Contents

Proof, Sets and the Real Numbers

Logical Statements.

The foundation of mathematics (especially analysis) is logic. The fundamental building block of logic is logical statements.

A logical statement is something which is either true or false.

If A and B are statements, then $(A \text{ and } B)$ is a statement, which is true if A and B are both true, otherwise it is false.

$(A \text{ or } B)$ is also a statement. It is true if A is true, or if B is true, or if both A and B are true. It is only false if A is false and B is false. In everyday life, we use or to mean “one or the other but not both”, but in mathematics, “or” includes the possibility that both things are true. We might use “xor” (“exclusive or”) to denote “one or the other but not both”, but we will not be using this in this course.

$(\text{not } A)$ also written $(\neg A)$ is a statement as well. It is true if A is false and it is false if A is true. We call (not A) the negation of A.

A very important idea is logical implications. If I have 2 statements A and B such that whenever A is true, B is also true, we write $A \Rightarrow B$ “A implies B”.

For example, “I live in England” ⇒ “I live in the UK” since England is in the UK. Notice however that “I live in the UK” ⇒ “I live in England” is false because not everybody who lives in the UK, lives in England, for example someone who lives in Wales.

To give a more mathematical example, $n > 10 \Rightarrow n > 1$ because if a number is greater than 10, it is also bigger than 1.

Something worth noting is that $(A \Rightarrow B)$ is itself a statement, as it can be true or false.

$A \Rightarrow B$ means that whenever A is true B is true, so the negation of $A \Rightarrow B$ is that there is at least one instance in which A is true and yet B is false.

Notice that if $A \Rightarrow B$ then whenever A is true B is true, meaning if B were false, the only possibility is that A must be false (because if A were true then B would have been true, but it’s not). This means we can say $(B \text{ false}) \Rightarrow (A \text{ false})$ or $\neg B \Rightarrow \neg A$. This is called the contrapositive of $A \Rightarrow B$. Any implication is equivalent to it’s contrapositive.

We say that 2 statements A and B are equivalent when

$$A \Rightarrow B \text{ and } B \Rightarrow A$$

In this case, we write $A \Leftrightarrow B$ or we say “A is true if and only if B is true”. We also write “iff” (with 2 f’s) to mean “if and only if”.

This means that whenever A is true B is true, and whenever B is true, A is true.

If you need to prove that 2 statements A and B are equivalent, you can first assume that A is true and then show that B follows logically from it (this shows that $A \Rightarrow B$) then assume that B is true and show that A logically follows (this shows that $B \Rightarrow A$).

Alternatively we can use contraposition. Assume that A is true, the show that B logically follows (the same as before, showing that $A \Rightarrow B$). Then we can assume that A is false, then show that B cannot be true (this shows that $\neg A \Rightarrow \neg B$ and therefore [by contrapositive] $B \Rightarrow A$). Since we now have $A \Rightarrow B$ and $B \Rightarrow A$ we are done.

This section might seem a bit abstract as we discuss the ideas, but they are quite simple, and any confusions should be cleared away when you see some examples.

Sets and Functions

Sets are perhaps the most fundamental type of object in mathematics.

A set is a collection of elements. Elements can be any kind of mathematical object, including other sets, numbers, or even statements. Most of the sets we will be considering in this course will be sets of real numbers. We will define the real numbers soon, but for now we will use our intuitive idea that the real numbers is the set of all numbers on the number line, leaving no gaps.

We can define sets by simply writing out the elements they contain in braces:

$$A = \{1,2,3\}$$

Is the set containing 1, 2, 3 and nothing else. This is an example of a finite set. Some sets are infinite (like the set of real numbers, More on this later).

We denote the set of real numbers as $\mathbb{R}$.

We also have special letters for other sets. The set of "natural numbers", numbers like $1,2,3,\cdots$ is denoted by

$$\mathbb{N} \coloneqq \{1,2,3,\cdots\}$$

"$\coloneqq$" means "is defined to be equal to".

The set of "whole numbers" is just the set of natural numbers, but also with zero.

$$\mathbb{W} \coloneqq \{0,1,2,\cdots\}$$

The set of "integers" includes numbers like $0, 1, -1, 2, -2, \cdots$ is denoted by

$$\mathbb{Z} \coloneqq \{0, 1, -1, 2, -2, \cdots\}$$

We will define these sets more precisely later.

If we want to say that a certain element x is a member of a set S, we write $x \in S$

If S does not contain the element x, we write $x \notin S$

For example, $p \in \mathbb{R}$ means, "p is a member of the set of real numbers", i.e., "p is a real number."

We can also define a set by a rule by writing something like

$$\mathbb{R}^+ \coloneqq \{x \in \mathbb{R} |\ x > 0\}$$

Is this set of all real numbers x such that $x > 0$, i.e., the set of positive real numbers. The vertical bar reads as "such that". Sometimes we use a colon instead of a vertical bar.

Next, we have subsets.

If A and B are sets such that every element of A is also in B, we say that A is a subset of B. We write this as $A \subseteq B$.

More formally,

$$\text{if } x \in A \Rightarrow x \in B$$

$$\text{then } A \subseteq B$$

This is simply saying that if x being in A, implies that x is in B (every x in A is also in B) then A is a subset of B.

We say that the sets A and B are equal $(A = B)$ if every element in A is in B and every element of B is in A. In other words.

$$\text{if } A \subseteq B \text{ and } B \subseteq A$$

$$\text{then } A = B$$

Finally, (though this is not that important), we say that A is a "proper subset of B" if A is a subset of B but A is not equal to B. We write $A \subset B$

More formally:

$$\text{if } A \subseteq B \text{ and } A \neq B$$

$$\text{then } A \subset B$$

The difference between subsets and proper subsets will rarely (if ever) be useful for this course, but I leave it above for completeness.

We can combine 2 sets, A and B together to make new sets.

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

is the set of all elements which are in both A and B. This is called the "intersection" of A and B.

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

Is the set of all elements which are in either A or B (or both). This is called the "union" of A and B.

$$A/B \coloneqq \{x: x \in A \text{ and } x \notin B\}$$

is read as "A minus B" and is the set of all values which are in A but are not in B.

A function

$$f: D \to C$$

is a map between 2 sets D and C. We call D the domain and C the codomain.

f associates every element of D with some element of C.

We might write, if $x \in D$ then $f(x) \in C$. You may have heard of something called the "range" and be wondering what the difference between the range and the codomain is.

Every element of the domain is mapped onto some element of the codomain, but not every element of the codomain will necessarily be mapped to. The set of elements in the codomain which are *actually* mapped onto is called the "range" of f. The range is always a subset of the codomain.

A function is said to be "injective" if every element of the codomain is mapped to by at most one element of the domain. In other words, a function with the property that no 2 different inputs give the same output. Injective functions are sometimes called one-to-one (but this is not a good name for them).

A function is said to be "surjective" if every element in the codomain is mapped to by at least one element of the domain. In other words, if the codomain is equal to the range. Surjective functions are sometimes called "onto".

A function is said to be "bijective" if it is both injective and surjective. Bijective functions are sometimes called one-to-one-onto.

If a function is bijective then it is injective, meaning every element of the codomain has at most one element which maps to it, and surjective meaning every element of the codomain has at least one element which maps to it. This means that for a bijective function, every element of the codomain has *exactly* one element that maps onto it. The name "one-to-one" function would probably be more fitting for bijective functions, but it's best to stick with the terms "injective", "surjective" and "bijective" to avoid ambiguity.

The inverse of a function, $f: D \to C$, denoted by $f^{-1}: C \to D$ is a function which tells you which input to f gives you a particular output. More formally:

$$f(x) = y \text{ iff } f^{-1}(y) = x$$

$$\text{for every } x \in D \text{ and every } y \in C$$

For this to make sense, we need f^{-1} to map each element of y in C to exactly one element x in D.

More precisely, we need:

$$\text{for every } y \in C, \text{there exists one and only one } x \in D \text{ such that } f(x) = y$$

This is precisely the definition of f being bijective.

This means that f is "invertible" meaning it has an inverse, f^{-1} if and only if f is bijective.

See the problem sheet for some examples of what we can do about this if f is not bijective.

There are 2 symbols which we will use again and again throughout this course. ∃ (a backwards E) and ∀ (an upside-down A). We also often abbreviate "such that", as s.t.,

∃ means "there exists".

$$\exists_{n\in\mathbb{Z}}\ \text{s.t.}, n^2 = 4$$

Reads "there exists an integer, n such that $n^2 = 4$". This is of course true since $n = 2$ is an integer and $2^2 = 4$. Notice that $n = -2$ also works. ∃ does not require that there is only a single element with the property, it only requires that there is at least one.

∀ means "for all".

$$\forall_{n\in\mathbb{N}}\ n > 0$$

Reads "For all natural numbers, $n, n > 0$". This is true because the smallest natural number is 1. ∀ requires that the property is true for **every* object, ∃ requires it to be true for only at least one.

To negate (find the negation of) a statement involving "∀" we need to find a single counter example. The negation of "$\forall_x\ A(x)$" is "$\exists_x$ such that $\neg A(x)$" where $A(x)$ is some statement which depends on x.

Similarly, to negate a statement like "$\exists_x$ such that $A(x)$" we need to show that there exists no x such that $A(x)$. To do this we need to show that $A(x)$ is false no matter which x you choose. To be precise, the negation of "$\exists_x\ \text{s.t.}, A(x)$" is "$\forall_x\ \neg A(x)$".

When you negate a statement involving a chain of ∀'s and ∃'s with some condition $A(x)$ at the end, you can simply replace each ∀ with ∃ and replace every ∃ with ∀ and negate the condition at the end (replace $A(x)$ with $\neg A(x)$). (Think about why this is true, this is a very important idea).

Finally, we define the **image** of a set A under a function f as being the set of all possble outputs of f from inputs in A. We write the image of A under f as $f(A)$.

$$\text{Def: } f(A) \coloneqq \{f(x) | x \in A\}$$

Similarly we define the **pre-image** of B under f as:

$$\text{Def: } f^{-1}(B) \coloneqq \{x | f(x) \in B\}$$

(The set of inputs which give an output which lies in the set B.)

Axioms of an Ordered Field

The title of this section seems intimidating, but it's actually something which you will be familiar with.

I will start by being imprecise, then we will formalize the ideas using these "Field Axioms".

We will start with the natural numbers, $\mathbb{N} \coloneqq \{1,2,3,\cdots\}$

If we add any 2 natural numbers, we get another natural number. We say that $\mathbb{N}$ is "closed" under addition. We can also multiply and 2 natural numbers and always get a natural number, so $\mathbb{N}$ is closed under multiplication.

It is possible, to subtract two natural numbers and get something which is not a natural number. We say that $\mathbb{N}$ is not closed under subtraction.

We define the integers: $\mathbb{Z} \coloneqq \{0, 1, -1, 2, -2, \cdots\}$ to solve this problem. The integers are closed under addition, subtraction and still multiplication.

The integers are still not closed under division though, I can divide 2 integers, e.g., $1 \div 2$ and get something which is not an integer. We then define the rational numbers to solve this problem,

$$\mathbb{Q} \coloneqq \left\{\frac{m}{n} : m \in \mathbb{Z}, n \in \mathbb{N}\right\}$$

Is the set of all numbers which can be written as a ratio of integers. The reason why I have $n \in \mathbb{N}$ instead of $n \in \mathbb{Z}$ is to avoid n being 0 (more on this in a moment).

We don't need n to be negative since if we want a negative rational number, we can just let m be negative.

The rational numbers are closed under addition, subtraction, multiplication, and almost division. The reason I say "almost" is because we still can't divide anything by zero. This is a fundamental unavoidable problem. I will show you why I little later on.

We are now ready to define a field.

A field, F is a set with some notion of addition $(+)$ and multiplication $(\times)$, (which is closed under both) and follows the following properties (axioms):

The following are true for all $a, b, c \in F$ unless otherwise stated.

Axioms of addition:

Commutativity: $a + b = b + a$

Associativity: $(a + b) + c = a + (b + c)$

Identity: $\exists_{0 \in F}$ s.t., $\quad \forall_{a \in F}, \quad a + 0 = a$

we call this "0" the "additive identity"

Inverses: $\forall_{a\in F}, \exists_{-a\in F}$ s.t., $a + (-a) = 0$

we call this "$-a$" the "additive inverse" of a

(we often write $a \times b$ as $a \cdot b$ or simply ab)

Axioms of multiplication:

Commutativity: $ab = ba$

Associativity: $(ab)c = a(bc)$

Identity: $\exists_{1\in F}$ s.t., $\quad \forall_{a\in F}, \quad a \cdot 1 = a$

We call this "1" the "multiplicitive identity"

Inverses: $\forall_{a\in F \text{ with } a\neq 0} \exists_{a^{-1}\in F}$ s.t., $\quad aa^{-1} = 1$

We call a^{-1} the "multiplicitive inverse" of a.

