicitly before using it.

by the monotone convergence theorem. Now note that $a_{n+1} = \frac{x}{n+1}a_n$. Since $a_{n+1} \to L$ as $n \to \infty$ and $\frac{x}{n+1} \to 0$ as $n \to \infty$, uniqueness of limits gives

$$L = 0 \times L = 0,$$

as required[46].

b) Note that if $x = 0$, then $a_n = 0$ for all n, when the result is immediate, while if $x < 0$, then $|a_n| = \frac{|x|^n}{n!} \to 0$, by part *a*), and hence $a_n \to 0$. □

[46] Note that $a_{n+1} = \frac{x}{n+1}a_n$ is a *recursion relation* defining a_{n+1} in terms of n, a_n, and the fixed constant x.

Note the technique of *taking limits in the recursion relation* $a_{n+1} = \frac{x}{n+1}a_n$ above. We see this technique again in the next example.

Example 3.24. Let $\alpha \in (2, 3)$ and define a sequence $(x_n)_{n=1}^{\infty}$ recursively by

$$x_1 = \alpha, \quad x_{n+1} = x_n^2 - 4x_n + 6, \quad n \in \mathbb{N}.$$

Show

a) $x_n \in (2, 3)$ for all n;

b) $(x_n)_{n=1}^{\infty}$ is decreasing;

c) $x_n \to 2$ as $n \to \infty$.

Solution. *a*) Since the sequence is defined recursively, it makes sense to prove *a*) by induction[47].

To carry out the proof by induction[48], we start by observing that $x_1 \in (2, 3)$ according to the assumptions. Assuming inductively that $x_n \in (2, 3)$ for some $n \in \mathbb{N}$, we have

$$x_{n+1} - 2 = x_n^2 - 4x_n + 4 = (x_n - 2)^2 > 0$$

and

$$x_{n+1} - 3 = x_n^2 - 4x_n + 3 = (x_n - 3)(x_n - 1) < 0,$$

as $x_n > 1$ and $x_n < 3$. Therefore, by induction, $x_n \in (2, 3)$ for all $n \in \mathbb{N}$.

b) We note that the recursion formula won't give us that much useful information about the ratio $\frac{x_{n+1}}{x_n}$, so we look at $x_{n+1} - x_n$ to see the sequence is decreasing. For $n \in \mathbb{N}$, we have

$$x_{n+1} - x_n = x_n^2 - 5x_n + 6 = (x_n - 2)(x_n - 3) < 0,$$

as $2 < x_n < 3$. Therefore $(x_n)_{n=1}^{\infty}$ is decreasing.

c) Finally, we use the monotone convergence theorem to learn that $(x_n)_{n=1}^{\infty}$ converges, then take limits in the recursion formula $x_{n+1} = x_n^2 - 4x_n + 6$ to find this limit. Since $(x_n)_{n=1}^{\infty}$ is decreasing and bounded below, it converges by the monotone convergence theorem to some number $L \in \mathbb{R}$. As $2 < x_n \leq x_1 < 3$ for all n, we have $2 \leq L \leq x_1 < 3$ by the order properties of limits. On the other hand, as $x_{n+1} \to L$ as $n \to \infty$, taking limits in the recursion formula gives $L = L^2 - 4L + 6$, so $L^2 - 5L + 6 = 0$, and hence $L = 2$ or $L = 3$. Since we have already noted that $L < 3$, we must have $L = 2$. □

[47] We are told that $x_1 = \alpha \in (2, 3)$, and have a formula for x_{n+1} in terms of x_n.

[48] Note that I'll not set my induction proof out by introducing the statement $P(n)$ to be $x_n \in (2, 3)$, and then using a dummy variable k. Setting out inductions using $P(n)$ and a dummy variable k is perfectly valid, though rather cumbersome — it's essentially designed to prevent people writing "let $n = n + 1$" which can't be true (but is a very popular thing to write in proofs by induction). Feel free to use which ever style you prefer when writing down a mathematical induction, provided your proof is correct!

Subsequences

Recall that a sequence is an infinite list of numbers, for example

$$1, 3, 5, 2, 7, 9, 15, 16, 18, 25, \ldots.$$

Subsequences of this sequence are obtained by deleting some (possibly infinitely many) of the numbers in the list, to obtain a new infinite list of numbers[49]. For example,

$$1, \cancel{3}, 5, 2, \cancel{7}, \cancel{9}, 15, 16, 18, \cancel{25}, \ldots$$

gives a subsequence of the sequence above.

Of course, we should give a precise definition of what a subsequence is, and we shall prepare for this as follows. Each term in the subsequence corresponds to a term in the original sequence. In the example above, 1 is the first term in the subsequence and the first term of the sequence; 5 is the second term of the subsequence and the third term of the sequence; 2 the third term of the subsequence and the fourth term of the sequence, and so forth.

Let us now write our original sequence more formally as $(x_n)_{n=1}^{\infty}$, and our subsequence as $(y_n)_{n=1}^{\infty}$. For each $n \in \mathbb{N}$, the n-th term y_n of the subsequence must be some term in the original sequence, say the k_n-th term in the original sequence. So we should have $y_n = x_{k_n}$ for some $k_n \in \mathbb{N}$. What properties must the indices k_n have? For each n, the $n+1$-st term in the subsequence $x_{k_{n+1}}$ must be located further along the original sequence than the n-th term x_{k_n}. That is, we must have $k_{n+1} > k_n$ for all $n \in \mathbb{N}$. We use this to make the following precise definition.

Definition 3.25. A *subsequence* of a sequence $(x_n)_{n=1}^{\infty}$ is a sequence of the form $(x_{k_n})_{n=1}^{\infty}$ for some strictly increasing natural numbers $k_1 < k_2 < k_3 < \ldots$.

Note that for strictly increasing natural numbers $k_1 < k_2 < k_3 < \ldots$ we have[50] $k_n \geq n$ for each $n \in \mathbb{N}$. In the example above, we have $k_1 = 1, k_2 = 3, k_3 = 4$, and so on.

There are a number of standard subsequences. Letting $(x_n)_{n=1}^{\infty}$ be a sequence, we have the subsequence $(x_{2n})_{n=1}^{\infty}$ consisting of the even terms of $(x_n)_{n=1}^{\infty}$; similarly, $(x_{2n-1})_{n=1}^{\infty}$ is the subsequence consisting of the odd terms of $(x_n)_{n=1}^{\infty}$; we also have a shifted sequence $(x_{n+1})_{n=1}^{\infty}$ which is obtained by removing the first term. You should be able to work out descriptions for other subsequences: how would you write down the subsequence which starts with the second term of $(x_n)_{n=1}^{\infty}$ and takes every fifth subsequence term? Finally, note that by taking $k_n = n$ for all $n \in \mathbb{N}$, we see that $(x_n)_{n=1}^{\infty}$ is a subsequence of itself.

Thinking about what convergence means, it should be clear that we expect that if $(x_n)_{n=1}^{\infty}$ converges to L, then every subsequence of $(x_n)_{n=1}^{\infty}$ converges too[51].

Proposition 3.26. *Let $(x_n)_{n=1}^{\infty}$ be a sequence and $L \in \mathbb{R}$. Then $x_n \to L$ if and only if every subsequence of $(x_n)_{n=1}^{\infty}$ converges to L.*

[49] So, deleting *all* the numbers is not allowed, and neither is deleting all the numbers after some fixed point, for instance.

[50] A formal proof of this statement can be given by induction.

[51] The converse is also true, and actually devoid of content: every sequence is a subsequence of itself.

Proof. As every sequence is a subsequence of itself the implication from right to left is immediate. For the implication from left to right, let $(x_n)_{n=1}^\infty$ be a sequence with $x_n \to L$ as $n \to \infty$. Let $k_1 < k_2 < \dots$ be an arbitrary strictly increasing sequence of natural numbers defining a subsequence $(x_{k_n})_{n=1}^\infty$. Let $\varepsilon > 0$ be arbitrary, so by definition of convergence, there exists $n_0 \in \mathbb{N}$ such that $n \geq n_0 \implies |x_n - L| < \varepsilon$. For $n \geq n_0$, we have $k_n \geq n \geq n_0$, and hence $|x_{k_n} - L| < \varepsilon$. That is, we have $n \geq n_0 \implies |x_{k_n} - L| < \varepsilon$. Hence we conclude that $(x_{k_n})_{n=1}^\infty$ converges to L as $n \to \infty$. □

The previous proposition provides a tool for quickly proving that various sequences do *not* converge: find either a non-convergent subsequence or two subsequences converging to different limits.

Example 3.27. Show that $(x_n)_{n=1}^\infty$ does not converge when $x_n = \frac{n-1}{n}(-1)^n$.

Solution. We consider the subsequences $(x_{2n})_{n=1}^\infty$ and $(x_{2n-1})_{n=1}^\infty$ of even and odd terms of $(x_n)_{n=1}^\infty$, respectively. Then we have

$$x_{2n} = \frac{2n-1}{2n} \to 1, \text{ and } x_{2n-1} = -\frac{2n-2}{2n-1} \to -1$$

as $n \to \infty$. As these two subsequences converge to different values, the sequence $(x_n)_{n=1}^\infty$ does not converge. □

The Bolzano–Weierstrass Theorem

I now want to turn to another consequence of the completeness axiom: the Bolzano–Weierstrass Theorem[52]. The key idea of the proof that we will study below is found in the following combinatorial lemma.

[52] Which, like the monotone convergence theorem is in fact equivalent to completeness.

Lemma 3.28 (A combinatorial lemma)**.** *Every real sequence has a monotone subsequence.*

Proof. Let $(x_n)_{n=1}^\infty$ be a real sequence. Define[53] a natural number n to be *farseeing*[54] if for all $m > n$, we have $x_m < x_n$.

[53] Definitions in proofs are normally made just for the purpose of the proof, and shouldn't be considered a general mathematical concept. That certainly applies here.

[54] The picture I imagine here is a series of towers, one at each natural number n, of height x_n. Then the condition that n is farseeing means that I can stand on the n-th tower, and when look to the right, my view is not obstructed by any subsequent towers.

Suppose that there are infinitely many farseeing natural numbers, and write $k_1 < k_2 < k_3 < \dots$ for these farseeing numbers. Then the subsequence $(x_{k_n})_{n=1}^\infty$ is (strictly) decreasing, as for each n we have $k_{n+1} > k_n$, so $x_{k_{n+1}} < x_{k_n}$ as k_n is farseeing.

Otherwise there are only finitely many farseeing natural numbers, so we can fix N such that for all $n \geq N$, the natural number n is not farseeing. We will inductively construct an increasing subsequence of $(x_n)_{n=1}^\infty$, say $(x_{k_n})_{n=1}^\infty$ by specifying a suitable strictly increasing sequence of naturals $k_1 < k_2 < k_3 < \dots$. Set $k_1 = N$. Suppose inductively that for some $n \in \mathbb{N}$, we have defined natural numbers $N = k_1 < k_2 < k_3 < \dots < k_n$ such that $x_{k_1} \leq x_{k_2} \leq \dots \leq x_{k_n}$. Then, as $k_n \geq N$, the number k_n is not farseeing. By definition, this means that there exists a natural number $k_{n+1} > k_n$ such that $x_{k_{n+1}} \geq x_{k_n}$, and so we obtain the required increasing subsequence of $(x_n)_{n=1}^\infty$ by induction. □

Theorem 3.29 (Bolzano–Weierstrass)**.** *Every bounded real sequence has a convergent subsequence.*

Proof. Let $(x_n)_{n=1}^{\infty}$ be a bounded real sequence. By the combinatorial lemma above, there exists a monotonic subsequence $(x_{k_n})_{n=1}^{\infty}$ of $(x_n)_{n=1}^{\infty}$. Since this subsequence is bounded it converges by the monotone convergence theorem. $\square$

We will see one quick application of Bolzano–Weierstrass in the following section, and then again right at the end of the course.

Cauchy Sequences

The last topic in this section is of Cauchy sequences, and this is really laying the scene for the 3H course Metric spaces and Basic Topology next year. In that course you will learn about an abstract setting for the study of convergence, just as the earlier courses this year developed the abstract notion of a vector space as a general setting for the study of vectors and linear maps. In order to define the notion of completeness in this setting one has to use properties which can be defined only using the distance function[55] of a metric space[56]. Cauchy sequences are the tool that will be used to do this.

Let us start with the formal definition.

Definition 3.30. Let $(x_n)_{n=1}^{\infty}$ be a sequence. We say that $(x_n)_{n=1}^{\infty}$ is *Cauchy* if

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N} \text{ s.t. } \forall (n, m \in \mathbb{N} \text{ with } n, m \geq n_0), |x_n - x_m| < \varepsilon.$$

You should think of the condition of being Cauchy as meaning that as n gets large, all the terms of the sequence get arbitrarily close together[57]. I'd encourage you to spend time unpacking this definition in just the same way I did with the definition of convergence earlier in the section to try to see why I think of Cauchy sequences in this way.

The process of verifying that a certain sequence $(x_n)_{n=1}^{\infty}$ is Cauchy directly from the definition is very much like verifying that a sequence converges to L directly from the definition. First, introduce the symbol ε with "Let $\varepsilon > 0$ be arbitrary". You will then need to manipulate the inequality $|x_n - x_m| < \varepsilon$, and see if you can find a suitably large n_0 which will guarantee that $|x_n - x_m| < \varepsilon$ for $n, m \geq n_0$. When you do this, you can always assume say that $m \geq n$ as the other case is handled by symmetry. More precisely, the statement

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N} \text{ s.t. } \forall (n, m \in \mathbb{N} \text{ with } m \geq n \geq n_0), |x_n - x_m| < \varepsilon.$$

is equivalent to $(x_n)_{n=1}^{\infty}$ is Cauchy.