0 is not required to have a multiplicitve inverse as this leads to problems (as we will see later).

Axiom of distributivity:

$$a \times (b + c) = ab + ac \ \forall_{a,b,c\in\mathbb{F}}$$

which connects multiplication and addition

The integers do not form a field because not every integer has a multiplicative inverse. E.g., There is no integer which I can multiply by 2 to get 1.

The rational numbers do form a field (CHECK YOURSELF!)

The set $\{0\}$ also forms a field. (so long as we let $0 = 1$). There is no rule which says that $0 \neq 1$. This is called the trivial field, and it is of no interest to us. To avoid this field, we will add a new axiom:

Axiom of non triviality:

$$0 \neq 1$$

Now, the “simplest” field is (in some sense) $\{0,1\}$ where we define:

$$0+0=0, \qquad 0+1=1, \qquad 1+0=1, \qquad 1+1=0$$

$$0\cdot 0=0, \qquad 0\cdot 1=0, \qquad 1\cdot 0=0, \qquad 1\cdot 1=1$$

When we defined fields I said we needed "some notion" of addition and multiplication. These don't need to be addition and multiplication in the traditional sense, so long as they obey the field axioms. One of the problems on the problem sheet is to verify that this is indeed a field.

This field is cyclic in the sense that repeatedly adding 1 causes you to go in a cycle $0,1,0,1,\cdots$

To avoid this we want some notion of ordering. We want some way to say which of 2 elements is less than the other. How should we define the statement $a<b$? Intuitively, we know what this means, but we need to make it precise. The key is to define "less than", not in terms of what it *is*, but rather in terms of what it *does*. We define an "ordered field" as a field with some notion of $(<)$ which satisfies the following 4 axioms of order:

Axioms of order:

Trichotomy: $\forall_{a,b\in F}$, exactly one of the following is true: $a<b, \qquad a=b, \qquad b<a$

Transitivity: $(a<b \text{ and } b<c) \Rightarrow a<c$

Monotony of addition: $\forall_{a,b,c\in F}, \qquad a<b \Rightarrow a+c<b+c$

Monotony of multiplication: $\forall_{a,b,c\in F}$ with $0<c, \qquad a<b \Rightarrow ac<bc$

The last 2 of these axioms essentially say that, you can add any number to both sides of an inequality, and you can multiply both sides of an inequality by any positive number.

We then can define $a>b$ as being equivalent to the statement $b<a$.

We also define $a\leq b$ as $(a<b$ or $a=b)$ and $a\geq b$ as $(a>b$ or $a=b)$

The simplest "ordered field" is in some sense the rational numbers, $\mathbb{Q}$, I will justify this claim in the next section.

Properties of an Ordered Field

From the axioms of an ordered field which we discussed previously, we can derive all of the familiar properties of the rational numbers. We will derive some of them in this section and the rest will be left as problems on the problem sheet.

One familiar property is that anything multiplied by 0 is 0. This will be the first theorem of our course.

$$\text{Theorem 1.1: } 0 \cdot a = 0\ \forall_{a \in F}$$

pf:

$$0 \cdot a$$

$$= (0 + 0) \cdot a \text{ (0 is an additive identity)}$$

$$= 0 \cdot a + 0 \cdot a \text{ (distributivity)}$$

So far we have:

$$0 \cdot a = 0 \cdot a + 0 \cdot a$$

$0 \cdot a \in F$ so it must have an additive inverse $-0 \cdot a$

We add this to both sides:

$$0 \cdot a + (-0 \cdot a) = 0 \cdot a + 0 \cdot a + (-0 \cdot a)$$

$$0 = 0 \cdot a + 0 \text{ (since an element plus it's additive inverse is 0)}$$

$$0 = a \cdot 0 \text{ (since 0 is the additive identity)}$$

And we have $a \cdot 0 = 0$ which was what we wanted.

You may be thinking, "if we need to derive everything from the axioms like this, this will become very tedious very quickly". The good news is that once we have proved a theorem, we can use it again whenever we like. We can now use the fact that $a \cdot 0 = 0$ without proof (because we already proved it). As we continue to prove more results, our life will become much easier because we can build upon what we have already done.

Deciding the correct order in which to prove statements can be tricky. This is a common theme in real analysis, and is one reason why it took many decades for mathematicians to first get this all done. That's also why I'm here, to guide you in the right direction.

The next thing we will prove is that $-1 \times a = -a$, (hopefully a familiar statement).

$$\text{Theorem 1.2: } -1 \times a = -a \text{ for every } a$$

pf:

The definition of -1 is that $1 + (-1) = 0$ so we ought to use that fact.

The definition of $-a$ is that $a + (-a) = 0$

Taking the first equation and multiplying by a:

$$1 + (-1) = 0$$

$$\big(1 + (-1)\big)a = 0a$$

$$1a + (-1)a = 0$$

$$\text{using distributivity and that } 0a = 0 \text{ (our first theorem)}$$

$$a + (-1)a = 0$$

But by definition $-a$ is the number such that $a + (-a) = 0$

$(-1)a$ fits this description and hence $-1 \times a = -a$ as required.

Next, I want you to pause and try this yourself before seeing what I do. As a hint, we will need to use the theorem we just proved, $-1 \times a = -a$

I want us to prove that $(-1)(-1) = 1$

$$\text{Theorem 1.3: } (-1)(-1) = 1$$

pf:

$$(-1) \times (-1) = -(-1), \text{which is, by definition, the additive inverse of } (-1)$$

$$1 + (-1) = 0 \text{ (by definition)}$$

$$(-1) + 1 = 0 \text{ (commutativity of addition)}$$

$$\text{so } 1 \text{ is the additive inverse of } (-1)$$

$$\text{i.e.}, -(-1) = 1$$

$$\text{meaning } (-1) \times (-1) = 1 \text{ as required.}$$

I now want you to use this result to prove that $(-a)(-a) = a^2$ where a^2 means $a \times a$.

Also try to prove this one, we will use it in a moment.

$$\text{Theorem 1.4: if } a > 0 \text{ then } -a < 0$$

$$\text{and if } a < 0 \text{ then } -a > 0$$

pf:

If $a > 0$, we can use monotony of addition to add $(-a)$ to both sides

$$a + (-a) > 0 + (-a)$$

$$0 > -a, \qquad \text{i.e.}, -a < 0$$

If $a < 0$, we do the same, $a + (-a) < 0 + (-a) \Rightarrow 0 < -a \Rightarrow -a > 0$

Finally:

Theorem 1.5:

If $a \neq 0$ then $a^2 > 0$

pf:

We want to use trichotomy, either $a < 0$ or $a > 0$. We will tackle each case separately.

If $a > 0$

$$a \cdot a > 0 \cdot a \text{ (monotony of multiplication)}$$

$$a^2 > 0$$

We can use monotony of multiplication since $a > 0$

if $a < 0$

We can't use monotony of multiplication since we can only multiply by something positive.

$-a > 0$ as we proved before

Now we can multiply by $-a$ which is positive.

$$(-a)(-a) > 0(-a)$$

$$a^2 > 0 \text{ (using } (-a)(-a) = a^2\text{)}$$

So, in both cases $a^2 > 0$, we are done.

Notice that $1 \neq 0$ and $1^2 = 1$ so $1 > 0$. This also means that $-1 < 0$.

There are other important facts for you to prove on the problem sheet, but for now we will move on.

We know that 0 and 1 must be elements of any ordered field with $0 \neq 1$.

Can we guarantee any more than this? Yes! As it will turn out, we can guarantee that our field consists of at least all the rational numbers, which is what I meant when I said that $\mathbb{Q}$ is the "simplest" ordered field.

First, we know that 0 must have an additive inverse. $0 + 0 = 0$ so it is it's own inverse. Nothing new yet. 0 need not have a multiplicative inverse. $1 \cdot 1 = 1$ so 1 is it's own multiplicative inverse. Still nothing new.

1 must have an additive inverse -1. We already showed that $1 > 0$ and hence $-1 < 0$. This means that (by transitivity) $-1 < 1$. Then by trichotomy, $-1 \neq 0$ and $-1 \neq 1$. This means that -1 is genuinely a brand-new number!

We have this ordering $-1 < 0 < 1$

I know that (since the field must be closed under addition, that $1 + 1 \in F$. Also (since $0 < 1$), $1 < 1 + 1$ meaning (again by transitivity and trichotomy) that $1 + 1$ must be a new number also. We will denote it by 2.

We can then similarly define $3 \coloneqq 2 + 1$ etc.,

Each of these must have additive inverses, $-2, -3, \cdots$

We now have integers!

Then each of these (except 0) must have a multiplicative inverse. In particular,

$$\frac{1}{1}, \frac{1}{2}, \frac{1}{3} \cdots \in F$$

Finally, since F is closed under multiplication, we must have, for example $-3 \times \frac{1}{2} \in F$

We denote this as $-\frac{3}{2}$. This covers all rational numbers. As we have already seen (from the problem sheet). The rational numbers are an ordered field, so we can stop here. This means that the rational numbers are the "simplest" ordered field in some sense.

Before finishing this section, I will demonstrate why allowing 0 to have a multiplicative inverse would be a bad idea.

$0 \cdot 0^{-1} = 1$ (by definition).

But we already showed that 0 times anything is 0. This is a contradiction.

Types of Proof and the Need for Real Numbers

Proving mathematical statements is the entire point of Analysis. So far we have seen direct proofs, where we start with some axioms, or information and show the conclusion follows logically from them.

This is what every proof does, but it isn't always so easy to prove something in such a direct way.

One very powerful proof technique is called induction. Induction allows us to prove statements about natural numbers. The idea is as follows.

Let $A(n)$ be a statement which depends on n.

$A(1)$ means the statement is true when $n = 1$

$A(n) \Rightarrow A(n+1)$ means that when the statement is true for some value, n, it is also true for $n + 1$.

Since it is true for 1, it must be true for $1 + 1 = 2$. Since it is true for 2, it must also be true for $2 + 1 = 3$, etc.,

It must be true for every natural number.

We show that the statement is true for 1. We assume that the statement is true for some n, then show that this implies it must be true for $n + 1$. This assumption is called the "inductive hypothesis".

One important use of induction is proving results about sums.

To introduce sigma notation:

$$\sum_{k=a}^{b} f(k) \coloneqq f(a) + f(a+1) + f(a+2) + \cdots + f(b-1) + f(b)$$

Where a and b are integers with $a \leq b$. To be more precise in this definition:

$$\sum_{k=a}^{a} f(k) = f(a), \qquad \sum_{k=a}^{b+1} f(k) = \sum_{k=a}^{b} f(k) + f(b+1)$$

Sigma notation itself has a definition which is reminiscent of induction, which is why induction works so well for it.

Here is an example.

$$\text{Prove } \sum_{k=1}^{n} 1 = n$$

Of course this is obvious, $1 + 1 + \cdots + 1$ with n 1's added together is n. But to prove it rigorously, we use induction. First $A(1)$. When $n = 1$:

$$\text{LHS} = \sum_{k=1}^{1} 1 = 1$$

$$\text{RHS} = 1$$

LHS means "left-hand side"

RHS, right-hand side

So, it is true for $n = 1$

Next, we assume it is true for some n,

$$\sum_{k=1}^{n} 1 = n$$

We are NOT assuming that it is true for every n, only for one particular arbitray value of n.

We next need to show that (given it is true for n) it must be true for $n + 1$.

$$\text{We want to show that } \sum_{k=1}^{n+1} 1 = n + 1$$

$$\sum_{k=1}^{n+1} 1 = \sum_{k=1}^{n} 1 + 1 = n + 1$$

As required.

As a harder example, we will prove the binomial theorem. You are probably familiar,

$$(a + b)^n = \sum_{k=0}^{n} a^k b^{n-k} \binom{n}{k}$$

Where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ (n choose k).

I should also define the factorial function.

$$0! = 1, \qquad n! = n(n-1)!, \qquad \text{for } n \in \mathbb{N}$$

And x^n:

$$x^0 = 1, \qquad x^n = x \cdot x^{n-1}, \qquad \text{for } n \in \mathbb{N}$$

Again, defined inductively.