Example. *Show directly from the definition that the sequence $(x_n)_{n=1}^{\infty}$ given by $x_n = \frac{n+1}{n+2}$ is Cauchy.*

[55] In the case of the real numbers the distance function is $d(x, y) = |x - y|$.

[56] The completeness axiom we use for the real numbers is defined in terms of the order structure, and this will not be an ingredient in the theory of metric spaces: indeed, there is no total ordering on $\mathbb{C}$ which is compatible with the algebraic axioms — still the complex numbers are a perfectly nice example of a metric space.

[57] It's important to note that this is not the same as saying that eventually x_n gets arbitrarily close to x_{n+1}. The statement

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N} \text{ s.t. } \forall (n \in \mathbb{N} \text{ with } n \geq n_0), |x_n - x_{n+1}| < \varepsilon$$

is *not* equivalent to $(x_n)_{n=1}^{\infty}$ being Cauchy: a counterexample is $x_n = \sum_{r=1}^{n} \frac{1}{r}$, as we'll see in the next chapter on series.

Solution. Let $\varepsilon > 0$ be arbitrary. For natural numbers $m \geq n$, we have

$$\begin{aligned} |x_m - x_n| &= \left| \frac{m+1}{m+2} - \frac{n+1}{n+2} \right| \\ &= \left| \frac{(m+1)(n+2) - (n+1)(m+2)}{(m+2)(n+2)} \right| = \left| \frac{m-n}{(m+2)(n+2)} \right| \end{aligned}$$

Now, we estimate[58]

$$\left| \frac{m-n}{(m+2)(n+2)} \right| \leq \frac{m}{mn} = \frac{1}{n},$$

so if we take $n_0 \in \mathbb{N}$ such that $n_0 > \frac{1}{\varepsilon}$, it follows that if $m \geq n \geq n_0$, then $|x_m - x_n| < \varepsilon$. Since $\varepsilon > 0$ was arbitrary, it follows that $(x_n)_{n=1}^{\infty}$ is Cauchy. □

[58] I'm no longer trying to keep track of $|x_m - x_n|$ exactly, but am happy to make this quantity larger, and simpler to work with, provided I can bound the simpler quantity by ε. Thus I use the estimates $|m - n| = m - n \leq m$ as $m \geq n$ and $(m+2)(n+2) \geq mn$. Other estimates would work here.

Now we turn to our main theoretical result, the general principle of convergence. As a preparation, we need two propositions which I'll leave as exercises. The first is a good exercise in making sure you understand the definitions of convergence and Cauchy and can manipulate these definitions[59]. For the second, look at the proof of Theorem 3.9 and make sure you get a bound which does not depend on n or m. We note that neither result uses the completeness axiom.

[59] As a hint, you'll probably want to consider $\frac{\varepsilon}{2}$ in the definition of convergence for the first one.

Proposition 3.31. *Every convergent sequence is Cauchy.*

Proposition 3.32. *Every Cauchy sequence is bounded.*

Now we reach the general principle of convergence, which says that convergent sequences and Cauchy sequences are actually the same thing. This will be used in later courses to exhibit elements of metric spaces with certain properties as limits of Cauchy sequences[60]. The key strategy in the hard direction of the proof is to use Bolzano–Weierstass to obtain a convergent subsequence of a Cauchy sequence. We then show that the entire sequence converges to the same limit as the convergent subsequence[61].

[60] For example, we can show that there exist solutions to various differential equations in this way.

[61] This should not come as a surprise: large terms in the subsequence get close to the limit, and large terms in the sequence get close together, suggesting that all large terms in the sequence should be close to the limit. You can see from that sentence that we're estimating the distance in two steps, so we use an "$\frac{\varepsilon}{2}$ argument" in our proof.

Theorem 3.33 (The general principle of convergence)**.** *Let $(x_n)_{n=1}^{\infty}$ be a real sequence. Then $(x_n)_{n=1}^{\infty}$ is Cauchy if and only if it converges.*

Proof. If $(x_n)_{n=1}^{\infty}$ converges, then it is Cauchy by Proposition 3.31. Conversely, let $(x_n)_{n=1}^{\infty}$ be Cauchy, so that $(x_n)_{n=1}^{\infty}$ is bounded by Proposition 3.32. By the Bolzano–Weierstrass Theorem, $(x_n)_{n=1}^{\infty}$ has a subsequence, say $(x_{k_n})_{n=1}^{\infty}$ converging to $L \in \mathbb{R}$, say (where $k_1 < k_2 < \ldots$ is a strictly increasing sequence of natural numbers, and so satisfy $n \leq k_n$ for all n).

We claim that $x_n \to L$ as $n \to \infty$, so let $\varepsilon > 0$ be arbitrary. As $(x_n)_{n=1}^{\infty}$ is Cauchy, there exists $n_0 \in \mathbb{N}$ such that $m, n \geq n_0 \implies |x_n - x_m| < \frac{\varepsilon}{2}$. Since $(x_{k_n})_{n=1}^{\infty}$ converges to L, there exists $n_1 \in \mathbb{N}$ such that $n \geq n_1 \implies |x_{k_n} - L| < \frac{\varepsilon}{2}$. Now take $n \geq n_0$, and let $m = \max(n_0, n_1)$, so that $m \geq n_1$. Then $|x_{k_m} - L| < \frac{\varepsilon}{2}$, and $k_m \geq m \geq n_0$, hence $|x_{k_m} - x_n| < \frac{\varepsilon}{2}$. Therefore

$$|x_n - L| < |x_n - x_{k_m}| + |x_{k_m} - L| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

As $\varepsilon > 0$ was arbitrary, we conclude $x_n \to L$, as claimed. □

As a final note, let us remark that the general principle of convergence is not quite equivalent to the completeness axiom[62]. The completeness axiom is equivalent to the statement that both Archimedes Axiom[63] and the general principle of convergence holds.

[62] unlike the monotone convergence theorem, and the Bolzano–Weierstrass Theorem which are.

[63] which says that the natural numbers are not bounded above.

2C Intro to real analysis 2020/21

Handout

Series

In this chapter we will study *series* of real numbers. This is the mathematical framework in which we can make precise the idea of *infinite sums*, like

$$1+\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\cdots.$$

Let's start with the following definition.

Definition 4.1. Let $(a_n)_{n=1}^{\infty}$ be a sequence of real numbers. The (infinite) series generated by $(a_n)_{n=1}^{\infty}$ is the sequence $(s_n)_{n=1}^{\infty}$ where

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

The number s_n is called the n-th partial sum of the series. We also use the notation

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

when referring to the sequence $(s_n)_{n=1}^{\infty}$ of partial sums.

In other words, series are a certain kind of sequences, obtained by summing the terms of a sequence. In particular, series are related to sequences in two ways. On the one hand, to define the series $\sum_{n=1}^{\infty} a_n$, we need the sequence $(a_n)_{n=1}^{\infty}$ as input. On the other hand, the series $\sum_{n=1}^{\infty} a_n$ itself is nothing but a sequence in disguise. By definition, it is the sequence $(s_n)_{n=1}^{\infty}$ of partial sums s_n.

In order to distinguish the sequences $(a_n)_{n=1}^{\infty}$ and $(s_n)_{n=1}^{\infty}$, the notation $\sum_{j=1}^{\infty} a_j$ introduced above as an abbreviation for $(s_n)_{n=1}^{\infty}$ is useful[1]. At this point, we should not interpret the symbol $\sum_{j=1}^{\infty} a_j$ as a real number, given by an actual (infinite) sum. This will only be done once we have discussed convergence of sequences below[2].

Let us examine in more detail the relationship between the sequences $(a_n)_{n=1}^{\infty}$ and $(s_n)_{n=1}^{\infty}$, that is, between a sequence $(a_n)_{n=1}^{\infty}$ and its associated series. By construction, the sequence $(s_n)_{n=1}^{\infty}$ is obtained from $(a_n)_{n=1}^{\infty}$ by summation. Conversely, one can entirely recover $(a_n)_{n=1}^{\infty}$ from $(s_n)_{n=1}^{\infty}$ since

$$\begin{aligned}
a_1 &= s_1 \\
a_2 &= (a_1+a_2)-a_1 = s_2-s_1 \\
a_3 &= (a_1+a_2+a_3)-(a_1+a_2) = s_3-s_2,
\end{aligned}$$

and more generally,

$$a_n = s_n - s_{n-1}$$

[1] Sometimes we use the shorthand notation $\sum a_j$ instead of $\sum_{j=1}^{\infty} a_j$.

[2] Unfortunately, the standard mathematical notation is ambiguous in this regard. Try not to get confused by this!

for any $n > 1$. In other words, the information contained in the sequences $(a_n)_{n=1}^{\infty}$ and $(s_n)_{n=1}^{\infty}$ is equivalent.

So why do we bother about series at all, and not just continue working with sequences? The main reason is that many familiar functions occur naturally in the form of series, and are best understood from this point of view[3]. Still, one should keep in mind that no information is lost by switching back and forth between series and sequences. When proving results about series, we will often use this to reduce matters to known results about sequences from chapter 3.

We will be flexible regarding the starting term of our series. For instance, it is sometimes useful to consider series of the form $\sum_{n=0}^{\infty} a_n$, starting at $n = 0$ instead of $n = 1$, or at other values.

[3] For instance, to rigorously define and study the exponential function exp and trigonometric functions like sin and cos, one needs *power series*. This will be explained in the 3H course Analysis of Differentiation and Integration.

Convergence of series

Since series are nothing but sequences in disguise, we define the notion of convergence for series as follows.

Definition 4.2. Let $\sum_{n=1}^{\infty} x_n$ be a series of real numbers and let $L \in \mathbb{R}$. We say that the series $\sum_{n=1}^{\infty} x_n$ converges towards $L \in \mathbb{R}$ if the sequence $(s_n)_{n=1}^{\infty}$ of partial sums $s_n = \sum_{j=1}^{n} x_n$ converges towards L as n tends to infinity. In this case we also write $L = \sum_{n=1}^{\infty} x_n$ for the corresponding limit[4], and call it the *sum of the series*.
A series is called *divergent* if it does not converge to any limit.

[4] Here the symbol $\sum_{n=1}^{\infty} x_n$ stands for a real number, in contrast to the usage of this symbol as an abbreviation for the sequence of partial sums.

This definition encapsulates the intuitive idea that the limit L of the sequence of partial sums $s_n = x_1 + \cdots + x_n$ of a series is the "sum to infinity" of the terms x_n.

Let us point out again that we use the symbol $\sum_{n=1}^{\infty} x_n$ ambiguously to mean both the series, which always exists, and its sum, which may not exist.

In a similar fashion as we discussed convergence of sequences, Definition 4.2 can be written concisely as a quantified statement: the series $\sum_{n=1}^{\infty} x_n$ converges to $L \in \mathbb{R}$ if and only if

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N} \text{ s.t. } \forall n \in \mathbb{N}, \left(n \geq n_0 \implies \left| \sum_{j=1}^{n} x_j - L \right| < \varepsilon \right).$$

Equivalently, we may write this condition as

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N} \text{ s.t. } \forall (n \in \mathbb{N} \text{ with } n \geq n_0), \left| \sum_{j=1}^{n} x_j - L \right| < \varepsilon.$$

Notice that leaving out the first few terms in a series does not affect convergence, but it will usually affect the sum of a series[5].

[5] This should be compared with the situation for sequences: leaving out the first few terms in a convergent sequence $(x_n)_{n=1}^{\infty}$ neither affects convergence, nor the value of the limit $\lim_{n\to\infty} x_n$ of the sequence.

Properties of series

In this section we examine basic properties of series, and how convergent series interact with the algebraic operations on the real numbers. These rules provide methods for establishing convergence or divergence, and also tools for calculating sums of convergent series.

Theorem 4.3. *If a series $\sum_{n=1}^{\infty} x_n$ converges, then the sequence $(x_n)_{n=1}^{\infty}$ converges to 0, that is, $x_n \to 0$ as $n \to \infty$.*

Proof. Let $(s_n)_{n=1}^{\infty}$ the sequence of partial sums of the series $\sum_{n=1}^{\infty} x_n$. By assumption this sequence converges to S for some $S \in \mathbb{R}$. Then also the sequence[6] $(t_n)_{n=2}^{\infty}$ given by $t_n = s_{n-1}$ converges[7] to S. Using algebraic properties of limits, we deduce

$$x_n = s_n - s_{n-1} \to S - S = 0$$

as n tends to infinity. □

[6] Recall that we are flexible regarding the starting point of sequences. Here we start at $n = 2$ because $t_1 = s_0$ is not defined.

[7] For $\varepsilon > 0$ let $n_0 \in \mathbb{N}$ be chosen such that $|s_n - S| < \varepsilon$ for all $n \geq n_0$. Then we have $|t_n - S| = |s_{n-1} - S| < \varepsilon$ for all $n \geq n_0 + 1$. This means $t_n \to S$ as $n \to \infty$.

The contrapositive of the statement in Theorem 4.3 can be formulated as follows: if $a_n \nrightarrow 0$ then $\sum_{n=1}^{\infty} a_n$ diverges. That is, the previous theorem provides us with a useful criterion to show that series do *not* converge: if the coefficient sequence fails to converge to 0, the series diverges.