The binomial theorem is a statement which involves a whole number n, so we can do induction (though we should start at 0 instead of 1 because n could be 0)

For this induction, we will first need to prove Pascal's identity, which is just to say, we need to prove that $\binom{n}{k}$ follows the pattern of Pascals' triangle:

$$1$$

$$1\ 1$$

$$1\ 2\ 1$$

$$1\ 3\ 3\ 1$$

$$1\ 4\ 6\ 4\ 1$$

$$\vdots$$

To construct pascals triangle, you add adjacent numbers and write the sum in the next row.

$$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$$

We can prove this identity using a direct proof, with some simple algebra (see problem sheet).

We are now ready to prove the binomial theorem. I urge you to try this yourself as this is a very good exercise in induction. I would have put this on the problem sheet, but I will do it here because it is such an important theorem:

$$\text{Binomial Theorem (1.6): } (a+b)^n = \sum_{k=0}^{n} a^k b^{n-k} \binom{n}{k} \ \forall_{n \in \mathbb{W}}$$

$$\text{Basis: when } n = 0:$$

$$\text{LHS} = (a+b)^0 = 1$$

$$\text{RHS} = (a+b)^n = \sum_{k=0}^{0} a^k b^{0-k} \binom{0}{k} = a^0 b^{0-0} \binom{0}{0}$$

$$= \binom{0}{0} = \frac{0!}{0!\,(0-0)!} = \frac{1}{1 \cdot 1} = 1 = \text{LHS}$$

$$\text{Hence true for } 0 \ \big(A(0)\big)$$

$$\text{Assume true for some } n,$$

$$\text{Show true for } n+1\text{: (It's often good to start with the more complicated side)}$$

$$\sum_{k=0}^{n+1} a^k b^{n+1-k} \binom{n+1}{k}$$

$$= \sum_{k=0}^{n} a^k b^{n+1-k} \binom{n+1}{k} + a^{n+1} b^0 \binom{n+1}{n+1} = \sum_{k=0}^{n} a^k b^{n+1-k} \binom{n+1}{k} + a^{n+1}$$

$$= b^{n+1} + \sum_{k=1}^{n} a^k b^{n+1-k} \binom{n+1}{k} + a^{n+1} \text{ to get some symmetry}$$

We want to get $\sum_{k=0}^{n} a^k b^{n-k} \binom{n}{k}$ so that we can apply the inductive hypothesis.

Replace k by $k+1$

$$b^{n+1} + \sum_{k=0}^{n-1} a^{k+1} b^{n-k} \binom{n+1}{k+1} + a^{n+1}$$

$\sum_{k=a}^{b} f(k) = \sum_{k=a-1}^{b-1} f(k+1) = f(a) + \cdots + f(b)$

Use Pascal's identity:

$$b^{n+1} + \sum_{k=0}^{n-1} a^{k+1} b^{n-k} \binom{n}{k} + \sum_{k=0}^{n-1} a^{k+1} b^{n-k} \binom{n}{k+1} + a^{n+1}$$

$$= b^{n+1} + \sum_{k=0}^{n-1} a^{k+1} b^{n-k} \binom{n}{k} + \sum_{k=1}^{n} a^{k} b^{n-k+1} \binom{n}{k} + a^{n+1}$$

$$= \left(a^{n+1} + \sum_{k=0}^{n-1} a^{k+1} b^{n-k} \binom{n}{k}\right) + \left(b^{n+1} + \sum_{k=1}^{n} a^{k} b^{n-k+1} \binom{n}{k}\right)$$

$$= \sum_{k=0}^{n} a^{k+1} b^{n-k} \binom{n}{k} + \sum_{k=0}^{n} a^{k} b^{n-k+1} \binom{n}{k}$$

$$= a \sum_{k=0}^{n} a^{k} b^{n-k} \binom{n}{k} + b \sum_{k=0}^{n} a^{k} b^{n-k} \binom{n}{k}$$

$$= (a+b) \sum_{k=0}^{n} a^{k} b^{n-k} \binom{n}{k}$$

$$= (a+b)(a+b)^n$$

$$= (a+b)^{n+1}$$

Hence $A(n) \Rightarrow A(n+1)$

And we are done.

$$(a+b)^n = \sum_{k=0}^{n} a^k b^{n-k} \binom{n}{k}, \qquad \forall_{n\in\mathbb{N}} \text{ as required}$$

Another important method of proof is proof by contradiction.

Sometimes it can be difficult to prove that a statement, A is true and so it might be easier to prove A is not false. We assume that the statement is untrue (that is, we assume $\neg A$), and then we reach some kind of contradiction (either to an axiom, a theorem we have already proved, the assumption itself, or something which follows from the assumption). Anything at all which makes A being false logically impossible. This means that A must be true.

A classic example is the proof that $\sqrt{2}$ is irrational. This is also a very important theorem for our purposes, as $\sqrt{2} \notin \mathbb{Q}$ is an undesirable property of the rational numbers and so we want to fix it. The solution is to extend the rational numbers to the real numbers (which ought to include $\sqrt{2}$) but we are getting ahead of ourselves.

$$\text{Theorem 1.7 }\left(\text{Irrationality of }\sqrt{2}\right)\text{: There exists no } x \in \mathbb{Q} \text{ such that } x^2 = 2$$

Proof:

$$\text{Assume for the sake of contradiction, that there does exists such an } x.$$

We can write:

$$x = \frac{p}{q}$$

where p and q are integers with no common factors.

If they did have common factors, we could just cancel out highest common factor to get it into this form. Then

$$x^2 = 2 = \left(\frac{p}{q}\right)^2$$

$$2 = \frac{p^2}{q^2}$$

$$2q^2 = p^2$$

$$2q^2 \text{ is even so } p^2 \text{ is even}$$

p is an integer so it is either odd or even.

Even numbers can be written as $2k$ and $(2k)^2 = 4k^2 = 2 \cdot 2\mathrm{k}^2$ which is even

Odd numbers can be written as $2k + 1$ and $(2k + 1)^2 = 4k^2 + 4k + 1$

$= 2(2k^2 + 2k) + 1$ which is odd.

This tells us that since p^2 is even, p must also be even (since if p were odd, p^2 would be odd).

Therefore we can write p^2 as $4k^2$ for some integer k.

$$2q^2 = 4k^2$$

$$q^2 = 2k^2$$

$2k^2$ is even so q^2 must be even. By the same reasoning as before, q is even

We now have that p and q each have a factor of 2 which contradicts the assumption that they had no common factors. This means that the assumption was wrong (remember the assumption was that there is a rational number who's square is 2). We have shown that there is no such number, in other words, $\sqrt{2} \notin \mathbb{Q}$.

The rational numbers are very well behaved in many regards, but we have now found a limitation with them. There are numbers which intuitively make sense to us, but the rational numbers are not able to describe them. We can approximate such numbers by rational numbers, but there will always be gaps. We want to (roughly speaking) define the real numbers as being a set which fills in these gaps.

The Real Numbers and the Axiom of Completeness

We so far have 14 axioms which we want any familiar set of numbers to follow. We called these the axioms of a non-trivial ordered field.

To tie this down to just the real numbers, we will need one more. The "Axiom of Completeness".

There are many different equivalent axioms we could use here, but I will use one which is commonly used because it is also one of the easiest to do things with (though they are all equally as reasonable). Before stating this axiom, I need to define a few concepts first.

$$\text{Defn: A set of real numbers } S \subseteq \mathbb{R} \text{ is } \textbf{bounded above} \text{ iff}$$

$$\exists_{M\in\mathbb{R}} \text{ s.t.,} \qquad \forall_{x\in S}, \qquad M \geq x$$

In other words, if there exists a real number which is greater than or equal to every element of the set. If this is the case, we call M an "upper bound" for S. Note that a set can have multiple upper bounds. Similarly, we define:

$$\text{Defn: A set of real numbers } S \subseteq \mathbb{R} \text{ is } \textbf{bounded below} \text{ iff}$$

$$\exists_{m\in\mathbb{R}} \text{ s.t.,} \qquad \forall_{x\in S}, \qquad m \leq x$$

And we call m a "lower bound" for S.

We call a set **bounded** iff it is both bounded above and bounded below.

An important idea is that of a "least upper bound", which is an intuitive idea:

$$\text{Defn: Given a set } S \subseteq \mathbb{R}, \text{we call } \alpha \text{ a } \textbf{least upper bound} \text{ for } S \text{ iff:}$$

$$\alpha \text{ is an upper bound for } S \text{ and for every upper bound } M, \alpha \leq M$$

[α is the Greek letter alpha]

The "least upper bound" is often referred to as the **supremum** of S, and we write $\alpha = \sup S$.

The **infimum** or "greatest lower bound" has a similar definition.

$$\text{Defn: Given a set } S \subseteq \mathbb{R}, \text{we call } \beta \text{ an } \textbf{infimum} \text{ for } S \text{ iff:}$$

$$\beta \text{ is a lower bound for } S \text{ and for every lower bound } m, \beta \geq m$$

And we write $\beta = \inf S$

Supremums and infimums are unique, meaning if a set has a supremum it has only one and likewise for infimum. (Problem sheet).

We are now ready for the "axiom of completeness":

Axiom of Completeness:

Every non-empty set of real numbers which is bounded above, has a least upper bound.

With this 15th axiom, we have now fully defined our real numbers.

We will progress by using this axiom to prove a very obvious fact.

$$\text{Theorem of Archimedes (1.8)}: \forall_{x \in \mathbb{R}}, \qquad \exists_{n \in \mathbb{N}} \text{s.t.}, \qquad n > x$$

That is, given any real number x, I can always find a natural number, n which is greater than x. This statement seems very obvious, but we will prove it from the axioms nonetheless.

pf:

We will proceed by contradiction; we start by supposing that the statement is not true.

Its negation is:

$$\exists_{x \in \mathbb{R}}, \qquad \text{such that } \forall_{n \in \mathbb{N}}, \qquad x \geq n$$

We swapped the ∃ with ∀ and vice-versa. We also negated the statement at the end. The negation of $n > x$ is $x \geq n$ (by trichotomy).

If we think about what this negated statement is saying, it says that there exists a real number which is greater than or equal to every natural number. This is precisely the definition of the natural numbers being bounded above. We know that $1 \in \mathbb{N}$ so the set of natural numbers is not empty.

Since $\mathbb{N}$ is now a non-empty set which is (supposedly) bounded above, we can apply the axiom of completeness:

$$\exists_{\alpha \in \mathbb{R}} \text{ such that } \alpha = \sup \mathbb{N}$$

Supremums have 2 properties, we will want to make use of them both.

First α is an upper bound, meaning $\alpha \geq n$ for every $n \in \mathbb{N}$.

Second, α is less than or equal to all upper bounds, meaning there are no upper bounds less than α.

$$\alpha - 1 < \alpha$$

Meaning $\alpha - 1$ is not an upper bound. What does that mean? It means there is at least one $n \in \mathbb{N}$ such that $\alpha - 1 < n$ (since if this weren't so, then $\alpha - 1$ would have been an upper bound).

$$\alpha - 1 < n \Rightarrow \alpha < n + 1$$

But now we use the fact that the natural numbers are closed under addition. $n \in \mathbb{N}$ and $1 \in \mathbb{N}$ which means that $n + 1 \in \mathbb{N}$ and so we have a natural number which is greater than α. This contradicts the first fact, that α is an upper bound for the natural numbers.

This contradiction means that the assumption (the negation of the theorem) was false, meaning the theorem is true.

i.e.,

$$\text{for every real number } x, \text{there exists a natural number } n \text{ such that } n > x$$

Before finishing this section, we want to ensure that the axiom of completeness actually solves the problem of $\sqrt{2}$ not being a rational number. We want to show that $\sqrt{2}$ is real.

$$\text{Theorem 1.9: there exists a positive } x \in \mathbb{R} \text{ such that, } x^2 = 2$$

pf:

Of course, to prove this, we will need to use our axiom of completeness.

It tells us that for any non-empty, bounded above set, there exists a real number which is it's least upper bound. We want to create a set who's upper bound is $\sqrt{2}$ and then, by completeness it must be real.

First, we need a candidate for our set which we think will have a least upper bound of $\sqrt{2}$. Can you think of such a set?

After some thinking a good suggestion might be

$$S = \{x \in \mathbb{R}: x^2 < 2\}$$

The set of real numbers whose square is less than 2. It doesn't seem unreasonable that $\sup S = \sqrt{2}$.

The idea is to let $\alpha = \sup S$, then to show that $\alpha^2 = 2$ (and $\alpha > 0$). Which will meet our conditions.

First, we need to show that S is non-empty and bounded above. To show non-empty, we just need a single element which is in S.

$$1 \in \mathbb{R} \text{ and } 1^2 = 1 < 2 \text{ so } 1 \in S$$

Hence non-empty.

To show bounded above, I need to find an upper bound. It seems clear that 2 should be an upper bound. I need to show that for every element x in S, that $x \leq 2$

$$\text{Let } x \in S$$

$$x^2 < 2 < 4 < 2^2$$

$$\text{so } x^2 < 2^2$$

$$\Rightarrow x < 2$$

So S is bounded above. We now have permission to let $\alpha = \sup S$.

Showing that $\alpha > 0$ is easy. We know that α is an upper bound for S and $1 \in S$ and so $\alpha \geq 1 > 0$.

We know that α is definitely a real number (by the axiom of completeness), so α^2 is also real. So how do we show that $\alpha^2 = 2$ (which completes the proof).

The idea will be to assume that $\alpha^2 < 2$ and that $\alpha^2 > 2$ and reach a contradiction in both cases. This means that (by trichotomy) $\alpha^2 = 2$ as required.

I will show how $\alpha^2 < 2$ leads to a contradiction. The other case will be left as an exercise on the problem sheet.

$$\text{Suppose } \alpha^2 < 2$$

We need to get a contradiction. Remember, α has 2 defining properties. The fact that it is an upper bound, and the fact that there are no upper bounds less than α. Which of these 2 properties does it seem like $\alpha^2 < 2$ would violate?

We supposed that $\alpha^2 < 2$ but we think this will lead to a contradiction, because it should actually be bigger. We think we have chosen a value of α too small, meaning that it is no longer an upper bound. So this is the property which we will try to violate.

To be precise, we want an element of S which is greater than α (and so α is not an upper bound which gets us our contradiction).

The key idea here is that we want to find a number which is bigger than α, but only by the tiniest amount, so that we can guarantee that it is still in S. To capture this idea of a number which is just **barely** bigger than α, I will consider $x = \alpha + \frac{1}{n} > \alpha$. We can make n as large as we like and hence make x as close to α as we need (but x will always be bigger than α no matter how large n is).

I now have a number which is greater than α, $(x > \alpha)$. I just need to show that (by choosing the appropriate value of n) we can get $x \in S$ and then we will be done.

For $x \in S$, we need $x^2 < 2$ meaning:

$$\left(\alpha + \frac{1}{n}\right)^2 < 2$$

Which is equivalent to (by expanding brackets):

$$\alpha^2 + \frac{2\alpha}{n} + \frac{1}{n^2} < 2$$

Remember our goal is to show that there is some n such that the above inequality is always true (whenever $\alpha^2 < 2$). The $\frac{1}{n^2}$ term is making things more difficult so we can get rid of it by noticing that $\frac{1}{n^2} \leq \frac{1}{n}$ which means

$$\alpha^2 + \frac{2\alpha}{n} + \frac{1}{n^2} \leq \alpha^2 + \frac{2\alpha}{n} + \frac{1}{n} = \alpha^2 + \frac{2\alpha + 1}{n}$$

Which is good because this new expression $\left(\alpha^2 + \frac{2\alpha+1}{n}\right)$ has no $\frac{1}{n^2}$'s involved which will make the next step much easier.

Crucially, notice that if I can show that

$$\alpha^2 + \frac{2\alpha + 1}{n} < 2$$

Then automatically (by transitivity):

$$\alpha^2 + \frac{2\alpha}{n} + \frac{1}{n^2} \leq \alpha^2 + \frac{2\alpha + 1}{n} < 2$$

$$\Rightarrow \alpha^2 + \frac{2\alpha}{n} + \frac{1}{n^2} < 2$$

$$\Rightarrow \left(\alpha + \frac{1}{n}\right)^2 < 2$$

Which was what we wanted. So, all I need to do now (which is surprisingly easy), is to show that (with the right choice of n),

$$\alpha^2 + \frac{2\alpha + 1}{n} < 2$$

Which is equivalent to:

$$\frac{2\alpha + 1}{n} < 2 - \alpha^2$$

We can multiply by n (since $n \in \mathbb{N}$ is positive) and we can divide by $2 - \alpha^2$ (because $\alpha^2 < 2$ meaning $2 - \alpha^2 > 0$).

$$\frac{2\alpha + 1}{2 - \alpha^2} < n$$

The question becomes, can I find an n which is greater than $\frac{2\alpha+1}{2-\alpha^2}$? The answer is immediately yes (by the theorem of Archimedes). I don't need to know what that n is, it is sufficient to just know that it exists.

To remind ourselves of what we have just done, using this value of n, we have that $\left(\alpha+\frac{1}{n}\right)^2 < 2$ which means that $\alpha+\frac{1}{n} \in S$ but $\alpha+\frac{1}{n} > \alpha$ which contradicts the fact that α is an upper bound, which means that the assumption (which was that $\alpha^2 < 2$) was false.

I will leave the rest of the proof to you, it is very similar. You will want to show that $\alpha^2 > 2$ contradicts that α is the *least* upper bound, by fining an upper bound less than α.

We then conclude that $\alpha^2 = 2$ and $\alpha > 0$ is a real number, we denote α by $\sqrt{2}$ and hence $\sqrt{2} \in \mathbb{R}$ which shows that the completeness axiom does at least fulfil the purpose it was created for (it will also do much more for us as we move forward).

Intervals

In this short section, I want to define a particularly useful type of set.

Def: An interval I is a set such that, $\forall_{a,b\in I}$ with $a < b$, $a < c < b \Rightarrow c \in I$

That is, for every a and b in the interval, anything between a and b is also in the interval.

Intervals might take the following forms:

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

$$(a, b] \coloneqq \{x \in \mathbb{R}: a < x \leq b\}$$

$$[a, b) \coloneqq \{x \in \mathbb{R}: a \leq x < b\}$$

$$[a, b] \coloneqq \{x \in \mathbb{R}: a \leq x \leq b\}$$

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

$$(-\infty, b] \coloneqq \{x \in \mathbb{R}: x \leq b\}$$

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

$$[a, \infty) \coloneqq \{x \in \mathbb{R}: a \leq x\}$$

$$(-\infty, \infty) \coloneqq \mathbb{R}$$

(Check from the definition that these are all intervals).

We use regular brackets, (),to exclude an endpoint. We use square brackets, [], to include an endpoint. Whenever we use $\pm\infty$ we use a regular bracket, because we cannot include $\pm\infty$ as it is not a real number. Also if both brackets are () we call it an open interval, if both are [], we call it a closed interval.

The Modulus Function

The modulus function is a very simple function which we will be using a *lot* throughout this course. It is very simple, but very versatile. We denote the modulus of x by $|x|$. An informal definition is that the modulus function keeps positive things the same, and makes negative things positive. A more formal definition is:

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

If I put in something non-negative, it just returns itself back, e.g., $|2| = 2$, $\left|\frac{17}{5}\right| = \frac{17}{5}$ and $|0| = 0$.

If I put in something negative, it gets multiplied by -1 (making it positive).

e.g., $|-3| = -(-3) = 3$ and $|-5| = 5$.

An important idea is that we can use the modulus function to consider the "distances" between two real numbers. We can find the difference between two numbers (a and b) by subtracting ($a - b$). But this will give us a negative number when $b > a$. If we don't care about which is bigger, only the difference between them, we can use the modulus function.

$|a - b|$ can be thought of as the "distance" between two real numbers a and b.

In this section we will discuss some very important (and useful) properties of this function which we will be using a lot later on.

Because the modulus function is defined in this "piecewise" way, which means we define it one way for some values and another way for other values, means we can easily prove facts about it just my considering different cases. For example:

$$\text{Theorem 1.10: } |a||b| = |ab|$$

This seems rather obvious, but to be certain, we ought to prove it. There are four cases to consider; $a, b \geq 0$; $a, b < 0$; $a < 0,\ b \geq 0$; $a \geq 0, b < 0$. We can easily prove the statement by handling these four cases separately.

pf:

$$\text{Case 1: } a \geq 0 \text{ and } b \geq 0 \text{ so } |a| = a \text{ and } |b| = b$$

$$\text{We have that } ab \geq 0 \text{ (monotony of multiplication) so } |ab| = ab$$

$$\text{LHS} = |a||b| = ab$$

$$\text{RHS} = |ab| = ab$$

$$\text{so LHS} = \text{RHS}$$

$$\text{Case 2: } a < 0 \text{ and } b < 0 \text{ so } |a| = -a \text{ and } |b| = -b$$

$$\text{in this case, } ab > 0 \text{ so } |ab| = ab$$

$$\text{LHS} = -a \times -b = ab$$

$$\text{RHS} = ab$$

$$\text{So LHS} = \text{RHS}$$

$$\text{Case 3: } a < 0 \text{ and } b \geq 0$$

$$|a| = -a, \qquad |b| = b$$

$$ab \leq 0 \Rightarrow |ab| = -ab$$

$$\text{LHS} = |a||b| = -a \times b = -ab$$

$$\text{RHS} = |ab| = -ab$$

$$\text{So LHS} = \text{RHS}$$

There is also the fourth case, where $a \geq 0$ and $b < 0$, but we can ignore this case because it is equivalent to case 3, because the original equation is symmetric in a and b (we can swap a with b and it doesn't change). $|a||b| = |ab|$ is equivalent to $|b||a| = |ba|$. We often say "without loss of generality" or "WLOG", in a situation like this where the two cases are equivalent and so we can get away with only considering one of them.

Proving other theorems involving the modulus function is very similar. I will prove one more important theorem here.

$$\text{Theorem 1.11: "Triangle Inequality"}$$

$$|a + b| \leq |a| + |b|$$

This statement becomes clear when you try some various numbers. The reason why it is called the "triangle inequality" is because of a geometric interpretation, that being that the shortest distance between two points is a straight line. The relation between that fact and this "triangle inequality" becomes more apparent when considering complex numbers (which we will not be doing in this particular course).

pf:

We consider cases. Again, since the statement does not change when you swap a and b, we can "WLOG" ignore the fourth case.

Case 1: $a, b \geq 0 \Rightarrow a + b \geq 0$

$a, b \geq 0 \Rightarrow |a| = a, \qquad |b| = b$

$a + b \geq 0 \Rightarrow |a + b| = a + b$

$\text{LHS} = |a + b| = a + b$

$\text{RHS} = |a| + |b| = a + b$

LHS = RHS so case 1 works.

Case 2: $a, b < 0 \Rightarrow a + b < 0$

$a, b < 0 \Rightarrow |a| = -a, \qquad |b| = -b$

$a + b < 0 \Rightarrow |a + b| = -(a + b) = -a - b$

$\text{LHS} = -a - b$

$\text{RHS} = -a - b$

LHS = RHS so case 2 works

Case 3: $a < 0$ and $b \geq 0$

$|a| = -a$ and $|b| = b$

We don't know whether $a + b$ is ≥ 0 or < 0, so we need to consider these 2 subcases separately

Case 3a: $a + b \geq 0$

$\text{LHS} = |a + b| = a + b$

$\text{RHS} = |a| + |b| = -a + b$

Since $a < 0, -a > 0$ meaning $a < -a$ and hence $a + b < -a + b$

LHS < RHS so case 3a works

Case 3b: $a + b < 0$

$\text{LHS} = |a + b| = -a - b$

$\text{RHS} = |a| + |b| = -a + b$

$b \geq 0$ meaning $b \geq -b$ and so $-a - b \leq -a + b$

again LHS $\leq$ RHS and so case 3b works.

All cases work, so we have finished the proof.

Further to the idea that we can use $|a-b|$ to represent the “distance” between a and b on the number line, if this distance is less than, say c, then this means that a is between $b-c$ and $b+c$. To make this more precise:

$$\text{Theorem 1.12: } |a-b|<c \Leftrightarrow b-c<a<b+c$$

pf:

We want to prove that this goes in both directions. If the LHS is true then the RHS is true, and if the RHS is true then the LHS is true. I will begin with the “⇒” direction.

$$\text{Let } |a-b|<c$$

We will deal with this by cases, as usual.

$$\text{Case 1: } a \geq b$$

$$\text{in this case, } a-b \geq 0 \Rightarrow |a-b| = a-b$$

$$\text{so } a-b<c \Rightarrow a<b+c$$

We have showed that $a<b+c$ but we also need that $a>b-c$

$$0 \leq |a-b| < c \Rightarrow 0<c$$

$$\text{then } b-c<b \leq a \text{ hence } b-c<a$$

$$\text{we now have } b-c<a<b+c$$

$$\text{Case 2: } a<b \Rightarrow a-b<0 \Rightarrow |a-b| = b-a$$

$$\text{so } b-a<c \Rightarrow b-c<a$$

We still need to show that $a<b+c$.

$$0 \leq |a-b| < c \Rightarrow 0<c$$

$$\text{then } a<b<b+c \text{ and hence } a<b+c$$

$$\text{so we have } b-c<a<b+c \text{ in both cases.}$$

Next for the “⇐” direction of this proof.

$$\text{Let } b-c<a<b+c$$

$$\Rightarrow -c<a-b<c$$

$$\text{if } a-b \geq 0$$

$$\text{then } a-b<c \Rightarrow |a-b|<c$$

$$\text{if } a - b < 0$$

$$\text{then } -c < a - b \Rightarrow b - a < c \Rightarrow |a - b| < c$$

$$\text{so in any case: } |b - a| < c$$

We are done.