Example 4.4. Show that the series

$$\sum_{n=1}^{\infty} \frac{3n+5}{7n+2}$$

diverges.

Solution. According to Theorem 4.3 it suffices to show that the sequence $(a_n)_{n=1}^{\infty}$ with $a_n = \frac{3n+5}{7n+2}$ does not converge to zero. By algebraic properties of sequence limits [8], we have

$$\frac{3n+5}{7n+2} \to \frac{3}{7} \neq 0.$$

Therefore[9], the series $\sum_{n=1}^{\infty} \frac{3n+5}{7n+2}$ diverges. □

[8] see Theorem 3.10 in Chapter 3.

[9] Alternatively, we could argue that

$$\frac{3n+5}{7n+2} \geq \frac{3n}{7n+2n} = \frac{3}{9}$$

for all $n \in \mathbb{N}$, which is enough to show that $(a_n)_{n=1}^{\infty}$ does not converge to zero.

We now come to an important example of a series.

Proposition 4.5. *The* geometric series

$$\sum_{n=0}^{\infty} x^n = 1 + x + x^2 + x^3 + \cdots$$

converges if and only if $|x| < 1$. For $|x| < 1$, the sum of the series is $\frac{1}{1-x}$.

Proof. Firstly, the geometric series $\sum_{n=0}^{\infty} x^n$ for $x = 1$ and $x = -1$ clearly diverges[10]. Now for any $x \in \mathbb{R}$ with $x \neq 1$ we have

$$\begin{aligned}(1-x)\sum_{j=0}^{n} x^j &= (1-x)(1+x+x^2+\cdots+x^n) \\ &= (1+x+x^2+\cdots+x^n) - (x+x^2\cdots - x^{n+1}) \\ &= 1 - x^{n+1},\end{aligned}$$

or equivalently,

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

Moreover $x^n \to 0$ as $n \to \infty$ if and only if $|x| < 1$ by one of our standard limits, see Theorem 3.18. This has two consequences: firstly, for $|x| < 1$ the geometric series is convergent with sum $\sum_{j=0}^{n} x^n = 1/(1-x)$. Secondly, for $|x| > 1$ the series $\sum_{j=0}^{n} x^n$ diverges according to Theorem 4.3 since the sequence $a_n = x^n$ is unbounded[11]. □

[10] For $x = 1$, the geometric series has partial sums $s_n = n$, so is unbounded and therefore divergent. For $x = -1$, we have

$$s_n = \frac{1}{2}(1 - (-1)^{n+1}),$$

which is a divergent sequence since the subsequences $s_{2n} = 1$ and $s_{2n+1} = 0$ converge to different limits.

[11] and therefore in particular not converging to zero.

Proposition 4.5 shows in particular that, say, the series $\sum_{n=0}^{\infty} \frac{1}{2^n}$ is convergent. This may come as a surprise at first sight: essentially, we are adding *infinitely many strictly positive terms*, and nonetheless get a *finite* total sum[12].

Let us next discuss general properties of convergent series, and in particular how sums of series interact with the algebraic operations of addition, multiplication, and the order structure of the real numbers.

[12] This fact puzzled already famous ancient Greek philosophers. Indeed, the series $\sum_{n=0}^{\infty} \frac{1}{2^n}$ plays a central role in one version of *Zeno's paradox*. If you're curious to learn more about this, try to search on the internet!

Theorem 4.6 (Properties of convergent series). *Let $\sum_{n=1}^{\infty} x_n$ and $\sum_{n=1}^{\infty} y_n$ be series with sums S and T respectively.*

a) For $\lambda \in \mathbb{R}$, we have $\sum_{n=1}^{\infty}(\lambda x_n) = \lambda S$;

b) The series $\sum_{n=1}^{\infty}(x_n + y_n)$ converges with sum $S + T$, that is,

$$\sum_{n=1}^{\infty}(x_n + y_n) = \sum_{n=1}^{\infty} x_n + \sum_{n=1}^{\infty} y_n.$$

c) If $x_n \leq y_n$ for all $n \in \mathbb{N}$, then $S \leq T$, that is, $\sum_{n=1}^{\infty} x_n \leq \sum_{n=1}^{\infty} y_n$.

d) If $\sum_{n=1}^{\infty} |x_n|$ also converges with sum U, then $|S| \leq U$, that is,

$$|\sum_{n=1}^{\infty} x_n| \leq \sum_{n=1}^{\infty} |x_n|.$$

I'm going to prove only the first part, and encourage you to prove the remaining three statements. The strategy is similar to the ideas used to establish algebraic properties of convergent sequences.

Proof of a). Let $\sum_{n=1}^{\infty} x_n$ be a convergent series with sum S. That is, the sequences of partial sums $(s_n)_{n=1}^{\infty}$, given by $s_n = x_1 + \cdots + x_n$, converges. The n-th partial sum for the series $\sum_{n=1}^{\infty} \lambda x_n$ is

$$\sum_{j=1}^{n} \lambda x_j = \lambda \sum_{j=1}^{n} x_j = \lambda s_n,$$

so by algebraic properties of limits[13], it is a convergent sequence. By definition, this means that the series $\sum_{n=1}^{\infty}(\lambda x_n)$ converges, with sum λS. □

[13] for convergent sequences, compare Theorem 3.10 in Chapter 3.

Note that the key idea in this proof was to reduce the claim to a familiar statement about convergent sequences.

Let us now have a look at further examples of series. A prominent example of a divergent series is the *harmonic series*.

Proposition 4.7. *The harmonic series*

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

is divergent.

Proof. Let us write

$$s_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n}$$

for the n-th partial sum of the harmonic series. By way of contradiction, we shall assume that $\sum_{n=1}^{\infty} \frac{1}{n}$ is convergent, and write $S = \lim_{n\to\infty} s_n$ for its sum.

Consider in addition the series $\sum_{n=1}^{\infty} y_n$ where

$$y_n = \begin{cases} 1 & n = 1 \\ 1/2 & n = 2 \\ 1/(n+1) & n \geq 3 \text{ and odd} \\ 1/n & n \geq 4 \text{ and even} \end{cases}$$

We claim that our assumption implies that $\sum_{n=1}^{\infty} y_n$ converges as well. Firstly, if t_n denotes the n-th partial sum of the series $\sum_{n=1}^{\infty} y_n$, then we have

$$\begin{aligned} t_{2n} &= 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{4} + \frac{1}{6} + \frac{1}{6} + \cdots + \frac{1}{2n} + \frac{1}{2n} \\ &= \frac{1}{2} + 1 + \left(\frac{1}{4} + \frac{1}{4}\right) + \left(\frac{1}{6} + \frac{1}{6}\right) + \cdots + \left(\frac{1}{2n} + \frac{1}{2n}\right) \\ &= \frac{1}{2} + 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} \\ &= \frac{1}{2} + s_n \end{aligned}$$

for all $n \in \mathbb{N}$. Since the strictly increasing sequence $(s_n)_{n=1}^{\infty}$ is convergent by assumption, it is bounded above by its limit S. Hence the sequence $(t_{2n})_{n=1}^{\infty}$ is bounded above by $\frac{1}{2} + S$. Moreover, the sequence $(t_n)_{n=1}^{\infty}$ is clearly strictly increasing as well, which means that $\frac{1}{2} + S$ is in fact an upper bound[14] for all terms t_n.

According to the monotone convergence theorem, the sequence $(t_n)_{n=1}^{\infty}$ is therefore convergent. Moreover, its limit T must satisfy $T \leq \frac{1}{2} + S$. Using the formula $t_{2n} = \frac{1}{2} + s_n$ and properties of limits, we obtain in fact[15] $T = \frac{1}{2} + S$.

However, by the very construction of y_n we have $y_n \leq 1/n$ for all n, which implies $T \leq S$ according to Theorem 4.6 *c*). Hence $\frac{1}{2} + S \leq S$, which is a contradiction. □

[14] For any $n \in \mathbb{N}$, we have $t_n < t_{2n} \leq \frac{1}{2} + S$.

[15] Recall that if $(x_n)_{n=1}^{\infty}$ is a convergent sequence, then also every subsequence of $(x_n)_{n=1}^{\infty}$ is convergent with the same limit. In particular, the subsequence $(t_{2n})_{n=1}^{\infty}$ of even terms of the sequence $(t_n)_{n=1}^{\infty}$ converges to T.

How can we decide whether an infinite series converges or not? In the following paragraphs we discuss several methods to answer this question.

Tests for convergence I: the comparison test

Let us start with the comparison test.

Theorem 4.8 (Comparison Test). *Let $(x_n)_{n=1}^{\infty}$ and $(y_n)_{n=1}^{\infty}$ be sequences, and assume that there exists an $N \in \mathbb{N}$ such that $0 \leq x_n \leq y_n$ for all $n \geq N$. Then*

$$\sum_{n=1}^{\infty} y_n \text{ converges} \implies \sum_{n=1}^{\infty} x_n \text{ converges,}$$

$$\sum_{n=1}^{\infty} x_n \text{ diverges} \implies \sum_{n=1}^{\infty} y_n \text{ diverges.}$$

Proof. The second statement is the contrapositive of the first, so it's enough to prove the first statement.

Since convergence of a series does not depend on the first N terms, we may assume $N = 1$. Let's write s_n and t_n for the n-th partial sums of $\sum_{n=1}^{\infty} x_n$ and $\sum_{n=1}^{\infty} y_n$, respectively. Then our assumption gives $0 \leq s_n \leq t_n$ for all $n \in \mathbb{N}$. Since (t_n) converges it is bounded above by M. Therefore $(s_n)_{n=1}^{\infty}$ is bounded as well. Moreover, we have

$$s_{n+1} = s_n + x_{n+1} \geq s_n,$$

so that $(s_n)_{n=1}^{\infty}$ is increasing. By the monotone convergence theorem, the sequence $(s_n)_{n=1}^{\infty}$ converges. That is, $\sum_{n=1}^{\infty} x_n$ is convergent. □

As a consequence of Theorem 4.8 we obtain the following variant of the comparison test.

Corollary 4.9 (Limit version of comparison test)**.** Let $(x_n)_{n=1}^{\infty}$ and $(y_n)_{n=1}^{\infty}$ be sequences of eventually positive[16] terms, and suppose that $x_n/y_n \to L$ as $n \to \infty$ for some $L \in (0, \infty)$. Then

$$\sum_{n=1}^{\infty} x_n \text{ converges} \iff \sum_{n=1}^{\infty} y_n \text{ converges.}$$

Proof. Since convergence of a series is not affected by the first few terms, we may assume that $x_n, y_n > 0$ for all n. By assumption, the sequence $\frac{x_n}{y_n}$ converges and hence is bounded above, say by M. Then we have $0 \leq \frac{x_n}{y_n} \leq M$, or equivalently $0 \leq x_n \leq My_n$ for all $n \in \mathbb{N}$. Thus convergence of $\sum_{n=1}^{\infty} y_n$ implies convergence of $\sum_{n=1}^{\infty} x_n$ by the comparison test.

Conversely, by algebraic properties of sequence limits we have $\frac{y_n}{x_n} \to L^{-1}$. In particular, the sequence $\frac{y_n}{x_n}$ is bounded above as well. In the same way as above, this shows that convergence of $\sum_{n=1}^{\infty} x_n$ implies convergence of $\sum_{n=1}^{\infty} y_n$. □

Let's turn to some applications of the comparison test. Note that it gives us a method for showing that a series converges without having to find the proposed limit first. After having established that the series converges you may be able to find the limit by other means, such as properties of limits.

Proposition 4.10. *Let $p \in \mathbb{N}$. Then[17] the series*

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

converges if and only if $p > 1$.

Proof. The case $p = 1$ corresponds to the harmonic series, which we already know to be divergent by Proposition 4.7.

Let us next consider $p = 2$ and define $x_n = \frac{1}{n(n+1)}$. Since

$$\begin{aligned} x_k = \frac{1}{k(k+1)} &= \frac{k+1}{k(k+1)} - \frac{k}{k(k+1)} \\ &= \frac{1}{k} - \frac{1}{k+1} \end{aligned}$$

[16] A sequence $(a_n)_{n=1}^{\infty}$ is called eventually positive if there exists $n_0 \in \mathbb{N}$ such that $a_n > 0$ for all $n \geq n_0$.

[17] This result holds in fact for any real number $p > 1$. Since we have not discussed how to rigorously define exponentials n^p for $p \in \mathbb{R}$ we restrict ourselves to the case $p \in \mathbb{N}$.

for all $k \in \mathbb{N}$ we see that

$$x_1 + x_2 + \cdots + x_n = \left(\frac{1}{1} - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \cdots + \left(\frac{1}{n} - \frac{1}{n+1}\right)$$
$$= 1 - \frac{1}{n+1}.$$

It follows that $\sum_{n=1}^{\infty} x_n$ is convergent with limit

$$\sum_{n=1}^{\infty} x_n = \lim_{n\to\infty} \left(1 - \frac{1}{n+1}\right) = 1.$$

Since $\frac{1}{(n+1)^2} < \frac{1}{n(n+1)}$ for all $n \in \mathbb{N}$, the comparison test shows that the series $\sum_{n=1} \frac{1}{(n+1)^2}$ is convergent. We conclude that

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

is convergent as well[18].

[18] If you're unsure why the last equality holds remind yourself that both sides of the equation are defined as limits of certain sequences of partial sums. Compare the partial sums on both sides and then take limits.

Let us now assume $p \geq 3$ is arbitrary. Since in this case $\frac{1}{n^p} \leq \frac{1}{n^2}$, the comparison test shows that $\sum_{n=1}^{\infty} \frac{1}{n^p}$ is convergent as well. □

Here are some further applications of the comparison test.