Sequences

The "Limit" of a Sequence

A sequence is a list of numbers in a particular order.

E.g.,

$$\left[1,2,3,17,-\frac{34}{11},-8,0\right]$$

Is a sequence.

We will be dealing with infinite sequences, meaning sequences with an infinite number of terms in them. One way to describe such a sequence is to define it's n^{th} term. For example, the sequence $\frac{1}{n}$ has a first term of 1, a second term of $\frac{1}{2}$ and a 27^{th} term of $\frac{1}{27}$.

$$\frac{1}{n}=\left[1,\frac{1}{2},\frac{1}{3},\frac{1}{4},\frac{1}{5},\cdots\right]$$

We often write something like $a_n=\frac{1}{n}$ to refer to such a sequence, and then (for example) $a_5=\frac{1}{5}$.

The main topic of our conversation has to do with whether such sequences "converge". We all have a vague idea of what it means for a sequence to "converge" to a particular value. For example, the sequence $\frac{1}{n}$ is clearly "converging" to 0. The terms get closer and closer to 0 as you make n larger and larger. The sequence $[1.49, 1.51, 1.499, 1.501, 1.4999, 1.5001, \cdots]$ (assuming the obvious pattern continues) ought to converge to 1.5 (if we make a sensible definition). The tough part is in deciding what a *precise* definition would be for such an abstract idea.

Suppose we have a sequence a_n, converging to some number L. Intuitively this means that by going far enough into the sequence, we can get the terms to be as close to L as we like. For example, there must be some term of the sequence which is within 0.1 of L. There must also be a term within 0.01 of L, or 0.0000000001 of L. The idea is that I can choose any positive number and there will be a term of the sequence which is within that distance of the limit L. Remember, we can define the "distance" between two numbers (in this case a_n and L) by using the modulus function as $|a_n-L|$.

Our current (incomplete) definition might be written as:

$$\lim a_n = L \text{ means:}$$

For any positive number, there exists a term of the sequence, a_n such that $|a_n - L|$ is less than the given number.

We may write this more formally by giving the positive number a name, let's call it ε (Greek letter epsilon).

$$\forall_{\varepsilon>0}, \qquad \exists_{n\in\mathbb{N}} \text{ s.t.}, \qquad |a_n - L| < \varepsilon$$

There is a problem with this definition, which is that it only requires a single term which is close to L. What we *really* want is not just that there exists a term which is close to L, but that the sequence *stays* that close to L.

So instead of saying "there exists a term, a_n, such that $|a_n - L| < \varepsilon$", we should be saying "there exists a term, a_N, such that for every term a_n after a_N, we have $|a_n - L| < \varepsilon$". This captures that idea of "staying close" to L. To make it more formal, this is our (final) definition.

$$\lim a_n = L \text{ means}$$

$$\forall_{\varepsilon>0}, \qquad \exists_{N\in\mathbb{N}} \text{ s.t.}, \qquad \forall_{n\geq N}, \qquad |a_n - L| < \varepsilon$$

This is called the ε-N Definition of the limit.

You should take some time to really make sure you understand this definition because we will be using it a lot, and having a good intuitive understanding will make it easier to understand what we do as we proceed.

If a sequence does not converge to any limit, we say that it diverges. This might mean that it "goes to $\pm\infty$", or that it oscillates between 2 or more values, or any number of other things. For the case of "diverging to ∞", we can be more precise. This definition is almost identical to the definition of convergence, but we replace the idea of a_n being "close to L" with the idea of a_n being "large".

$$\text{Defn: A sequence } a_n, \text{diverges to } \infty \text{ iff:}$$

$$\forall_{\varepsilon>0}, \qquad \exists_{N\in\mathbb{N}} \text{ s.t.}, \qquad \forall_{n\geq N}, \qquad a_n > \varepsilon$$

The important thing to note here is that this works for any $\varepsilon > 0$. So far we have interpreted this as meaning "no matter how small, as long as it is positive". Here however, we interpret is as "no matter how large". Mathematically it makes no difference, but often we use a different symbol to remind ourselves that it represents a large number, rather than a small one. Also, there is no need for the number to necessarily be positive. Hence, we typically write, for divergence to ∞:

$$\forall_{M\in\mathbb{R}}, \qquad \exists_{N\in\mathbb{N}} \text{ s.t.}, \qquad \forall_{n\geq N}, \qquad a_n > M$$

If a sequence, a_n, diverges to ∞, we write $\lim a_n = \infty$. Remember ∞ is not a real number, we are just using it here as a symbol to represent the idea of divergence to infinity.

Using the Definition of Convergence

Here, we will go use the definition of Convergence to prove that certain functions converge to their expected limits, and to prove that particular sequences diverge (do not converge to any limit). We will also demonstrate an important theorem which makes determining the limits of many sequences much easier.

$$\text{e.g.}, \lim \frac{1}{n} = 0$$

pf: The sequence of $\frac{1}{n}$ clearly converges to 0 (if our definition is to reflect our intuitive idea). So, we shall prove this fact.

Proofs involving the ε-N definition of the limit often follow a similar route to what I will demonstrate here.

Recall the definition:

$$\forall_{\varepsilon>0}, \qquad \exists_{N\in\mathbb{N}} \text{ s.t.}, \quad \forall_{n\geq N}, \qquad |a_n - L| < \varepsilon$$

Where $a_n = \frac{1}{n}$ and $L = 0$, so the inequality becomes:

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

Which is the same thing as

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

We need to show that a particular statement is true for any $\varepsilon > 0$. To do this, we start by saying,

$$\text{Let } \varepsilon > 0$$

Then we want to show that the rest of the statement must follow, regardless of which ε was chosen.

We need to show that there exists a natural number N with a certain property. We don't have enough information yet to determine what this N is for now. We will simply write:

$$\text{choose } N \text{ such that} \cdots$$

and then come back later to fill the gap once we have a better idea.

Then we need something to be true for all $n \geq N$, so I will say:

$$\text{Let } n \geq N$$

Finally, I need to show that $|a_n - L| < \varepsilon$. Of course, I already explained that this is the same as $\frac{1}{n} < \varepsilon$.

I know that $\frac{1}{n} \leq \frac{1}{N}$ (by dividing $n \geq N$ by nN) and so if I can get $\frac{1}{N} < \varepsilon$, I will have

$$\frac{1}{n} \leq \frac{1}{N} < \varepsilon$$

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

And we will be done. The question then is, how can I ensure that $\frac{1}{N} < \varepsilon$ (this is where we decide what the condition on N must be. This condition will be met if $N > \frac{1}{\varepsilon}$ which is always possible by the theorem of Archimedes. We use this to fill in the gap from before, and we are done.

This is what the proof looks like when written out in full:

$$\text{Let } \varepsilon > 0$$

$$\text{Choose } N > \frac{1}{\varepsilon} \text{ (possible by archimedes)}$$

$$\text{Let } n \geq N$$

$$|a_n - L| = \left|\frac{1}{n} - 0\right| = \frac{1}{n} \leq \frac{1}{N} < \varepsilon$$

$$\Rightarrow |a_n - L| < \varepsilon$$

$$\text{Hence } \lim \frac{1}{n} = 0$$

The hardest part of ε-N proofs is often determining how to choose N, and how to simplify the inequality at the end into something easier to work with.

$$\text{e.g., } \lim k = k$$

Where k is some constant

pf: I will write down the definition again, this time with $a_n = k$ and $L = k$.

$$\forall_{\varepsilon>0}, \quad \exists_{N\in\mathbb{N}} \text{s.t.}, \quad \forall_{n\geq N}, \quad |k - k| < \varepsilon$$

Again, we simplify the inequality at the end, which becomes

$$0 < \varepsilon$$

Which is always true (since $\varepsilon > 0$). It doesn't matter which value of N we choose, so let's just

$$\text{Choose } N = 1$$

$$\text{Let } n \geq N$$

$$|a_n - L| = 0 < \varepsilon$$

$$\text{Hence} \lim k = k$$

Not every sequence will converge to a particular limit. Some sequences may grow to infinity (more on this later), or they might oscillate between various values (or neither of the above).

$$\text{e.g.}, \lim(-1)^n \text{ does not exist}$$

When a sequence does not converge, we say that it's limit "does not exist" or "DNE".

$(-1)^n = -1$ when n is odd and 1 when n is even. This certainly doesn't seem to converge to anything in particular, so let's make certain by using the definition.

pf:

We want to show that $(-1)^n$ doesn't converge to anything, we want to show that the ε-N definition does not hold for any potential limit L.

We want:

$$\forall_{\varepsilon>0}, \qquad \exists_{N\in\mathbb{N}} \text{ s.t.}, \qquad \forall_{n\geq N}, \qquad |(-1)^n - L| < \varepsilon$$

to be false, for every $L \in \mathbb{R}$ which means we need to prove that the *negation* of the above is true. Recall that we get the negation by swapping the ∀'s and ∃'s and negating the final statement. We want to show that:

$$\forall_{L\in\mathbb{R}}, \qquad \exists_{\varepsilon>0} \text{ s.t.}, \qquad \forall_{N\in\mathbb{N}}, \qquad \exists_{n\geq N} \text{ s.t.}, \qquad |(-1)^n - L| \geq \varepsilon$$

We have a statement for all real numbers L, so

$$\text{Let } L \in \mathbb{R}$$

I only need to show that there exists a single $\varepsilon > 0$ such that the rest is the case. I just need to choose an appropriate ε. Since a_n can only take on values of -1 and 1, the reason why we suspect that no limit can exist is because it would need to be close to 1 and -1 at the same time. The difference between -1 and 1 is 2 and so if ε is half of that ($\varepsilon = 1$), it should work (since L can't be within that distance of -1 and 1 at the same time). Any smaller value of ε would work just as well.

$$\text{Let } \varepsilon = 1 > 0$$

I now need to show that the rest is true for every $N \in \mathbb{N}$ and so:

$$\text{Let } N \in \mathbb{N}$$

Then we have "there exists $n \geq N$" so I need to choose an appropriate one.

$$\text{Choose } n \cdots$$

I will come back to fill in the gaps once we have a better idea.

Finally, I need to show that for this n,

$$|(-1)^n - L| \geq 1$$

[Since $\varepsilon = 1$]

If $L = 0$, this is automatically true since $|(-1)^n - 0| = |(-1)^n| = 1 \geq 1$

If $L > 0$, we can make it true if $(-1)^n = -1$ (n is odd).

If $L < 0$, we can make it true *if* $(-1)^n = 1$ (n is even)

This tells us how we should have chosen n in the last step, so I will fill in the gap now.

$$\text{If } L \geq 0 \text{ then choose } n \geq N \text{ such that } n \text{ is odd}$$

$$\text{If } L < 0 \text{ then choose } n \geq N \text{ such that } n \text{ is even}$$

Then we have 2 cases:

$$L \geq 0 \Longrightarrow n \text{ is odd} \Rightarrow (-1)^n = -1, \qquad \text{also } L + 1 > 0$$

$$|(-1)^n - L| = |-1 - L| = |L + 1| = L + 1 \geq 1 \text{ as required}$$

$$L < 0 \Longrightarrow n \text{ is even} \Rightarrow (-1)^n = 1, \qquad \text{also } 1 - L > 0$$

$$|(-1)^n - L| = |1 - L| = 1 - L > 1 \text{ as required}$$

$$\text{Hence } \lim(-1)^n \text{ does not exist.}$$

ε-N proofs can be difficult. Whilst they are necessary to start with (as we have done here), we can make our lives easier by proving the following (relatively clear) statements about limits.

$$\text{If } \lim a_n = a \;\&\; \lim b_n = b \text{ Then:}$$

$$\lim(a_n + b_n) = a + b$$

$$\lim(a_n b_n) = ab$$

$$\text{If Additionally, } b \neq 0 \text{ and } b_n \neq 0 \text{ for any } n \text{ (or at least for any suffiently large } n\text{):}$$

$$\lim \frac{a_n}{b_n} = \frac{a}{b}$$

Where the $\neq 0$ conditions are simply to prevent division by zero.

The cases of $[\lim(a_n - b_n) = a - b]$ and $[\lim ka_n = ka]$ follow immediately from these results.

I will show you the proof for the addition and multiplication. The rest will be left as exercises.

$$\text{Theorem 2.1: } \lim a_n = a, \qquad \lim b_n = b, \qquad \Longrightarrow \lim(a_n + b_n) = a + b$$

pf:

First step (as always):

$$\text{Let } \varepsilon > 0$$

We know that $\lim a_n = a$ which (loosely) means we can make $|a_n - a|$ as small as we like. We can also make $|b_n - b|$ as small as we like.

Our goal is to show that we can also make $|a_n + b_n - a - b|$ as small as we like.

As typical, we will try to simplify the final inequality $|a_n + b_n - a - b| < \varepsilon$ into something easier to use. It would be useful if we could write it in terms of $|a_n - a|$ and $|b_n - b|$ as we have control over those (we can make them both small). For this we can use the triangle inequality.

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

So if I can show that $|a_n - a|$ and $|b_n - b|$ can each be made $< \frac{\varepsilon}{2}$, then it will follow that $|a_n + b_n - a - b| < \varepsilon$.

I know that since $\lim a_n = a$ there is an $N_1 \in \mathbb{N}$ such that $\forall_{n \geq N_1}$ $|a_n - a| < \frac{\varepsilon}{2}$

Also, since $\lim b_n = b$, there exists an $N_2 \in \mathbb{N}$ such that $\forall_{n \geq N_2}$ $|b_n - b| < \frac{\varepsilon}{2}$

Now I'm ready to choose my N.

$$\text{Let } N = \max(N_1 N_2)$$

This means that $\forall_{n \geq N}$, we have that $n \geq N_1$ and $n \geq N_2$ which means we can use both that $|a_n - a| < \frac{\varepsilon}{2}$ and $|b_n - b| < \frac{\varepsilon}{2}$.

Finally, we have:

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

$$\text{i.e.}, |a_n + b_n - (a + b)| < \varepsilon$$

Hence $\lim(a_n + b_n) = a + b$

The multiplication case it a little bit more tricky:

As you will see when we do the proof, there is another (rather clear) fact which we will need to use, which is that all convergent sequences are bounded.

$$\text{Lemma 2.2: if } a_n \text{ is convergent then } a_n \text{ is bounded.}$$

(A Lemma is just a small theorem which we are proving because it will make a future theorem much easier).

pf:

$$\text{Let } a_n \text{ be a convergent sequence with limit } L.$$

Intuitively it seems that a_n ought to be bounded because there are infinitely many terms which are "close to" L and only finitely many which are not close to L. By "close to", I mean "within ε of" where we can choose any $\varepsilon > 0$. Let's go with $\varepsilon = 1$. (You can use any, but I am just picking something).

We know now that (by definition of the limit) there exists an $N \in \mathbb{N}$ such that $\forall_{n \geq N}: |a_n - L| < 1$.

This means that $L - 1 < a_n < L + 1$ whenever $n \geq N$.

This gives us an upper and lower bound on all of the terms within the "tail end" of the sequence. We can form an easy upper bound for the rest, because the rest is only finitely many.

Let $M = \max(\{a_n: 1 \leq n \leq N - 1\} \cup \{L + 1\})$.

Then $M \geq a_n$ for $1 \leq n \leq N - 1$, also $M \geq L + 1 > a_n$ for $n \geq N$. This means $M \geq a_n$ for all n and hence M is an upper bound for a_n.

Similarly, $m \coloneqq \min(\{a_n: 1 \leq n \leq N - 1\} \cup \{L - 1\})$ is a lower bound for a_n. Hence a_n is bounded, as required.

Notice also that if I let $C = \max\{|M|, |m|\}$ then $C \geq 0$ and $|a_n| \leq C$, meaning if a sequence is bounded we can say $\exists_{C \geq 0}$ such that $|a_n| < C$. Notice that if C works for this, anything $> C$ would also work. If $C = 0$, then $C = 1$ would also have worked. We can always find a $C > 0$.

$\exists_{C > 0}$ such that $|a_n| < C$ for all $n \in \mathbb{N}$.

$$\text{Theorem 2.3: } \lim a_n = a, \lim b_n = b \Rightarrow \lim a_n b_n = ab$$

pf:

$$\text{Let } \varepsilon > 0$$

Before we choose an appropriate N, let's consider the final inequality

$$|a_n b_n - ab| < \varepsilon$$

And how we can simplify it. We want to get something in terms of $|a_n - a|$ and $|b_n - b|$ since these are two expressions which we know we can make arbitrarily small.

To achieve this, we will use a common trick, which is to add something, and immediately subtract it away.

$$|a_n b_n - ab| = |a_n b_n - a_n b + a_n b - ab|$$

We could have used ab_n, both would work. Next, we factor:

$$|a_n b_n - ab| = |a_n(b_n - b) + b(a_n - a)|$$

And use the triangle inequality:

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

Don't forget that our goal is to make this thing $< \varepsilon$.

We can make $|a_n - a|$ as small as we like, so let's choose $N_1 \in \mathbb{N}$ such that $\forall_{n \geq N_1}$, we have $|a_n - a| < \frac{\varepsilon}{2|b|}$ (which is okay so long as $b \neq 0$. We will handle the case with $b = 0$ separately at the end). This means that $|b||a_n - a| < \frac{\varepsilon}{2}$ whenever $n \geq N_1$.

We can also make $|b_n - b|$ as small as we like, but we have a problem, which is that $|a_n|$ is not a constant. How can we be sure that, as n gets larger, $|a_n|$ doesn't grow without bound? This is not a problem because we already know that convergent sequences are bounded. The means there exists a real number $C > 0$ such that $|a_n| \leq C$ for every n.

$$|a_n||b_n - b| \leq C|b_n - b|$$

We can then choose an $N_2 \in \mathbb{N}$ such that $\forall_{n \geq N_2}$ we have that $|b_n - b| < \frac{\varepsilon}{2C}$

This means that $|a_n||b_n - b| < C\frac{\varepsilon}{2C} = \frac{\varepsilon}{2}$

Finally, I choose $N = \max(N_1, N_2)$. Then $\forall_{n \geq N}$, we have $n \geq N_1$ and $n \geq N_2$ which means

$$|a_n b_n - ab| \leq |a_n||b_n - b| + |b||a_n - a| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

Hence $\lim a_n b_n = ab$

Now to handle the case with $b = 0$.

The inequality $|a_n b_n - ab| < \varepsilon$ becomes $|a_n||b_n| < \varepsilon$

Now since a_n is convergent, it is bounded, so again we have $C > 0$ with $|a_n| \leq C$ for every n. Also $\lim b_n = b = 0$ so there exists an $N \in \mathbb{N}$ such that $\forall_{n \geq N}$ we have that $|b_n - 0| = |b_n| < \frac{\varepsilon}{C}$

Then:

$$|a_n b_n - ab| = |a_n b_n - a \cdot 0| = |a_n||b_n| \leq C|b_n| < \frac{C\varepsilon}{C} = \varepsilon$$

Meaning $\lim a_n b_n = 0 = ab$ as required.

For the case of $\frac{a_n}{b_n}$, (which I will leave to you), it is sufficient to show that $\lim \frac{1}{b_n} = \frac{1}{b}$, because then you can just use the multiplication of limits to get $\lim \frac{a_n}{b_n} = \lim \left(a_n \cdot \frac{1}{b_n}\right) = a \cdot \frac{1}{b} = \frac{a}{b}$.

We can use these theorems to evaluate all kinds of limits which would have otherwise been very difficult.

$$\text{e.g.}, \lim \frac{1}{n^3} = \lim \frac{1}{n} \cdot \frac{1}{n} \cdot \frac{1}{n} = 0 \cdot 0 \cdot 0 = 0$$

You can now easily prove by induction that $\lim\limits_{n \to \infty} \frac{1}{n^k} = 0$ for any $k \in \mathbb{N}$.

$$\text{e.g.}, \lim \frac{n^4 - 3n^3 + 2n^2 + 17n - 18}{4n^4 + 11n^3 - 13n + 1} = \lim \frac{1 - \frac{3}{n} + \frac{2}{n^2} + \frac{17}{n^3} - \frac{18}{n^4}}{4 + \frac{11}{n} - \frac{13}{n^3} + \frac{1}{n^4}}$$

$$= \frac{1 - 0 + 0 + 0 - 0}{4 + 0 - 0 + 0} = \frac{1}{4}$$

Important Theorems for Convergence

So far, we have looked at how to determine the limits of particular sequences, but using the ε-N definition is only possible if we already have a candidate for what we think the limit should be. Using the theorems from the end of the last section, we can combine known limits together, but we can't do this if the limits which make up the larger expression are not known.

This next theorem will allow us to deduce that a certain type of sequence converges, without necessarily knowing what the limit is.

First, I need make a quick definition.

A sequence a_n is said to be $\mathbf{monotone\ increasing}$ if $a_n \leq a_m$ whenever $n \leq m$. This means that as you move forward in the sequence, the terms get larger (or can stay the same), but they can never get smaller.

Similarly, a_n is $\mathbf{monotone\ decreasing}$ if $a_n \geq a_m$ whenever $n \leq m$. As you move forward in the sequence, terms can never get larger.

We say that a_n is $\mathbf{monotone}$ if it is either monotone increasing or monotone decreasing.

We can now discuss the theorem:

$$\mathbf{Monotone\ Convergence\ Theorem}\ (2.4):$$

$$\text{Every monotone, bounded sequence is convergent.}$$

This seems obvious at first glance. It can't diverge to infinity (bounded), and it can't oscillate between different places because it can only go in one direction (monotone).

This theorem will be a very useful one.

pf:

I will assume that a_n is increasing. The proof for the case where it is decreasing is very similar and so I will leave it as an exercise to you.

Of course, the reason why this theorem will be useful is because it will allow us to say that certain sequences diverge even if we have no idea what the limit might be. In order to prove the theorem itself though, we will need a limit candidate. The least upper bound of a_n seems like a reasonable candidate.

$$\text{Let } a_n \text{ be bounded and monotone increasing}$$

$$\text{Since } a_n \text{ is bounded, it is bounded above.}$$

$$\text{Let } S = \{a_n : n \in \mathbb{N}\}$$

$$a_1 \in S \text{ so } S \text{ is non empty and bounded above}$$

$$\text{Let } \alpha = \sup S \;\; \text{(by axiom of completeness)}$$

We now need to show that a_n converges with limit α. We proceed by using the definition of the limit.

$$\text{Let } \varepsilon > 0$$

$$\text{Choose } N \cdots$$

We will come back later to choose the right N.

$$\text{Let } n \geq N$$

I need to show that $|a_n - \alpha| < \varepsilon$

First, we use that α is an upper bound, which means $\alpha \geq a_n$ and hence $a_n - \alpha \leq 0$

Our inequality becomes

$$\alpha - a_n < \varepsilon$$

I still need to use the fact that α is a least upper bound, which means that anything $< \alpha$ (even by a tiny amount) is not an upper bound. This gives me the idea to rearrange into:

$$a_n > \alpha - \varepsilon$$

Remember, we don't know that this inequality is true, we want to prove that it is true (whenever $n \geq N$).

I know that $\alpha - \varepsilon$ is not an upper bound, and so $\exists_{N \in \mathbb{N}}$ such that $a_N > \alpha - \varepsilon$ (this is how we should choose our N).

Then when I Let $n \geq N$, we have $a_n \geq a_N > \alpha - \varepsilon$ (since a_n is monotone increasing).

Finally, I rearrange $a_n > \alpha - \varepsilon$ into

$$\alpha - a_n (= |a_n - \alpha|) < \varepsilon$$

Hence, we are done.

The next important theorem I will prove is called the **squeeze theorem** sometimes called the sandwich or pinching theorem. It says that,

Squeeze Theorem (2.5):

$$\text{if } a_n \leq b_n \leq c_n \text{ for every } n$$

$$\text{and } \lim a_n = \lim c_n = L$$

$$\text{then } b_n \text{ converges and } \lim b_n = L$$

Or equivalently,

$$\text{if } |a_n| \leq b_n \text{ for every } n$$

$$\text{and } \lim b_n = 0$$

$$\text{then } \lim a_n = 0$$

I will prove the first version of this theorem, and I will leave it to you (problem sheet) to show that the second version immediately follows.

pf:

$$\text{Let } a_n, b_n, c_n \text{ be sequences with } a_n \leq b_n \leq c_n \text{ for every } n$$

$$\text{and with } L \coloneqq \lim a_n = \lim b_n$$

$$\text{Let } \varepsilon > 0$$

Our goal, as usual, is to find an $N \in \mathbb{N}$ such that $n \geq N \Rightarrow |b_n - L| < \varepsilon$, that is that we can get b_n to be within ε of L.