Example 4.11. Discuss the convergence of

$$\sum_{n=1}^{\infty} \frac{n^2 - 5n + 1}{n^4 + n^2 + 1}.$$

Solution. For $n \geq 5$ we have

$$0 \leq \frac{n^2 - 5n + 1}{n^4 + n^2 + 1} \leq \frac{n^2 + 1}{n^4} \leq \frac{n^2 + n^2}{n^4} = \frac{2}{n^2}.$$

Using the comparison test and Proposition 4.10 we conclude that $\sum_{n=1}^{\infty} \frac{n^2-5n+1}{n^4+n^2+1}$ is convergent.

Alternatively, set $x_n = \frac{n^2-5n+1}{n^4+n^2+1}$ and $y_n = \frac{1}{n^2}$. Then the terms x_n and y_n are eventually positive, and we have

$$\frac{x_n}{y_n} = n^2 \cdot \frac{n^2 - 5n + 1}{n^4 + n^2 + 1} = \frac{n^4 - 5n^3 + n^2}{n^4 + n^2 + 1} \to 1$$

as $n \to \infty$. Since $\sum_{n=1}^{\infty} \frac{1}{n^2}$ converges by Proposition 4.10, the limit version of the comparison test shows that $\sum_{n=1}^{\infty} \frac{n^2-5n+1}{n^4+n^2+1}$ is convergent. □

Tests for convergence II: the ratio test

A useful test for convergence is obtained by comparison with the geometric series. More precisely, we have the following result.

Theorem 4.12 (Ratio Test)**.** *Let $(x_n)_{n=1}^{\infty}$ be a sequence.*

a) Assume there exists $\lambda < 1$ and $n_0 \in \mathbb{N}$ such that $x_n > 0$ and $\frac{x_{n+1}}{x_n} \leq \lambda$ for all $n \geq n_0$. Then $\sum_{n=1}^{\infty} x_n$ converges.

b) *Assume there exists* $n_0 \in \mathbb{N}$ *such that* $x_n > 0$ *and* $\frac{x_{n+1}}{x_n} \geq 1$ *for all* $n \geq n_0$. *Then* $\sum_{n=1}^{\infty} x_n$ *diverges.*

Proof. a) Since convergence of a series is not affected by the first few terms, we may assume that $n_0 = 1$. Then for $n \in \mathbb{N}$ we have

$$x_{n+1} \leq \lambda x_n \leq \lambda^2 x_{n-1} \leq \cdots \leq \lambda^n x_1.$$

That is, we have $0 < x_n \leq x_1 \lambda^{n-1}$ for every $n \in \mathbb{N}$. Since the geometric series $\sum_{n=1}^{\infty} \lambda^{n-1}$ converges, the same holds for $\sum_{n=1}^{\infty} \lambda^{n-1} x_1$, and hence $\sum_{n=1}^{\infty} x_n$ converges by the comparison test.

b) Again we may assume $n_0 = 1$. If $\frac{x_{n+1}}{x_n} \geq 1$ for all n we have

$$x_{n+1} \geq x_n \geq \cdots \geq x_1 > 0$$

It follows that the sequence $(x_n)_{n=1}^{\infty}$ does not converge to zero. Therefore, $\sum_{n=1}^{\infty} x_n$ is divergent by Theorem 4.3. □

Note that in order to apply the ratio test for showing convergence it is *not* sufficient to have $\frac{x_{n+1}}{x_n} < 1$ for all $n \geq n_0$. Indeed, we have $\frac{x_{n+1}}{x_n} < 1$ for $x_n = 1/n$, but according to Proposition 4.7 the harmonic series $\sum_{n=1}^{\infty} 1/n$ diverges.

Example 4.13. Show that

$$\sum_{n=1}^{\infty} \frac{n}{3^n}$$

converges.

Solution. Writing $x_n = \frac{n}{3^n}$ we have

$$\frac{x_{n+1}}{x_n} = \frac{(n+1)3^n}{3^{n+1}n} = \frac{n+1}{3n} \leq \frac{2}{3} < 1.$$

for all $n \in \mathbb{N}$. Hence $\sum_{n=1}^{\infty} \frac{n}{3^n}$ converges by the ratio test. □

Sometimes the following variant of the ratio test is useful.

Theorem 4.14 (Limit version of the Ratio Test). *Let* $(x_n)_{n=1}^{\infty}$ *be a sequence. Assume that there exists* $n_0 \in \mathbb{N}$ *such that* $x_n > 0$ *for all* $n \geq n_0$ *and that* $\frac{x_{n+1}}{x_n} \to L$ *as* $n \to \infty$ *for some* $L \in [0, \infty)$.

a) *If* $L < 1$ *then* $\sum_{n=1}^{\infty} x_n$ *converges.*

b) *If* $L > 1$ *then* $\sum_{n=1}^{\infty} x_n$ *diverges.*

Proof. Again, we may assume $n_0 = 1$. Let $L \in [0, 1)$ and pick $\lambda \in (L, 1)$. Then $\varepsilon = \lambda - L > 0$, so by the definition of convergence, there exists $n_1 \in \mathbb{N}$ such that

$$n \geq n_1 \implies \left| \frac{x_{n+1}}{x_n} - L \right| < \varepsilon = \lambda - L.$$

In particular, we conclude $\frac{x_{n+1}}{x_n} - L < \lambda - L$, or equivalently $\frac{x_{n+1}}{x_n} < \lambda$ for $n \geq n_1$. Since $\lambda < 1$ we see that $\sum_{n=1}^{\infty} x_n$ converges according to the ratio test.

Now let $L \in (1, \infty)$. Then since $\epsilon = L - 1 > 0$, there exists $n_2 \in \mathbb{N}$ such that

$$n \geq n_2 \implies \left| \frac{x_{n+1}}{x_n} - L \right| < \varepsilon = L - 1.$$

This means $1 - L = -(L-1) < \frac{x_{n+1}}{x_n} - L$, or equivalently, $1 < \frac{x_{n+1}}{x_n}$ for $n \geq n_2$. It follows that $\sum_{n=1}^{\infty} x_n$ diverges according to the ratio test. $\square$

Example 4.15. Determine whether the series

$$\sum_{n=1}^{\infty} \frac{2^n}{n\sqrt{n}}, \quad \sum_{n=0}^{\infty} \frac{x^n}{n!} \text{ for } x > 0, \quad \sum_{n=1}^{\infty} \frac{4^n}{\binom{2n}{n}}$$

are convergent or divergent.

Solution. Consider the first series and write $x_n = \frac{2^n}{n\sqrt{n}}$. Then for any $n \in \mathbb{N}$ we have $x_n > 0$, and

$$\frac{x_{n+1}}{x_n} = \frac{2^{n+1}}{(n+1)\sqrt{n+1}} \frac{n\sqrt{n}}{2^n} = 2\left(\frac{n}{n+1}\right)^{3/2} = 2\left(\frac{1}{1+\frac{1}{n}}\right)^{3/2} \to 2$$

as $n \to \infty$. We conclude that $\sum_{n=1}^{\infty} \frac{2^n}{n\sqrt{n}}$ diverges by the limit version of the ratio test.

For the second series write $y_n = \frac{x^n}{n!}$. Then we have

$$\frac{y_{n+1}}{y_n} = \frac{x^{n+1}}{(n+1)!} \frac{n!}{x^n} = \frac{x}{n+1} = \frac{\frac{x}{n}}{1+\frac{1}{n}} \to 0$$

as $n \to \infty$. Hence $\sum_{n=1}^{\infty} \frac{x^n}{n!}$ converges by the limit version of the ratio test.

Finally, for the third series we write $z_n = \frac{4^n}{\binom{2n}{n}}$. Then for all $n \in \mathbb{N}$,

$$\begin{aligned}\frac{z_{n+1}}{z_n} &= \frac{4^{n+1}}{\binom{2(n+1)}{(n+1)}} \frac{\binom{2n}{n}}{4^n} = \frac{4(n+1)^2}{(2n+2)(2n+1)} = \frac{(2n+2)^2}{(2n+2)(2n+1)} \\ &= \frac{2n+2}{2n+1} \geq 1.\end{aligned}$$

Hence $\sum_{n=1}^{\infty} \frac{4^n}{\binom{2n}{n}}$ diverges by the ratio test. $\square$

Tests for convergence III: the Leibniz test

Let's study another useful criterion for convergence.

Theorem 4.16 (Leibniz's Test)**.** *Let $(x_n)_{n=1}^{\infty}$ be a sequence and suppose that there exists an $n_0 \in \mathbb{N}$ such that*

a) $x_n \geq 0$ for all $n \geq n_0$, that is, the sequence is eventually positive;

b) $x_{n+1} \leq x_n$ for all $n \geq n_0$, that is, the sequence is eventually decreasing;

c) $x_n \to 0$ as $n \to \infty$.

Then the series $\sum_{n=1}^{\infty} (-1)^n x_n$ converges[19]*.*

[19] The same could also be said for the series $\sum_{n=1}^{\infty} (-1)^{n-1} x_n$, by multiplying through by -1 and using algebraic properties of limits. So, as one would expect, it doesn't matter whether it's the odd terms which are eventually negative and evens which are eventually positive, or the other way around, what's important is that the terms are eventually alternating. Alternatively, you can easily check

For the proof of this theorem we need the following fact about sequence limits. Suppose $(s_n)_{n=1}^{\infty}$ is a sequence and $L \in \mathbb{R}$ has the property that both $s_{2n} \to L$ and $s_{2n-1} \to L$ as $n \to \infty$. That is, the subsequence of even terms converges to L, and the subsequence of odd terms converges to L, too. Then[20] $s_n \to L$ as $n \to \infty$.

[20] You should try and prove this, directly from the definition of convergence.

Proof. Without loss of generality, by changing the initial terms of $(x_n)_{n=1}^{\infty}$ if necessary, we may assume that $n_0 = 1$. As usual we set $s_n = \sum_{j=1}^{n} (-1)^j x_j$. Then

$$s_{2(n+1)} = s_{2n+2} = s_{2n} - x_{2n+1} + x_{2n+2} \leq s_{2n}$$

for all $n \in \mathbb{N}$, since $0 \leq x_{2n+2} \leq x_{2n+1}$. Thus the sequence $(s_{2n})_{n=1}^{\infty}$ of even partial sums is decreasing. Similarly, we have

$$s_{2(n+1)-1} = s_{2n+1} = s_{2n-1} + x_{2n} - x_{2n+1} \geq s_{2n-1}$$

for all $n \in \mathbb{N}$, since $0 \leq x_{2n+1} \leq x_{2n}$, so that $(s_{2n-1})_{n=1}^{\infty}$ is increasing. Furthermore, $s_{2n} \geq s_{2n-1}$ for all $n \in \mathbb{N}$ and hence

$$s_2 \geq s_4 \geq s_6 \geq \cdots \geq s_{2n} \geq s_{2n-1} \geq \cdots \geq s_1.$$

It follows that $(s_{2n})_{n=1}^{\infty}$ is bounded below by s_1 and $(s_{2n-1})_{n=1}^{\infty}$ is bounded above by s_2. According to the monotone convergence theorem, we have $\lim_{n\to\infty} s_{2n} = S$ and $\lim_{n\to\infty} s_{2n-1} = T$ for some $S, T \in \mathbb{R}$. Moreover $s_{2n} = s_{2n-1} + x_{2n} \to T$ since $x_n \to 0$ as $n \to \infty$. Thus $S = T$, and hence $(s_n)_{n=1}^{\infty}$ converges. $\square$

Example 4.17. Show that

$$\sum_{n=1}^{\infty} (-1)^{n-1} \frac{n^2}{n^3 - 5}$$

converges.

Solution. Let $x_n = \frac{n^2}{n^3-5}$. Then

$$x_n = \frac{n^2}{n^3 - 5} \geq 0$$

for $n \geq 2$. Moreover,

$$\begin{aligned} x_n - x_{n+1} &= \frac{n^2}{n^3 - 5} - \frac{(n+1)^2}{(n+1)^3 - 5} \\ &= \frac{n^4 + 2n^3 + n^2 + 10n + 5}{(n^3 - 5)(n^3 + 3n^2 + 3n - 4)} > 0 \end{aligned}$$

for $n \geq 2$. Hence the sequence $(x_n)_{n=1}^{\infty}$ is eventually decreasing. Finally, we have

$$x_n = \frac{n^2}{n^3 - 5} = \frac{1/n}{1 - 5/n^3} \to 0$$

as $n \to \infty$. Hence, by Leibniz's test, the series $\sum_{n=1}^{\infty} (-1)^{n-1} \frac{n^2}{n^3-5}$ converges. $\square$

Absolute convergence

In this section we study absolute convergence, a property which turns out to be stronger than convergence. Let's start by giving some definitions.

Definition 4.18. A series $\sum_{n=1}^{\infty} x_n$ is said to be absolutely convergent if the series $\sum_{n=1}^{\infty} |x_n|$ is convergent. We say that $\sum_{n=1}^{\infty} x_n$ is conditionally convergent if it is convergent, but not absolutely convergent.

Here are a few examples of absolutely convergent series.

Example. *The series $\sum_{n=1}^{\infty} \frac{1}{n^2}\cos(n)$ is absolutely convergent.*

Solution. If we write $x_n = \frac{1}{n^2}\cos(n)$, then clearly $|x_n| \leq \frac{1}{n^2}$ for all $n \in \mathbb{N}$. Hence $\sum_{n=1}^{\infty} |\frac{1}{n^2}\cos(n)|$ converges by the comparison test. By definition, this means $\sum_{n=1}^{\infty} \frac{1}{n^2}\cos(n)$ is absolutely convergent. □

Example. *The series $\sum_{n=1}^{\infty} \frac{x^n}{n!}$ is absolutely convergent for all $x \in \mathbb{R}$.*

Solution. Let us write $y_n = \frac{x^n}{n!}$. Then we have $|y_n| = \frac{|x|^n}{n!}$. Hence $\sum_{n=1}^{\infty} |y_n|$ is convergent according to the second example in 4.15. That is, $\sum_{n=1}^{\infty} \frac{x^n}{n!}$ is absolutely convergent. □

We will see next that absolute convergence of a series implies its convergence. That is, absolute convergence is a stronger property than convergence.

Theorem 4.19. *Let $\sum_{n=1}^{\infty} x_n$ be an absolutely convergent series. Then $\sum_{n=1}^{\infty} x_n$ converges.*

Proof. For any $n \in \mathbb{N}$ we have $-|x_n| \leq x_n \leq |x_n|$, so that $0 \leq x_n + |x_n| \leq 2|x_n|$. Since $\sum_{n=1}^{\infty} |x_n|$ converges by assumption, also $2\sum_{n=1}^{\infty} |x_n|$ converges. Hence $\sum_{n=1}^{\infty} x_n + |x_n|$ converges by the comparison test. Finally, since

$$x_n = (x_n + |x_n|) - |x_n|,$$

also $\sum_{n=1}^{\infty} x_n$ converges using algebraic properties. □

The converse of Theorem 4.19 does not hold. That is, there are convergent series which are not absolutely convergent.

Example. *The series $\sum_{n=1}^{\infty}(-1)^{n-1}\frac{1}{n}$ is convergent, but not absolutely convergent.*

Solution. The series $\sum_{n=1}^{\infty}(-1)^{n-1}\frac{1}{n}$ is convergent by Leibniz's test, but it is not absolutely convergent because the harmonic series $\sum_{n=1}^{\infty} \frac{1}{n}$ is divergent. □

Let's have a look at some further examples.

Example. *Determine whether the series*

$$\sum_{n=1}^{\infty}(-1)^{n-1}\frac{n^2}{n^3-5}$$

is absolutely convergent, conditionally convergent or divergent.

Solution. According to Leibniz's test, the series converges. Since

$$\frac{n^2}{n^3-5} \geq \frac{n^2}{n^3} = \frac{1}{n}$$

for $n \geq 3$, the series $\sum_{n=1}^{\infty} \frac{n^2}{n^3-5}$ diverges by the comparison test. Hence $\sum_{n=1}^{\infty} (-1)^{n-1} \frac{n^2}{n^3-5}$ is conditionally convergent. □

Example. *Determine whether the series*

$$\sum_{n=1}^{\infty} \frac{(-2)^n}{n3^n}$$

is absolutely convergent, conditionally convergent or divergent.

Solution. Writing $x_n = \frac{(-2)^n}{n3^n}$ we have $|x_n| = \frac{2^n}{n3^n}$ and

$$\frac{|x_{n+1}|}{|x_n|} = \frac{2^{n+1}}{(n+1)3^{n+1}} \frac{n3^n}{2^n} = \frac{2}{3} \frac{n}{n+1} \to \frac{2}{3}$$

as $n \to \infty$. Hence $\sum_{n=1}^{\infty} \frac{(-2)^n}{n3^n}$ is absolutely convergent according to the limit version of the ratio test. □

Guide for testing convergence

We have seen a number of criteria for testing convergence. So, given a series $\sum_{n=1}^{\infty} a_n$, how should we decide which of these criteria to apply?

Here is a guide for how to systematically make use of the methods we have discussed.

1. Does $a_n \to 0$?

No The series diverges.

Yes Proceed to 2.

2. Are the a_n eventually positive?

Yes

- Is the form of the a_n's amenable to the ratio test (do we find expressions like $x^n, n!, \binom{2n}{n}$ in the definition of a_n, so that there will be cancellations in a_{n+1}/a_n)? If so consider the *ratio test*.
- Otherwise look at the *comparison test*. You may compare the series to one of the standard series we have seen above. Remember in particular that $\sum_{n=1}^{\infty} \frac{1}{n^p}$ converges if and only if $p > 1$, and that $\sum_{n=0}^{\infty} x^n$ converges if and only if $|x| < 1$.

No

- Do the terms alternate in sign $+, -, +, -, +, \ldots$? If so consider Leibniz's test.
- Otherwise try to see if the series is absolutely convergent. Use the above methods for $\sum_{n=1}^{\infty} |a_n|$.

It worthwhile to point out that this is only a rough guide, in the sense that it is not guaranteed that we'll find out about convergence/divergence using the above methods — still it helps in many examples, in particular in the examples you'll find on the exercise sheets.

Rearrangements

Given a series $\sum_{n=1}^{\infty} x_n$, we can form a new series by *rearranging* the terms x_n. For instance, we could consider

$$x_1 + x_2 + x_4 + x_3 + x_6 + x_8 + x_5 + x_{10} + x_{12} + \cdots,$$

A rearrangement corresponds to a relabeling of the indices of the terms of the series[21]. To make that precise we give the following definition.

[21] In particular, all terms x_n appear exactly once in a rearrangement, we are neither allowed to write e.g. $x_1 + x_2 + x_2 + x_4 + x_3 + \cdots$, nor to leave out any x_n.

Definition 4.20. Let $\sum_{n=1}^{\infty} x_n$ be a series. We call a a series $\sum_{n=1}^{\infty} y_n$ a rearrangement of $\sum_{n=1}^{\infty} x_n$ if there exists a bijection $\theta : \mathbb{N} \to \mathbb{N}$ such that $y_n = x_{\theta(n)}$.

A natural question one may ask is how a rearrangement affects convergence. We will present two results in this direction.

Theorem 4.21. *Let $\sum_{n=1}^{\infty} x_n$ be an absolutely convergent series. Then any rearrangement $\sum_{n=1}^{\infty} x_{\theta(n)}$ of $\sum_{n=1}^{\infty} x_n$ is also absolutely convergent and converges to the same sum.*

Proof. Let us write $y_n = x_{\theta(n)}$, so that $\sum_{n=1}^{\infty} x_{\theta(n)} = \sum_{n=1}^{\infty} y_n$.

Assume first that $x_n \geq 0$ for all $n \in \mathbb{N}$. Given any $k \in \mathbb{N}$, let m_k be the maximum of the numbers $\theta(1), \ldots, \theta(k)$. Then

$$\sum_{j=1}^{k} y_j = \sum_{j=1}^{k} x_{\theta(j)} \leq \sum_{j=1}^{m_k} x_j \leq \sum_{j=1}^{\infty} x_j,$$

that is, the sequence of partial sums $s_k = \sum_{j=1}^{k} y_j$ is bounded. Moreover it is increasing because all terms x_n are positive. Hence the series $\sum_{n=1}^{\infty} y_n$ is convergent by the monotone convergence theorem. By ordering properties of limits, we also see that $\sum_{n=1}^{\infty} y_n \leq \sum_{n=1}^{\infty} x_n$. Since $\sum_{n=1}^{\infty} x_n$ may also be considered as a rearrangement of $\sum_{n=1}^{\infty} y_n$, a symmetric argument to the above shows that $\sum_{n=1}^{\infty} x_n \leq \sum_{n=1}^{\infty} y_n$, so these sums are equal.

Now consider the general case. For $x \in \mathbb{R}$ we set

$$x^+ = \max\{0, x\}, \quad x^- = \max\{-x, 0\}.$$

Then[22] $x^+ - x^- = x$ and $x^+ + x^- = |x|$. Since

[22] Check this claim by distinguishing the cases $x \geq 0$ and $x < 0$.

$$0 \leq x_n^+ \leq |x_n|, \qquad 0 \leq x_n^- \leq |x_n|,$$

the series $\sum_{n=1}^{\infty} x_n^+$ and $\sum_{n=1}^{\infty} x_n^-$ converge according to the comparison test. Note that y_n^+ is a rearrangement of x_n^+, and similarly y_n^- is a rearrangement of x_n^-. By the argument above, it follows that $\sum_{n=1}^{\infty} y_n^+$ and $\sum_{n=1}^{\infty} y_n^-$ converge as well and sum to the same corresponding values. This implies that $\sum_{n=1}^{\infty} |y_n| = \sum_{n=1}^{\infty} y_n^+ + y_n^-$ converges too, so $\sum_{n=1}^{\infty} y_n$ converges absolutely. Moreover, we have that

$$\sum_{n=1}^{\infty} x_n = \sum_{n=1}^{\infty} x_n^+ - \sum_{n=1}^{\infty} x_n^- = \sum_{n=1}^{\infty} y_n^+ - \sum_{n=1}^{\infty} y_n^- = \sum_{n=1}^{\infty} y_n,$$

where the first and last equalities above follow from Theorem 4.6. □

We conclude this section with two remarks on rearrangements of conditionally convergent sequences. It turns out that if the requirement of absolute convergence of $\sum_{n=1}^{\infty} x_n$ is weakened to conditional convergence, then things go badly wrong. In fact, the following result holds[23].

[23] A rather strange and surprising result at first sight!

Theorem 4.22. *Let $\sum_{n=1}^{\infty} x_n$ be conditionally convergent and let $T \in \mathbb{R}$ be arbitrary. Then there exists a rearrangement of $\sum_{n=1}^{\infty} x_n$ which converges to T.*

We will not prove this theorem here[24]. For conditionally convergent series, there are also rearrangements which are divergent.

[24] For a proof and more information see chapter 3 in Rudin's book.

2C Intro to real analysis 2020/21

HANDOUT

Continuity

In this chapter we study the notion of *continuity* for real functions. Intuitively, continuity of a function f means that one can draw the graph of f without lifting the pen from the paper. Most functions you have seen before are continuous, but there are also some rather basic examples of functions which fail to be continuous.

From a general mathematical perspective, continuity is an important regularity property for functions[1]. From the continuity of a function f a number of nice properties can be deduced, and this plays a role in a wide range of problems in analysis and its applications. We will see some of this below, most notably the intermediate value theorem and the extreme value theorem.

Before entering the discussion of continuity, let us first recall what we mean by a *function*. Although we often simply write a formula like

$$f(x) = \frac{x^3 - 2}{x + 1}$$

and refer to this as a "function", the precise mathematical definition of a function is as follows: If X and Y are sets, then a function from X to Y is an object $f : X \to Y$ which assigns a unique element $f(x) \in Y$ to each $x \in X$. We also refer to X as the *domain* of the function and Y as the *codomain*, and sometimes write $X = \mathrm{dom}(f)$ and $Y = \mathrm{codom}(f)$.

In practice, one often omits the domain and codomain from the definition of a function. For instance, $f(x) = 3 + 2x^2 + \sin(x)$ is a function, tacitly with domain and codomain $\mathbb{R}$. Note that we could also view this as a function with domain, say, $[0, 1]$ and codomain $\mathbb{R}$. In our previous example $f(x) = \frac{x^3-2}{x+1}$ we should similarly choose domain and codomain appropriately[2]. Simply writing down a formula does not specify a function completely, and one should in principle always indicate both the domain and codomain[3].

Let us also point out that a function need not be given by an explicit formula[4]. For instance,

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

is a perfectly valid function $f : \mathbb{R} \to \mathbb{R}$.

By a *real function* we mean a function f whose domain $\mathrm{dom}(f)$ is some subset of $\mathbb{R}$, possibly all of $\mathbb{R}$, and whose codomain is $\mathbb{R}$, or possibly some subset thereof. Thus, for each $x \in \mathrm{dom}(f)$ of a real function f there is a unique element $f(x) \in \mathbb{R}$ associated to x.

[1] Another crucial regularity property is *differentiability*, this will be one of the main topics of the 3H course on Analysis of Differentiation and Integration next year.

[2] Here one could consider the domain $X = \mathbb{R} \setminus \{-1\}$, or any other subset X of $\mathbb{R}$ not containing -1. We have to exclude $x = -1$ from the domain since the expression for $f(-1)$ contains an illegal division by zero. As codomain we can simply take $Y = \mathbb{R}$.

[3] In practice, however, this is often neglected.

[4] We already indicated this in the chapter on sequences. Recall that sequences are particular examples of functions, with domain $\mathbb{N}$ and codomain $\mathbb{R}$.

The definition of continuity

Let us now give the precise definition of continuity.

Definition 5.1. Let $f : \mathrm{dom}(f) \to \mathbb{R}$ be a real function. If $c \in \mathrm{dom}(f)$, we say that f is continuous[5] at c if for all $\varepsilon > 0$ there exists $\delta > 0$ such that for all $x \in \mathrm{dom}(f)$ with $|x - c| < \delta$ we have $|f(x) - f(c)| < \epsilon$.

We say that f is continuous if f is continuous at all points of its domain $\mathrm{dom}(f)$.

[5] In ERA the definition of continuity at c is slightly different, as it additionally asks for $\mathrm{dom}(f)$ to contain a neighbourhood of c. However, our definition is standard. Ignore those parts of ERA where it discusses the domain containing a neighbourhood!