We know that we can make a_n and c_n as close to L as we like. The key idea of the proof is this:

If we can get both a_n and c_n to simultaneously be within ε of L then, since b_n is between a_n and c_n, this will imply that b_n is also within ε of L, which is what we want. To make this more precise:

$$\text{Take } N_1 \text{ such that } n \geq N_1 \Rightarrow |a_n - L| < \varepsilon$$

$$\text{Take } N_2 \text{ such that } n \geq N_2 \Rightarrow |c_n - L| < \varepsilon$$

We can now choose N,

$$\text{Let } N = \max(N_1, N_2)$$

$$\text{Let } n \geq N$$

$$\Rightarrow n \geq N_1 \text{ and } n \geq N_2$$

We know then, that $|a_n - L| < \varepsilon$ and $|c_n - L| < \varepsilon$

We want to show that this implies that $|b_n - L| < \varepsilon$ and we will be done.

If you consider visually what is happening here, we have a_n is between $L - \varepsilon$ and $L + \varepsilon$.

c_n is also between these values. We know that b_n is between a_n and c_n so it must also be in this interval. To be more formal:

$$|a_n - L| < \varepsilon \Leftrightarrow L - \varepsilon < a_n < L + \varepsilon$$

$$|c_n - L| < \varepsilon \Leftrightarrow L - \varepsilon < c_n < L + \varepsilon$$

We also know that $a_n \leq b_n \leq c_n$ so:

$$L - \varepsilon < a_n \leq b_n \leq c_n < L + \varepsilon$$

$$\Rightarrow L - \varepsilon < c_n < L + \varepsilon$$

$$\Rightarrow |c_n - L| < \varepsilon$$

We are done.

One final theorem I want to prove is very subtle, yet important. We already know that every convergent sequence is bounded, but specifically, I want to show that the limit must be within the same bounds as the sequence. To be more precise,

$$\text{Theorem 2.6: If } x_n \in [a, b] \text{ and } x_n \text{ converges to } L, \text{then } L \in [a, b]$$

There are probably many ways to go about proving this, but I am thinking contradiction might work out nicely.

pf:

$$\text{Let } x_n \text{ be a sequence with } \lim x_n = L \text{ and } x_n \in [a, b]$$

$$\text{Suppose (for contradiction) that } L > b$$

The idea is that since we can get x_n arbitrarily close to L, if $L > b$ we can find an $x_n > b$ which would contradict that $x_n \in [a, b]$.

$$\text{Let } \varepsilon = L - b \text{ (which is allowed since } L > b)$$

$$\text{There is an } N \in \mathbb{N} \text{ such that}, \forall_{n \geq N},$$

$$|x_n - L| < L - b$$

$$\Leftrightarrow b - L < x_n - L < L - b$$

$$\Leftrightarrow b < x_n < 2L - b$$

$$\text{Critically, we have } b < x_n \text{ which is a contradiction since } x_n \in [a, b]$$

$$\text{The assumption } (L > b) \text{ was false, meaning } L \leq b$$

$$\text{The next step is to show that } L \geq a$$

$$\text{We suppse then, than } L < a, \text{and proceed similarly}$$

I leave the second part as an exercise to you.

A corollary of this theorem ("corollary" means a theorem which follows quickly from a more difficult theorem which you just proved) is that if $x_n \in (a, b)$ then $L \in [a, b]$.

$$\text{pf: } x_n \in (a, b) \Rightarrow x_n \in [a, b] \Rightarrow L \in [a, b]$$

You may be wondering whether we can say that $x_n \in (a, b) \Rightarrow L \in (a, b)$. In fact, we cannot, as demonstrated by a simple counter-example: $\frac{1}{n} \in (0,2)$ but $\lim \frac{1}{n} = 0 \notin (0,2)$ (but $0 \in [0,2]$).

An important type of sequence is a geometric sequence. A geometric sequence is any sequence of the form:

$$x_n = ar^n$$

Intuitively, it seems like this sequence will diverge for $|r| > 1$, converge to 0 for $|r| < 1$, converge to a for $r = 1$ and oscillate between $-a$ and a (so diverge) for $r = -1$.

This is all assuming that $a \neq 0$. If $a = 0, x_n = 0$ is a constant sequence with limit 0.

Theorem 2.7 (Convergence of Geometric Sequences)

Let $x_n = ar^n$ with $a \neq 0$

Case 1: $|r| = 1$

Either $r = \pm 1$

Case 1a: $r = 1$

$x_n = a \cdot 1^n = a$ is a constant sequence with $\lim x_n = a$

Case 2a: $r = -1, \qquad x_n = a(-1)^n$

Suppose x_n converges with limit L, then $\frac{x_n}{a} = (-1)^n$ converges with limit $\frac{L}{a}$

we know that $(-1)^n$ diverges, this contradiction means x_n diverges.

Case 2: $|r| > 1$

We want to show that x_n is unbounded.

$$|x_n| = |ar^n| = |a||r|^n$$

$$|r| > 1 \Rightarrow |r| - 1 > 0$$

$$\text{Let } \delta \coloneqq |r| - 1 \Rightarrow |r| = 1 + \delta$$

$$|x_n| = |a|(1+\delta)^n \geq |a|(1+n\delta) \text{ (first 2 terms from binomial theorem)}$$

Which is unbounded above (diverges to ∞ in fact), hence x_n is unbounded and so it diverges.

Case 3: $|r| < 1$

I want to show that x_n converges to 0.

$$|r| < 1 \Rightarrow \frac{1}{|r|} > 1$$

$$\text{We know (from Case 2) that } \frac{1}{|r|^n} \to \infty \text{ which means that } |r|^n \to 0$$

[We have seen from the problem sheet that $a_n \to \infty \Rightarrow \frac{1}{a_n} \to 0$]

$$\text{Hence } r^n \to 0 \text{ and } \lim x_n = \lim ar^n = a \lim r^n = a \cdot 0 = 0$$

We are done.

I will finish this section with a miscellanious theorem relating to supremums, which will be slightly useful later on.

Theorem 2.8:

Let S be a set with $\sup S = \alpha$

Then there exists a sequence $x_n \in S$ with $\lim x_n = \alpha$

We can prove this constructively, meaning we will show that a sequence exists by showing how you could construct it.

pf:

We use the fact that $\alpha = \sup S$ means that anything $< \alpha$ is not an upper bound. This means there exists some $x \in S$ such that $\alpha - 1 < x \leq \alpha$. Call this x_1.

There is also some $x_2 \in S$ such that $\alpha - \frac{1}{2} < x_2 \leq \alpha$.

In fact since $\alpha - \frac{1}{n} < \alpha$, $\alpha - \frac{1}{n}$ is not an upper bound for S and so there exists some $x_n \in S$ such that $\alpha - \frac{1}{n} < x_n \leq \alpha$. There may be many (potentitally infinitely many) choices for each x_n but this does not matter, any choice will do. We now have some sequence $x_n \in S$ with $\alpha - \frac{1}{n} \leq x_n \leq \alpha$ for every $n \in \mathbb{N}$. We can then use the squeezing theorem:

$$\lim\left(\alpha - \frac{1}{n}\right) = \alpha, \qquad \lim \alpha = \alpha$$

$$\Longrightarrow \lim x_n = \alpha$$

Which was what we wanted.

Subsequences

You can form a **subsequence** of a sequence by removing any number of terms from the original sequence.

For example, $[1,13,57,109,\cdots]$ is a subsequence of $a_n = n = [1,2,3,4,5,\cdots]$, and $\left[\frac{1}{2},\frac{1}{4},\frac{1}{6},\cdots\right]$ is a subsequence of $b_n = \frac{(-1)^n}{n} = \left[-1,\frac{1}{2},-\frac{1}{3},\frac{1}{4},-\frac{1}{5},\cdots\right]$.

More formally:

A **subsequence** of x_n is any sequence x_{n_j} where n_j is a strictly increasing sequence of positive integers.

Strictly increasing means each term is strictly greater than the last, as in they cannot be equal.

To make this definition make more sense, I will give an example. Let $n_j = 2j - 1 = [1,3,5,7,\cdots]$ (which is strictly increasing).

Then $a_{n_j} = [a_1, a_3, a_5, \cdots]$ is subsequence of a_n (it is the subsequence consisting of the odd terms of a_n.

Note that the terms of the subsequence a_{n_j} are dependent on n_j which itself depends on j, meaning a_{n_j} depends on j.

There are 2 important theorems about subsequences which we will consider in this section. The first is a relatively obvious fact. The second is not so obvious, but isn't terribly counterintuitive either.

Firstly,

Theorem 2.9: The subsequence of a convergent sequence is also convergent and converges to the same limit as the original sequence.

pf:

We are given a convergent sequence;

Let a_n be a sequence with $\lim a_n = L$

Next we have a subsequence of a_n;

Let a_{n_j} be a subsequence of a_n (then n_j is a strictly increasing sequence of positive integers).

We need to show that $\lim\limits_{j\to\infty} a_{n_j} = L$. We use the ε-N definition.

$$\text{Let } \varepsilon > 0$$

We know that the sequence a_n converges to L, which means $\exists_{N\in\mathbb{N}}$ such that $\forall_{n\geq N}$ we have $|a_n - L| < \varepsilon$.

To complete the proof, we need to choose $J \in \mathbb{N}$ such that $\forall_{j\geq J}$, $\left|a_{n_j} - L\right| < \varepsilon$.

We need to choose J such that $n_J \geq N$, but is this always possible? We can prove by induction that $n_j \geq j$ for every j and so $n_N \geq N$. This means by choosing $J = N$ we get the following.

$$\text{Let } j \geq J$$

$$\Rightarrow n_j \geq n_J \geq J = N$$

$$n_j \geq N \Rightarrow \left|a_{n_j} - L\right| < \varepsilon$$

$$\text{Hence } \lim_{j\to\infty} a_{n_j} = L$$

And we are done. There is one gap to fill in this proof. I will leave it to you to prove that $n_j \geq j$ for every j (think first about why this is obvious and then prove it by induction).

We have seen before that every sequence which is convergent is bounded. It is not the case that every bounded sequence is convergent (e.g., $(-1)^n$).

But is there anything we can say regarding convergence and bounded sequences? Thinking vaguely for a moment, if I have a bounded sequence, we have a finite amount of space and an infinite number of terms. It seems that these points must all, in at least one place, cluster together.

The way to express this formally is:

The Bolzano-Weierstrass Theorem:

Every bounded sequence has a convergent subsequence

A common proof which I have seen for this theorem many times is to repeatedly split the interval on which the sequence is bounded in half infinitely many times, forming a subsequence which is what we call “Cauchy” (we will talk about Cauchy sequences later), which ends up giving us a convergent subsequence.

I will use a different approach, which I think is nicer, it also avoids any need for these so-called “Cauchy sequences”.

We have no idea what the subsequence we find will converge to, but we do know that it will be bounded (because the original sequence is bounded). The **Monotone Convergence Theorem** then, might be a good call. We need only to find a single **monotone** subsequence. It turns out that not only does every bounded sequence have a monotone subsequence, but in fact every *sequence* as a monotone subsequence.

$$\text{Lemma 2.10: Every sequence has a monotone subsequence.}$$

$$\text{pf:}$$

For this proof I will define something called a peak index.

$$\text{Def: a peak index of } a_n \text{ is a term } a_N \text{ such that } n > N \Rightarrow a_N > a_n$$

In other words, a term of the sequence is a peak index if it is bigger than every term which comes after it.

There are 2 possibilities. There are infinitely many peak indices or there are not.

Remember we are trying to find a monotone subsequence.

Case 1: There are infinitely many peak indices.

Let a_{N_j} be the jth peak index. This is an infinite sequence (since there are infinitely many peak indices). Since each a_{N_j} is a peak index, every term after it (including the next peak index) must be less than a_{N_j}. This means a_{N_j} is a (monotone) decreasing subsequence.

We have found a monotone decreasing subsequence in this case. In the case of finitely many peak indices, we will be able to find an increasing sequence.

Case 2: There are finitely many peak indices. (This includes their being none).

If there are no peak indices, let $a_{n_1} = a_1$. If there is at least one, let a_{n_1} be the term after the final peak index.

a_{n_1} is not a peak index (since either there are none, or it comes after the last one).

Critically there are no more peak indices ever again after this point in the sequence.

$$a_{n_1} \text{ is not a peak index} \Longrightarrow \text{there exists some } n_2 > n_1 \text{ with } a_{n_2} \geq a_{n_1}.$$

We also know that $a_{n_2}$is not a peak index so there is some $n_3 > n_2$ with $a_{n_3} \geq a_{n_2}$

Continue in this way:

$$a_{n_j} \text{ is not a peak index} \Rightarrow \exists_{n_{j+1} > n_j} \text{ s.t.,} \qquad a_{n_{j+1}} \geq a_{n_j}.$$

We form this subsequence a_{n_j} which is (though not strictly) monotone increasing.