The definition of continuity at c can also be rewritten using quantifiers. More precisely, the function f is continuous at $c \in \mathrm{dom}(f)$ if

$$\forall \varepsilon > 0, \exists \delta > 0 \text{ s.t. } \forall x \in \mathrm{dom}(f), |x - c| < \delta \implies |f(x) - f(c)| < \varepsilon.$$

As we will see in examples, the value of δ will usually depend on ε. If you make ε smaller you should expect the corresponding value of δ to be smaller, too. The general idea behind the concept of continuity is that *small changes in the argument* x of a continuous function f lead to *small changes in the value* $f(x)$.

Let us consider some examples of continuous functions.

Example 5.2. Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = 3x + 7$ for $x \in \mathbb{R}$. Show directly from the definition[6] that f is continuous at $c = 2$.

[6] As usual, *directly from the definition* means you shall verify the condition in the definition of continuity, that is Definition 5.1.

Solution. Let $\varepsilon > 0$. Then for $x \in \mathbb{R}$ we have

$$|f(x) - f(2)| = |3x + 7 - 13| = |3x - 6| = 3|x - 2|.$$

Therefore, $|f(x) - f(2)| < \varepsilon$ is equivalent to $3|x - 2| < \varepsilon$, or $|x - 2| < \varepsilon/3$. If we choose $\delta = \varepsilon/3$ we conclude that $|x - 2| < \delta$ implies $|f(x) - f(2)| < \varepsilon$. Thus f is continuous at 2. □

Example 5.3. Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x^2 + x + 1$ for $x \in \mathbb{R}$. Show that f is continuous at $c = 3$.

Solution. Let $\varepsilon > 0$. For $x \in \mathbb{R}$ we compute

$$|f(x) - f(3)| = |x^2 + x + 1 - 13| = |x^2 + x - 12| = |x - 3||x + 4|.$$

We therefore have to find $\delta > 0$ such that $|x - 3| < \delta$ implies $|x - 3||x + 4| < \varepsilon$. Notice that we can pick the number δ as small as we like. So we can try to find a suitable δ in the interval $(0, 1]$. Now for[7] $|x - 3| < \delta \leq 1$ we have

[7] The inequality $\delta \leq 1$ I'm using here is just a convenient choice, other choices are possible.

$$-1 < x - 3 < 1 \Rightarrow 6 < x + 4 < 8,$$

which means $|x + 4| < 8$. Inserting this into the relation above yields

$$|x - 3||x + 4| < \delta 8$$

provided[8] $|x - 3| < \delta$ and $\delta \leq 1$. Therefore, if we set $\delta = \min(1, \varepsilon/8)$ we obtain $|x - 3| < \delta \Rightarrow |f(x) - f(3)| = |x - 3||x + 4| < \epsilon$. Hence f is continuous at 3. □

[8] What happens if we replace the bound 1 for δ by some other value? What happens e.g. if we consider $\delta \leq 2$?

Example 5.4. Let $f : \mathbb{R} \setminus \{1\} \to \mathbb{R}$ be given by

$$f(x) = \frac{x+1}{x-1}.$$

Show that f is continuous at 2.

Solution. Let $\varepsilon > 0$. For $x \in \mathbb{R}$ we have

$$|f(x) - f(2)| = \left| \frac{x+1}{x-1} - 3 \right| = \left| \frac{x+1-3x+3}{x-1} \right| = \frac{2|x-2|}{|x-1|}.$$

Now $|x-2| < 1/2$ implies $-1/2 < x-2 < 1/2$, and hence[9] $1/2 < x-1 < 3/2$. Therefore we obtain

$$\frac{2|x-2|}{|x-1|} < 4|x-2|$$

provided $|x-2| < 1/2$. Taking $\delta = \min(1/2, \varepsilon/4)$, we get that $|x-2| < \delta$ implies $|f(x) - f(2)| < \varepsilon$. □

[9] Here I start from $|x-2| < 1/2$ instead of $|x-2| < 1$, compare with the previous example. Can you see why taking $|x-2| < 1$ would not work?

Example 5.5. Let $f : [0, \infty) \to \mathbb{R}$ be given by $f(x) = \sqrt{x}$. Show directly from the definition that f is continuous.

Solution. Here we have to show continuity at all points in the domain of f. So we start by fixing $c \in [0, \infty)$, and let $\varepsilon > 0$ be arbitrary. For $x \in [0, \infty)$ we have

$$|\sqrt{x} - \sqrt{c}|^2 = |\sqrt{x} - \sqrt{c}||\sqrt{x} - \sqrt{c}| \le |\sqrt{x} - \sqrt{c}||\sqrt{x} + \sqrt{c}| = |x-c|$$

since both $\sqrt{x}$ and $\sqrt{c}$ are positive[10]. Therefore

$$|f(x) - f(c)| = |\sqrt{x} - \sqrt{c}| \le |x-c|^{1/2}.$$

Therefore $|f(x) - f(c)| < \varepsilon$ provided $|x-c|^{1/2} < \varepsilon$, or equivalently, $|x-c| < \varepsilon^2$. Hence if we set $\delta = \varepsilon^2$ we conclude that $|x-c| < \delta$ implies $|f(x) - f(c)| < \varepsilon$. Thus f is continuous at c. Since $c \in \mathrm{dom}(f)$ was arbitrary this means that f is continuous. □

[10] and hence $|\sqrt{x} - \sqrt{c}| \le |\sqrt{x} + \sqrt{c}| = \sqrt{x} + \sqrt{c}$.

The next example shows that continuous functions can exhibit a certain kind of singular behaviour, in the sense that they need not be *smooth*[11].

[11] That is, there can be "corners" in the graph of f.

Example 5.6. Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = |x|$. Show that f is continuous[12].

[12] Intuitively, the function f "changes its slope" drastically at zero. It will be shown in the 3H course on Analysis of Differentiation and integration that f fails to be *differentiable* at 0.

Solution. Again we have to prove continuity at c for every $c \in \mathrm{dom}(f) = \mathbb{R}$. We consider first the case $c > 0$, and let $\varepsilon > 0$ be arbitary. In this case, for $|x-c| < c$ the number x is positive as well, and hence

$$|f(x) - f(c)| = |x-c|.$$

Therefore, if we take $\delta = \min(\varepsilon, c)$, then[13] $|x-c| < \delta$ implies $|f(x) - f(c)| < \varepsilon$.

In a similar way on proves[14] continuity at c for $c < 0$.

The most interesting part is to show that f is continuous at 0. Let $\varepsilon > 0$. We have

$$|f(x) - f(0)| = ||x| - 0| = |x|,$$

[13] In a similar way as before, we take $\delta = \min(\varepsilon, c)$, and not simply $\delta = \varepsilon$ in order to ensure that our previous calculation $|f(x) - f(c)| = |x-c|$ is valid.

[14] Try to write down the details!

so if we set $\delta = \varepsilon$ then $|x - 0| < \delta$ implies $|f(x) - f(0)| < \epsilon$ as required.

Summarising, we have shown that f is continuous at c for all $c \in \operatorname{dom}(f)$, and hence f is continuous. $\square$

Let us now discuss an example where continuity fails. We say that a real function f is *discontinuous* at $c \in \operatorname{dom}(f)$ if f fails to be continuous at c.

Example 5.7. Let $f : \mathbb{R} \to \mathbb{R}$ be given by

$$f(x) = \begin{cases} 1 & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

Show that f is discontinuous at 0.

Solution. Let[15] $\varepsilon = 1/2$. Then for any $\delta > 0$ consider $x = -\delta/2$. We have $|x - 0| = |-\delta/2| = \delta/2 < \delta$ but $|f(x) - f(0)| = |0 - 1| = 1 > \epsilon$. Therefore, f is not continuous at 0. $\square$

[15] Can you see why I chose ε in this way? Which other choices would also work in this argument?

We can even produce real functions which are discontinuous *everywhere*. The following example is an exercise.

Example 5.8. Let $\chi_{\mathbb{Q}} : \mathbb{R} \to \mathbb{R}$ be defined by

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

Then $\chi_{\mathbb{Q}}$ is not continuous at any $c \in \mathbb{R}$.

Intuitively, it should be clear that $\chi_{\mathbb{Q}}$ must be discontinuous: We cannot draw the graph of $\chi_{\mathbb{Q}}$ without lifting the pen from the paper; in fact, it is hard to draw the graph of f at all!

Properties of continuous functions

In this section we collect some results which will help us to check continuity, and prove that a large variety of familiar functions are continuous.

The first result which we'll present allows us to reduce matters to the study of certain sequences and their limits, that is, things we have already discussed in detail before.

Theorem 5.9 (Sequential characterisation of continuity)**.** *Let $f : \operatorname{dom}(f) \to \mathbb{R}$ be a real function and $c \in \operatorname{dom}(f)$. Then f is continuous at c if and only for every sequence $(x_n)_{n=1}^{\infty}$ in $\operatorname{dom}(f)$ with $x_n \to c$ as $n \to \infty$ we have $f(x_n) \to f(c)$ as $n \to \infty$.*

Proof. Assume first that f is continuous at c, and let $x_n \to c$ be a sequence in[16] $\operatorname{dom}(f)$ converging to c. Moreover let $\varepsilon > 0$ be arbitrary. Then, by the definition of continuity, there exists $\delta > 0$ such that $|x - c| < \delta$ implies $|f(x) - f(c)| < \varepsilon$. Since $x_n \to c$ and $\delta > 0$, there exists $n_0 \in \mathbb{N}$ such that $|x_n - c| < \delta$ for all $n \geq n_0$. In particular,

[16] We say that a sequence $(x_n)_{n=1}^{\infty}$ is *in* $\operatorname{dom}(f)$ if $x_n \in \operatorname{dom}(f)$ for all $n \in \mathbb{N}$, that is, if all its terms are contained in $\operatorname{dom}(f)$.

this implies $|f(x_n) - f(c)| < \varepsilon$ for $n \geq n_0$. That is, $f(x_n) \to f(c)$ as $n \to \infty$.

Conversely, assume that $f(x_n) \to f(c)$ as $n \to \infty$ for every sequence $(x_n)_{n=1}^{\infty}$ in $\mathrm{dom}(f)$ satisfying $x_n \to c$. Moreover suppose that f is not continuous at c. Then there exists an $\epsilon > 0$ such that for any $\delta > 0$ we find $x \in \mathrm{dom}(f)$ with $|x - c| < \delta$ and $|f(x) - f(c)| \geq \epsilon$. In particular, for each $n \in \mathbb{N}$ we can choose $x_n \in \mathrm{dom}(f)$ such that $|x - c| < 1/n$ and $|f(x_n) - f(c)| \geq \epsilon$. Then $x_n \to c$ as $n \to \infty$, so our assumption implies $f(x_n) \to f(c)$ as $n \to \infty$. This is a contradiction to $|f(x_n) - f(c)| \geq \epsilon$ for all $n \in \mathbb{N}$. □

We can rephrase Theorem 5.9 by saying that

$$f(\lim_{n\to\infty} x_n) = \lim_{n\to\infty} f(x_n)$$

whenever $(x_n)_{n=1}^{\infty}$ is a convergent sequence in $\mathrm{dom}(f)$, provided f is continuous.

Let us revisit one of our previous examples using Theorem 5.9.

Example 5.10. Use the sequential characterisation of continuity to show that $f : \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

is discontinuous at 0.

Solution. Consider $x_n = -\frac{1}{n}$. Then $x_n \to 0$, but since $f(x_n) = 0$ for all n we have $f(x_n) \to 0$, and this differs from $f(0) = 1$. By the sequential characterisation of continuity, the function f is not continuous at 0. □

We now come to algebraic properties of continuous functions. You should compare the following theorem and its proof with algebraic properties of sequence limits in chapter 3.

Theorem 5.11 (Algebraic properties of continuity)**.** *Let $f : \mathrm{dom}(f) \to \mathbb{R}$ and $g : \mathrm{dom}(g) \to \mathbb{R}$ be real functions and $\lambda \in \mathbb{R}$. Moreover let $c \in \mathrm{dom}(f) \cap \mathrm{dom}(g)$ and assume that f and g are continuous at c. Then*

a) $f + g$ is continuous at c;

b) λf is continuous at c;

c) fg is continuous at c;

d) If $g(c) \neq 0$, then f/g is continuous at c.

Let us first clarify what the functions in these statements are. If f and g are functions, then $f + g$ and fg are defined[17] by

$$(f+g)(x) = f(x) + g(x), \qquad (fg)(x) = f(x)g(x),$$

respectively. Moreover, we define the function λf by

$$(\lambda f)(x) = \lambda f(x)$$

[17] Here we tacitly pick the domains of definition appropriately: Both $f + g$ and fg are defined for $x \in \mathrm{dom}(f) \cap \mathrm{dom}(g)$, so it is natural to set $\mathrm{dom}(f + g) = \mathrm{dom}(f) \cap \mathrm{dom}(g) = \mathrm{dom}(fg)$. For λf we do not need to restrict the domain, we can simply take $\mathrm{dom}(\lambda f) = \mathrm{dom}(f)$. Notice however that in principle we could also choose smaller domains for all these functions.

for any $\lambda \in \mathbb{R}$. Finally, f/g is defined by $(f/g)(x) = f(x)/g(x)$ for $x \in \operatorname{dom}(f) \cap \operatorname{dom}(g)$ such that $g(x) \neq 0$.

I'm going to prove the first and last statement of the theorem, and encourage you to prove the other parts.

Proof of a). Let $\epsilon > 0$. By assumption, there exists $\delta_1, \delta_2 > 0$ such that[18]

$$|x - c| < \delta_1 \implies |f(x) - f(c)| < \frac{\varepsilon}{2}$$
$$|x - c| < \delta_2 \implies |g(x) - g(c)| < \frac{\varepsilon}{2}.$$

Take $\delta = \min(\delta_1, \delta_2)$. Then, for $|x - c| < \delta$, we have

$$\begin{aligned} |(f(x) + g(x)) - (f(c) + g(c))| &= |(f(x) - f(c)) + (g(x) - g(c))| \\ &\leq |f(x) - f(c)| + |g(x) - g(c)| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon, \end{aligned}$$

using the triangle inequality. Since $\varepsilon > 0$ was arbitrary, $f + g$ is continuous at c. □

The key thing that made this proof work was the use of the triangle inequality, in a similar way as we have already seen a few times before. The triangle inequality enables me to control the distance from $(f+g)(x) = f(x) + g(x)$ to $(f+g)(c) = f(c) + g(c)$ in terms to the two distances I understand, namely $|f(x) - f(c)|$ and $|g(x) - g(c)|$. I had found this inequality before embarking on writing my proof.

In the above proof, I verified directly the definition of continuity for the function $f + g$. For part d), I will use Theorem 5.9 in order to reduce the claim to a known result about sequence limits.

Before we embark on this, we notice that $g(x)$ is nonzero for x sufficiently close to c, so that $f(x)/g(x)$ makes sense[19].

In particular, if $(x_n)_{n=1}^{\infty}$ is a sequence in $\operatorname{dom}(g)$ converging to c, then the terms $g(x_n)$ will be nonzero except for a finite number of indices $n \in \mathbb{N}$.

With this in mind, let us now write down a proof.

Proof of d). Let $(x_n)_{n=1}^{\infty}$ be a sequence in $\operatorname{dom}(f) \cap \operatorname{dom}(g)$ converging to c. Since f and g are continuous, we have $f(x_n) \to f(c)$ and $g(x_n) \to g(c)$ as $n \to \infty$ due to Theorem 5.9. Using algebraic properties of sequence limits [20], this implies $(f/g)(x_n) = f(x_n)/g(x_n) \to f(c)/g(c) = (f/g)(c)$ as $n \to \infty$. Using again Theorem 5.9, this means that f/g is continuous at c. □

In this proof we have made use of results from earlier on, and that's why our argument becomes quite short and simple. Trying to write down a direct proof as for part a) above would actually be rather cumbersome[21].

Using Theorem 5.11, we can prove that polynomial functions and rational functions are always continuous[22].

[18] I'm going to use $\frac{\varepsilon}{2}$ in the definition of continuity for f and g. Can you see why I want to take $\frac{\varepsilon}{2}$ in this definition?

[19] Indeed, consider $\varepsilon_0 = |g(c)|$. By continuity of g at c there exists $\delta_0 > 0$ such that $|x - c| < \delta_0$ implies $|g(x) - g(c)| < \varepsilon_0$. This means $-|g(c)| < g(x) - g(c) < |g(c)|$. If $g(c) > 0$ we conclude $0 < g(c) - |g(c)| < g(x)$, and if $g(c) < 0$ we have $g(x) < |g(c)| + g(c) = 0$. Hence in both cases $g(x) \neq 0$ for all x with $|x - c| < \delta_0$.

[20] More precisely, part d) of Theorem 3.10 in chapter 3.

[21] This is a key idea to keep in mind when thinking about how to construct proofs: can I use known results to reduce the argument to something simpler? Working this way will often save time and effort.

[22] You should compare this to the proofs *directly from the definition* in examples 5.2, 5.3 and 5.4 of the previous subsection.

Theorem 5.12. *Let*

$$g(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

be a polynomial with real coefficients $a_0, \ldots, a_n$. Then the function $g : \mathbb{R} \to \mathbb{R}$ mapping x to $g(x)$ is continuous.

More generally, let

$$f(x) = \frac{g(x)}{h(x)}$$

be a rational function where $g(x), h(x)$ are polynomials with real coefficients, and let $N = \{x \in \mathbb{R} \mid h(x) = 0\}$ be the set of zeros of $h(x)$. Then the function $f : \mathbb{R} \setminus N \to \mathbb{R}$ mapping x to $f(x)$ is continuous.

Proof. Notice first that the identity function $x \mapsto x$ is continuous[23]. Using parts b) and c) of Theorem 5.11 we see that $x \mapsto a_j x^n$ is continuous on $\mathbb{R}$ for any $a_j \in R$. Then part a) of Theorem 5.11 implies that $g : \mathbb{R} \to \mathbb{R}, g(x) = \sum_{j=0}^{n} a_j x^j$ is continuous.

Now consider the rational function $f(x) = \frac{g(x)}{h(x)}$. Using part d) of Theorem 5.11 we see that $f(x)$ is continuous on its maximal domain of definition[24] $\mathbb{R} \setminus N$. Notice here that the expression $\frac{g(x)}{h(x)}$ is only defined when the denominator $h(x)$ is nonzero. □

[23] Formally, we consider $id : \mathbb{R} \to \mathbb{R}$ defined by $id(x) = x$. Given $\epsilon > 0$ choose $\delta = \epsilon$. Then obviously $|x - c| < \delta$ implies $|id(x) - c| = |x - c| < \delta = \epsilon$.

[24] Recall that if X is a set and $Y \subset X$ is any subset, then $X \setminus Y$ is defined by $X \setminus Y = \{x \in X \mid x \notin Y\}$.

Next we discuss how continuity behaves with respect to the composition of functions. Assume that $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ are functions. Recall that the composition $g \circ f : \mathbb{R} \to \mathbb{R}$ is the real function defined by

$$(g \circ f)(x) = g(f(x)).$$

More generally, if f and g are not necessarily defined on all of $\mathbb{R}$, the previous formula defines a function $g \circ f : \mathrm{dom}(f) \to \mathbb{R}$ provided the image[25] $\mathrm{im}(f)$ of f satisfies $\mathrm{im}(f) \subset \mathrm{dom}(g)$.

[25] The image of g is defined by $\mathrm{im}(g) = \{g(x) : x \in \mathrm{dom}(g)\}$.

Theorem 5.13. *Let f, g be real functions and assume that f is continuous at c and g is continuous at $f(c)$. Then $g \circ f$ is continuous at c.*

Proof. Let $\varepsilon > 0$ be arbitrary. Since g is continuous at $f(c)$, there exists $\eta > 0$ such that $|y - f(c)| < \eta$ implies $|g(y) - g \circ f(c)| < \varepsilon$. Since f is continuous at c, there exists $\delta > 0$ such that $|x - c| < \delta$ implies $|f(x) - f(c)| < \eta$. This means that for $|x - c| < \delta$ we have $|(g \circ f)(x) - (g \circ f)(c)| < \varepsilon$. Hence $g \circ f$ is continuous at c. □

Combining Theorem 5.13 with the results above we obtain further examples of continuous functions.

Example 5.14. Let $f : \mathbb{R} \setminus \{5\} \to \mathbb{R}$ be the function

$$f(x) = \sqrt{\frac{1 + x^2 + x^6}{x^2 - 10x + 25}}$$

Show that f is continuous.

Solution. First of all notice that the denominator of the above expression can be written as $x^2 - 10x + 25 = (x-5)^2$, so that the function f is well-defined[26] on its domain $\mathrm{dom}(f) = \mathbb{R} \setminus \{5\}$.

We observe that f can be written as the composition $f = g \circ h$ of the rational function $h : \mathbb{R} \setminus \{5\} \to \mathbb{R}$ given by $h(x) = \frac{1+x^2+x^6}{x^2-10x+25}$ with the square root function $g : [0, \infty) \to \mathbb{R}$ givenby $g(x) = \sqrt{x}$. Here we use that $h(x) \geq 0$ for all $x \in \mathrm{dom}(f)$, so that $\mathrm{im}(h) \subset \mathrm{dom}(g)$.

Now h is continuous according to Theorem 5.12, and we have shown in Example 5.5 that g is continuous. Hence Theorem 5.13 implies that f is continuous. □

[26] The term "well-defined" means that our definition makes sense, in particular, that it does not involve any invalid operations like dividing by zero.

Let us discuss another application of Theorem 5.13. If $f, g : \mathbb{R} \to \mathbb{R}$ are functions, then we define new functions $\max(f,g), \min(f,g) : \mathbb{R} \to \mathbb{R}$ by

$$\max(f,g)(x) = \max(f(x), g(x)), \quad \min(f,g)(x) = \min(f(x), g(x)),$$

respectively.

Example 5.15. Let $f, g : \mathbb{R} \to \mathbb{R}$ be continuous functions. Show that $\max(f,g)$ and $\min(f,g)$ are continuous.

Solution. We have

$$\begin{aligned} \max(f,g)(x) &= \frac{1}{2}(f(x) + g(x) + |f(x) - g(x)|), \\ \min(f,g)(x) &= \frac{1}{2}(f(x) + g(x) - |f(x) - g(x)|). \end{aligned}$$

Since $x \mapsto |f(x) - g(x)|$ is the composition of the continuous functions $f - g$ and absolute value $x \mapsto |x|$, we see using algebraic properties that $\max(f,g)$ is continuous.

The argument for $\min(f,g)$ is analogous. □

The intermediate value theorem

In this subsection we shall study one of the most central results about continuous functions, namely the intermediate value theorem. Essentially, this theorem says that if we are given a continuous function f defined on an interval $[a,b]$, such that $f(a) \neq f(b)$, then f attains all *intermediate values* between $f(a)$ and $f(b)$.

A special case of this situation is when $f(a) < 0$ and $f(b) > 0$. In this case, the theorem says that there must exist some $c \in (a,b)$ such that $f(c) = 0$. This should be no surprise: if we are able to draw the graph of f without lifting the pen, starting with a negative value of $y = f(x)$ at $x = a$, and ending with a positive value at $x = b$, then somewhere in between we'll have to cross the line $y = 0$.

Here is the precise formulation of the intermediate value theorem.

Theorem 5.16 (Intermediate value theorem). *Let $f : [a,b] \to \mathbb{R}$ be a continuous function and assume that d is a number such that $f(a) < d < f(b)$ or $f(b) < d < f(a)$. Then there exists a point $c \in (a,b)$ such that $f(c) = d$.*

Proof. Consider the case $f(a) < d < f(b)$. We define

$$S = \{x \in [a, b] \mid f(x) \leq d\}.$$

Since $f(a) < d$ by assumption we have $a \in S$, so that S is nonempty. By construction, the set S is bounded above by b. Therefore $c = \sup(S)$ exists. We claim that $f(c) = d$.

In order to show this, choose a sequence[27] $(x_n)_{n=1}^{\infty}$ in S converging to c. Since f is continuous at c, we have $f(x_n) \to f(c)$ as $n \to \infty$, and $f(c) \leq d$ because $f(x_n) \leq d$ for all n. Since $d < f(b)$ we have $c \neq b$. Similarly, choose a sequence $(y_n)_{n=1}^{\infty}$ in $(c, b]$ converging to c. Since $(c, b] \cap S$ is empty, we have $f(y_n) > d$ for all $n \in \mathbb{N}$. Again by continuity, this implies $f(c) \geq d$. Hence we obtain $f(c) = d$ as desired.

The case $f(b) < d < f(a)$ is analogous. □

[27] Why can we find such a sequence?

We also record the following variant of the intermediate value theorem: If $f : [a, b] \to \mathbb{R}$ is a continuous function such that $f(a) \leq d \leq f(b)$ or $f(a) \geq d \geq f(b)$, then there exists $c \in [a, b]$ with $f(c) = d$. Indeed, if $d = f(a)$ we can take $c = a$, if $d = f(b)$ we may take $c = b$. In the case that d lies strictly between $f(a)$ and $f(b)$ we obtain the desired point c using Theorem 5.16.

Let's turn to some applications of the intermediate value theorem. Firstly, the theorem can be used to prove the existence of solutions to certain equations without having to find them explicitly.

Example 5.17. Show that the equation $x^3 - 3x + 1 = 0$ has a solution in the interval $(1, 2)$.

Solution. We consider the function $f(x) = x^3 - 3x + 1$. Due to Theorem 5.12 we know that f is continuous, and we calculate $f(1) = -1 < 0$ and $f(2) = 3 > 0$. Hence by the intermediate value theorem, there exists a number $c \in (1, 2)$ such that $f(c) = 0$. This means exactly that c solves the equation $x^3 - 3x + 1 = 0$ as desired. □

In a similar spirit, we can revisit our proof of the existence of the real number $\sqrt{2}$ in chapter 2. More precisely, consider the real function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2 - 2$. Due to Theorem 5.12, the function f is continuous. Moreover $f(0) = -2 < 0$ and $f(2) = 2 > 0$. By the intermediate value theorem, there exists $c \in (0, 2)$ such that $f(c) = 0$. In other words, the number c thus obtained is positive and satisfies $c^2 = 2$, which is precisely saying that $c = \sqrt{2}$. [28]

[28] In chapter 2 we already gave a proof for the existence of $\sqrt{2}$, based on the completeness axiom. We now have another proof of this result, which is shorter and less technical than the first one. If you trace through the ingredients in the second proof, you'll find however that eventually we're again relying completeness; things are just organised in a different way, using more advanced ideas like the concept of continuity.

Another application of the intermediate value theorem provides the existence of *fixed points* for certain maps.

Example 5.18. Let $f : [a, b] \to [a, b]$ be a continuous function. Then there exists a fixed point for f, that is, a point $c \in [a, b]$ such that $f(c) = c$.

Solution. We consider the function $g(x) = f(x) - x$. Then g is continuous. Moreover, we have $g(a) = f(a) - a \geq 0$ as $a \leq f(a)$, and similarly[29] $g(b) = f(b) - b \leq 0$ as $f(b) \leq b$. Hence by the intermediate value theorem, there exists a number $c \in [a, b]$ such that $g(c) = 0$. This means $f(c) = c$, so c is a fixed point of f. □

[29] Here we are using $a \leq f(x) \leq b$ for any $x \in [a, b]$, which holds by assumption.

Apart from the intermediate value theorem, another central result about continuous functions is the following fact.

Theorem 5.19. *Let $f : [a, b] \to \mathbb{R}$ be continuous. Then f is bounded.*

Proof. Suppose f is not bounded above. Then for any $n \in N$, there exists $x_n \in [a, b]$ such that $f(x_n) > n$. By Bolzano-Weierstrass, the sequence $(x_n)_{n=1}^{\infty}$ has a subsequence $(x_{k_n})_{n=1}^{\infty}$ converging to some $c \in [a, b]$. Therefore, by continuity of f at c, we obtain that $f(x_{k_n}) \to f(c)$ as $n \to \infty$. This means in particular that $(f(x_{k_n}))_{n=1}^{\infty}$ is bounded, since we know that convergent sequences are bounded[30]. If M is an upper bound for this sequence, then we have $f(x_{k_n}) \leq M$ for all n. This contradicts the fact that

$$f(x_{k_n}) > k_n \geq n > M$$

for $n \in \mathbb{N}$ with $n > M$. Hence our assumption was wrong, which means that f is bounded above.

In a similar way one shows that f is bounded below. □

[30] Again we are referring to known results on sequences, compare chapter 3.

The extreme value theorem

As a consequence of Theorem 5.19 we obtain the *extreme value theorem*.

Theorem 5.20 (Extreme value theorem)**.** *Let $f : [a, b] \to \mathbb{R}$ be continuous. Then there exist $u, v \in [a, b]$ such that $f(u) \leq f(x) \leq f(v)$ for all $x \in [a, b]$.*

Proof. Let us first show that there exists $v \in [a, b]$ such that $f(x) \leq f(v)$ for all $x \in [a, b]$. We consider the set

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

Clearly S is nonempty, and according to Theorem 5.19 it is bounded. Therefore $M = \sup(S)$ exists. We claim that there exists a $v \in [a, b]$ with $f(v) = M$.

Suppose this is not the case. Then we have $f(x) < M$ for all $x \in [a, b]$. Therefore, we can define a function $g : [a, b] \to \mathbb{R}$ by

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

The function g is continuous according to Theorem 5.12, and therefore bounded by Theorem 5.19. Let $K > 0$ be an upper bound for g, so that $g(x) \leq K$ for all $x \in [a, b]$. This means

$$\frac{1}{M - f(x)} \leq K \iff f(x) \leq M - \frac{1}{K}$$

for all $x \in [a, b]$. Thus $M - \frac{1}{K}$ is an upper bound for f, which contradicts the fact that M is the least upper bound of f.

This contradiction establishes that there exists some $v \in [a, b]$ such that $f(v) = M$. In particular, we have $f(x) \leq f(v)$ for all $x \in [a, b]$ by the fact that M is an upper bound for f.

The existence of u is proved in a similar way. □

Combining the intermediate value theorem and the extreme value theorem, it follows that the image $f([a,b]) = \{f(x) \mid x \in [a,b]\}$ of any bounded interval $[a,b]$ under a continuous function $f : \mathbb{R} \to \mathbb{R}$ is again some bounded interval.

Example 5.21. Let $f : [a,b] \to \mathbb{R}$ be continuous. Show that

$$f([a,b]) = [f(u), f(v)]$$

for some $u, v \in [a,b]$.

Proof. By the extreme value theorem, there exist $u, v \in [a,b]$ such that $f(u) \le f(x) \le f(v)$ for all $x \in [a,b]$. Hence

$$f([a,b]) \subset [f(u), f(v)].$$

If $u = v$ then the set $[f(u), f(v)]$ consists of a single point, and the previous inclusion is clearly an equality.

If $u < v$, then the intermediate value theorem, applied to the continuous function $f : [u,v] \to \mathbb{R}$, shows that for any $d \in [f(u), f(v)]$ there exists $c \in [a,b]$ with $f(c) = d$. Hence we have $f([a,b]) = [f(u), f(v)]$.

In the case $v < u$ one argues similarly, by considering the continuous function $f : [v,u] \to \mathbb{R}$. $\square$

Recall that a map $f : X \to Y$ is a bijection if and only if f is injective and surjective. This is equivalent to the existence of an inverse map $g : Y \to X$ to f. The inverse is uniquely determined by the requirements

$$f \circ g = \mathrm{id}_Y, \qquad g \circ f = \mathrm{id}_X$$

where id_X and id_Y are the identity maps of X and Y, respectively[31].

[31] Following standard notation, we'll write $f^{-1} : Y \to X$ for the inverse map of f in the sequel.

Lemma 5.22. *Let $f : [a,b] \to [c,d]$ be a continuous bijection. Then $f(a) = c$ and $f(b) = d$ or $f(a) = d$ and $f(b) = c$.*

Proof. According to Example 5.21, surjectivity of f implies that there exist $u, v \in [a,b]$ with $f(u) = c$ and $f(v) = d$, respectively.

Assume first $u \le v$ and consider $f([u,v]) \subset [c,d]$. By the Intermediate Value Theorem, for any $y \in [f(u), f(v)] = [c,d]$ we find $x \in [u,v]$ such that $f(x) = y$. In other words, we have $f([u,v]) = [c,d]$. In particular, if $a < u$ then $y = f(a) \in [c,d]$ is equal to $f(x)$ for some $x \in [u,v]$. That is, $f(a) = f(x)$ and $a < x$, in particular, $a \ne x$. This contradicts injectivity of f. Similarly, if $v < b$ we get $y = f(b)$ equals $f(x)$ for $x \in [u,v]$, which contradicts injectivity of f again. We conclude $a = u$ and $b = v$.

The case $u \ge v$ is dealt with in a similar way, in this case one obtains $a = v$ and $b = u$.

$\square$

Theorem 5.23. *Let $f : [a,b] \to [c,d]$ be a bijective continuous map. Then*

a) f and f^{-1} are both strictly increasing or strictly decreasing;

b) *The inverse map f^{-1} is continuous.*

Proof. Consider first part a). We'll consider the case that $f(a) < f(b)$. The according to Lemma 5.22 we have $f(a) = c$ and $f(b) = d$. Now let $x, y \in [a, b]$ with $x < y$. If $f(x) \geq f(y)$ then $f([a, x]) \cap f([y, b])$ is nonempty by the intermediate value theorem, contradicting injectivity of f. Hence $f(x) < f(y)$, which means that f is strictly increasing. The case $f(a) > f(b)$ is analogous.

Now assume that f is strictly increasing but f^{-1} is not. Then there exists $x, y \in [c, d]$ such that $x < y$ but $f^{-1}(x) \geq f^{-1}(y)$. Applying the strictly increasing function f to $f^{-1}(x)$ and $f^{-1}(y)$ yields $x = f(f^{-1}(x)) \geq f(f^{-1}(y)) = y$, which is a contradiction. Again, the case of strictly decreasing functions is analogous.

For part b) let us consider the case that f is strictly increasing. Moreover let $(y_n)_{n=1}^{\infty}$ be a monotonic sequence in $[c, d]$ converging to s. Then $(f^{-1}(y_n))_{n=1}^{\infty}$ is a monotonic sequence in $[a, b]$, and by the monotone convergence theorem we have $(f^{-1}(y_n))_{n=1}^{\infty} \to r$ for some $r \in [a, b]$. Since f is continuous we obtain $y_n = f(f^{-1}(y_n)) \to f(r)$. Moreover we have $y_n \to s$, which means $s = f(r)$ since f is bijective. Hence $f^{-1}(s) = r$.

Now let $\varepsilon > 0$ be arbitrary. Considering $y_n = s - \frac{1}{n}$ we obtain $n_1 \in \mathbb{N}$ such that $f^{-1}(s - \frac{1}{n_1}) \geq r - \varepsilon$. Similarly, considering $y_n = s + \frac{1}{n}$ we obtain $n_2 \in \mathbb{N}$ such that $f^{-1}(s + \frac{1}{n_2}) \leq r + \varepsilon$. If we pick $\delta = \min(\frac{1}{n_1}, \frac{1}{n_2})$, then using monotonicity of f^{-1} we get $f^{-1}((s - \delta, s + \delta)) \subset (r - \varepsilon, r + \varepsilon)$. That is, for $x \in [c, d]$ with $|x - s| < \delta$ we get $|f^{-1}(x) - f^{-1}(s)| < \varepsilon$. In other words, f^{-1} is continuous at s. Since s was arbitrary, this shows that f^{-1} is continuous. □

Uniform continuity

In this subsection we discuss *uniform continuity*. Uniform continuity is a stronger property than continuity, its definition is obtained by slightly reordering the ingredients in the definition of continuity.

Definition 5.24. A real function $f : \mathrm{dom}(f) \to \mathbb{R}$ is called uniformly continuous if for every $\varepsilon > 0$ there exists $\delta > 0$ such that for all $c, x \in \mathrm{dom}(f)$ we have that $|c - x| < \delta$ implies $|f(c) - f(x)| < \epsilon$.

We can rewrite this using quantifiers: the function f is uniformly continuous if

$$\forall \varepsilon > 0, \exists \delta > 0 \text{ s.t. } \forall x, c \in \mathrm{dom}(f), |x - c| < \delta \implies |f(x) - f(c)| < \varepsilon.$$

As before, the value of δ will usually depend on ε. However, δ needs to be independent of $c \in \mathrm{dom}(f)$. In contrast, in the definition of continuity the number δ may depend not only on ε, but also on the point c.

From this discussion we obtain the following theorem[32].

[32] I'll leave it to you to write down a detailed proof: this is a good excercise in working with quantifiers!

Theorem 5.25. *Any uniformly continuous function is continuous.*

Let us have a look at some examples.

Example. *Check* directly from the definition *that the function* $f : \mathbb{R} \to \mathbb{R}$ *given by* $f(x) = 3x + 7$ *is uniformly continuous.*

Solution. In order to check the definition of uniform continuity let $\varepsilon > 0$. Then for $c, x \in \mathbb{R}$ we have

$$|f(x) - f(c)| = |3x + 7 - (3c + 7)| = |3x - 3c| = 3|x - c|.$$

Therefore, $|f(x) - f(c)| < \varepsilon$ is equivalent to $3|x - c| < \varepsilon$, or $|x - c| < \varepsilon/3$. If we chose $\delta = \varepsilon/3$ we conclude that $|x - c| < \delta$ implies $|f(x) - f(c)| < \varepsilon$. Thus f is uniformly continuous. □

We already know that all polynomial functions are continuous. However, such functions need not be uniformly continuous, as the following example illustrates.

Example. *Show that the function* $f : \mathbb{R} \to \mathbb{R}$ *given by* $f(x) = x^2$ *is not uniformly continuous.*

Solution. Let $\varepsilon = 1$. We claim that there is no $\delta > 0$ such that $|x - c| < \delta$ for $x, c \in \mathbb{R}$ implies $|f(x) - f(c)| < 1$. Indeed, notice that

$$|f(x) - f(c)| = |x^2 - c^2| = |x - c||x + c|.$$

So if we set $x = \frac{1}{\delta}, c = \frac{1}{\delta} + \delta/2$ we obtain

$$|f(x) - f(c)| = \frac{\delta}{2}\left(\frac{2}{\delta} + \frac{\delta}{2}\right) = 1 + \frac{\delta^2}{4} > 1,$$

and this shows that f is not uniformly continuous. □

If we restrict attention to functions defined on *bounded intervals* the situation changes significantly: continuous functions on bounded intervals are automatically uniformly continuous.

Theorem 5.26. *Any continuous function* $f : [a, b] \to \mathbb{R}$ *is uniformly continuous.*

Proof. Suppose that f is not uniformly continuous. Then there exists an $\varepsilon > 0$ such that for each $\delta > 0$ we find $x, c \in [a, b]$ such that $|c - x| < \delta$ and $|f(c) - f(x)| \geq \varepsilon$. In particular, applying this to $\delta = 1/n$ we find $x_n, c_n \in [a, b]$ such that $|c_n - x_n| < 1/n$ and $|f(c) - f(x)| \geq \varepsilon$. Since the sequence $(x_n)_{n=1}^{\infty}$ is bounded, the Bolzano-Weierstrass theorem shows that there exists a subsequence $(x_{k_n})_{n=1}^{\infty}$ such that $x_{k_n} \to c$ as $n \to \infty$ for some $c \in [a, b]$.

We claim that $(c_{k_n})_{n=1}^{\infty}$ also converges to c. To see this, consider

$$0 \leq |c_{k_n} - c| \leq |c_{k_n} - x_{k_n}| + |x_{k_n} - c| < \frac{1}{k_n} + |x_{k_n} - c| \to 0$$

as $n \to \infty$, using the triangle inequality in the first step. By the sandwich principle, the sequence $|c_{k_n} - c|$ converges to zero, which means precisely $c_{k_n} \to c$ as $n \to \infty$.

Since f is continuous at c, we have $f(x_{k_n}) \to f(c)$ and $f(c_{k_n}) \to f(c)$ as $n \to \infty$. This contradicts the fact that $|f(x_{k_n}) - f(c_{k_n})| \geq \varepsilon$ for all n according to our initial construction. Hence our assumption that f is not uniformly continuous was wrong. □