We have now found a monotone increasing subsequence.

In any case, we can find a monotone subsequence of any sequence.

Corollary 2.11: **The Bolzano-Weierstrass Theorem**:

Every bounded sequence has a convergent subsequence

pf:

Let a_n be a bounded sequence.

There exists a monotone subsequence, a_{n_j} of a_n (by the Lemma we just proved).

a_{n_j} is bounded (since a_n is bounded)

By Monotone Convergence Theoerm, a_{n_j} converges.

And we are done.

Cauchy Sequences

I briefly mentioned Cauchy sequences earlier. In this section I will define them and prove an important fact about them.

We often find ourselves wanting to show that a sequence is convergent, despite not knowing what the limit is. This motivates us to come up with a new way of determining when a sequence is convergent without needing to know the value of the limit.

You may notice that the terms of a convergent sequence get closer and closer to each other as you progress in the sequence. Not only do consecutive terms get closer and closer, but in fact all terms get closer together. To be more precise, we define:

$$\text{defn: } a_n \text{ is a } \textbf{Cauchy Sequence} \text{ iff:}$$

$$\forall_{\varepsilon>0}, \quad \exists_{N\in\mathbb{N}}, \text{s.t.}, \quad \forall_{n,m\geq N}, \quad |a_n - a_m| < \varepsilon$$

This is almost the same as the definition of convergence, but instead of saying that all terms after a_N are close to L, it says that all terms after a_N are close to each other. Our aim for this section is to show that this is indeed the same thing as a sequence being convergent, which will give us a way to determine whether a sequence is convergent without knowing the limit.

$$\text{Theorem 2.12: Every Convergent Sequence is a Cauchy Sequence.}$$

pf:

Going in this direction is relatively straightforward.

$$\text{Let } a_n \text{ be a convergent sequence with } \lim a_n = L$$

The goal is to show that:

$$\forall_{\varepsilon>0}, \quad \exists_{N\in\mathbb{N}}, \text{s.t.}, \quad \forall_{n,m\geq N}, \quad |a_n - a_m| < \varepsilon$$

First as usual,

$$\text{Let } \varepsilon > 0$$

$$\text{Choose } N \cdots$$

As usual, I will come back later to make the right choice of N.

$$\text{Let } n, m \geq N$$

I now need to show that $|a_n - a_m| < \varepsilon$. We need to use the fact that a_n converges, meaning we can make $|a_n - L|$ and $|a_m - L|$ as small as we like.

Using the triangle inequality:

$$|a_n - a_m| = |a_n - L + L - a_m| \leq |a_n - L| + |a_m - L|$$

If I want to make this less than ε I can choose N such that $\forall_{n \geq N}, |a_n - L| < \frac{\varepsilon}{2}$ (which is possible since $\lim a_n = L$). This is our choice of N.

Then I have (since $n, m \geq N$, $|a_n - L| < \frac{\varepsilon}{2}$ and $|a_m - L| < \frac{\varepsilon}{2}$):

$$|a_n - a_m| \leq |a_n - L| + |a_m - L| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

Hence we are done. Every convergent sequence is Cauchy.

Going the other direction, showing that every Cauchy sequence is convergent, is a bit trickier. We will start with a warm-up exercise (also a relevant result):

$$\text{Theorem 2.13: Every Cauchy Sequence is bounded.}$$

It seems relatively clear, even from the formal definition, why Cauchy sequences would be bounded. The proof is very similar to that of convergent sequences being bounded, so give this one a go yourself before I do it here. It might help to look back at what we did for convergent sequences.

pf:

$$\text{Let } a_n \text{ be a Cauchy sequence}$$

$$\text{Let } \varepsilon = 1$$

Then apply the definition:

$$\text{there exists an } N \in \mathbb{N} \text{ such that for every } n, m \geq N, \text{we have } |a_n - a_m| < \varepsilon = 1$$

$$\text{Let } m = N \text{ (allowed since } m \geq N)$$

$$|a_n - a_N| < 1 \text{ for every } n \geq N$$

$$\Leftrightarrow a_N - 1 < a_n < a_N + 1$$

Considering the set $\{a_n : n < N\} \cup \{a_N + 1\}$

This is a finite set since $\{a_n : n < N\}$ contains less than N elements and $\{a_N + 1\}$ contains 1 element.

Therefore, this set has a maximum value, call this value M.

By definition, $M \geq a_n$ for all $n < N$. Also $M \geq a_N + 1 > a_n$ for all $n \geq N$.

This means $M \geq a_n$ for every n and so M is an upper bound, and a_n is bounded above.

Similarly, $m \coloneqq \min(\{a_n : n < N\} \cup \{a_N - 1\})$ is a lower bound for a_n and hence a_n is bounded below.

We are now ready to show the (more difficult) fact that:

$$\text{Theorem 2.14: Cauchy Sequences are convergent.}$$

pf:

$$\text{Let } a_n \text{ be a Cauchy sequence.}$$

The reason why this is difficult is because showing that a sequence is convergent requires that we already know what it is going to converge to. We need a limit candidate. For a Cauchy sequence there is no clear candidate and so we need to find one.

A good exercise to do now is to take some time to think about how we might choose a limit candidate for our Cauchy sequence (it will be useful that Cauchy sequences are bounded).

The key is that if a_n is convergent (which we don't yet know that it is) then any subsequence of a_n must also converge (and to the same limit). This means that if I can find a convergent subsequence, then it's limit is a good candidate for the limit of a_n.

$$a_n \text{ is Cauchy} \Rightarrow a_n \text{ is bounded}$$

$$\Rightarrow \text{There exists a convergent subsequence, } a_{n_j} \text{ (by Bolzano-Weierstrass).}$$

$$\text{Let } L \coloneqq \lim_{j \to \infty} a_{n_j}$$

We, now how our limit candidate! With the (most) difficult part out of the way, we proceed with a typical ε-N style proof.

$$\text{Let } \varepsilon > 0$$

$$\text{Choose } N \cdots$$

I will fill the gap later.

$$\text{Let } n \geq N$$

I want to show that $|a_n - L| < \varepsilon$. I can use the two facts, that a_n is Cauchy and that a_{n_j} converges to L. This means I can make $\left|a_{n_j} - L\right|$, and $|a_n - a_m|$ as small as I like, where m can be any number which is $\geq N$.

To write $|a_n - L|$ in terms of these things (which I know I can make small), I will use the triangle inequality:

$$|a_n - L| = \left|a_n - a_{n_j} + a_{n_j} - L\right| \leq \left|a_n - a_{n_j}\right| + \left|a_{n_j} - L\right|$$

Where $n_j \geq N$

Next, we figure out an appropriate value for N would have been:

First, using the fact that a_n is Cauchy:

$$\exists_{N_1 \in \mathbb{N}} \text{ such that } \forall_{n,m \geq N_1}, \qquad |a_n - a_m| < \frac{\varepsilon}{2}$$

Then I use the fact that $\lim a_{n_j} = L$:

$$\exists_{J \in \mathbb{N}} \text{ such that } \forall_{j \geq J}, \qquad \left|a_{n_j} - L\right| < \frac{\varepsilon}{2}$$

Then I choose $N_2 = n_J$.

Finally, I can choose $N = \max(N_1, N_2)$.

$$n, n_j \geq N_1 \Rightarrow \left|a_n - a_{n_j}\right| < \frac{\varepsilon}{2}$$

$$n_j \geq N_2 = n_J \Rightarrow j \geq J \Rightarrow \left|a_{n_j} - L\right| < \frac{\varepsilon}{2}$$

Putting these together:

$$|a_n - L| \leq \left|a_n - a_{n_j}\right| + \left|a_{n_j} - L\right| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

And we are done. Cauchy Sequences are convergent.

Limit Superior and Limit Inferior

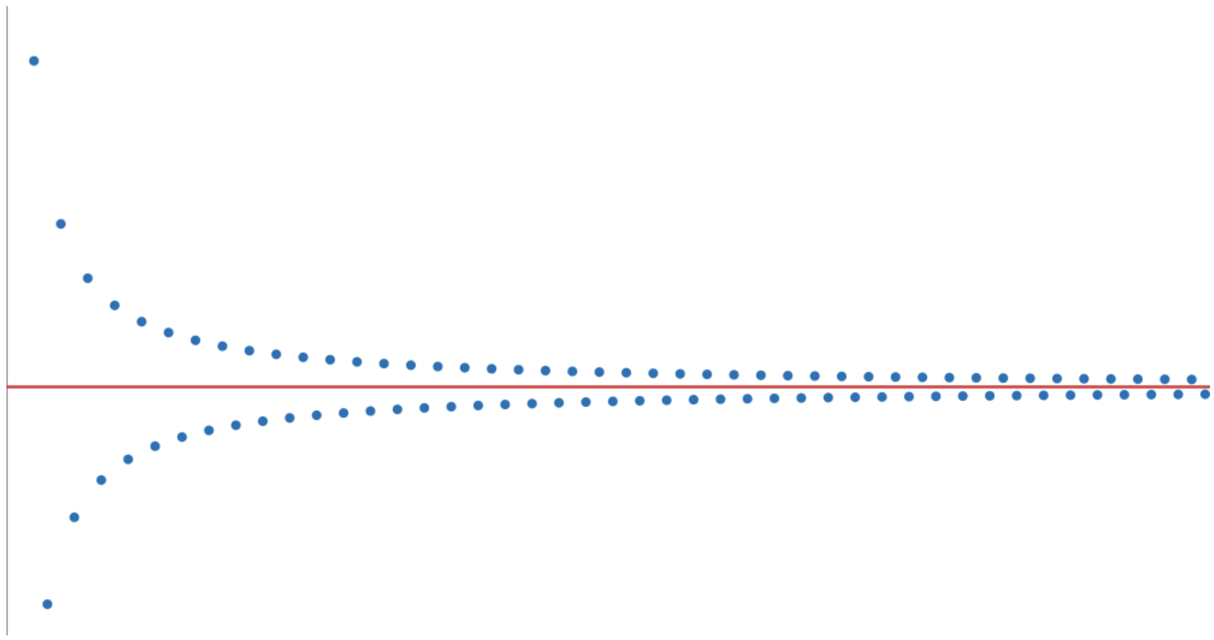
We have seen that every convergent sequence is bounded, but not every bounded sequence is convergent.

Say I have a bounded sequence which does not converge. I may not be able to say what the limit of the sequence is (since there is no limit), but what can I say about it?

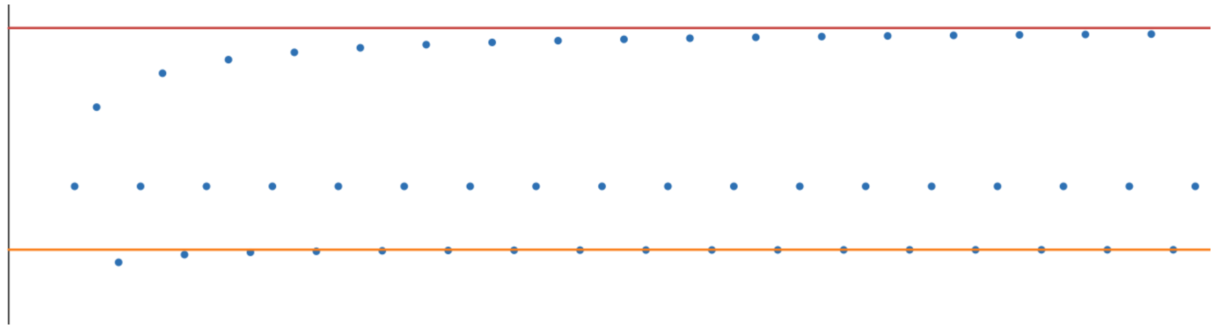
This is where the concept of limit superior/inferior comes in.

One way to think of a limit is that it is the number which the sequence "closes in" onto as n gets larger. If the sequence doesn't converge, but is bounded, the sequence doesn't close in on any particular value, but the lower and upper bound of the terms might do. A good way to illustrate what I mean by this is with a diagram.

This is a sequence which is converging to a limit (the red line)



This is a bounded sequence which is not convergent:



We call the red line the "**limit superior**" of the sequence and the orange line the "**limit inferior**".

As a more precise definition, first we define the sequence of upper and lower bounds of our sequence.

The sequence of least upper bounds:

$$\overline{x}_n \coloneqq \sup\{x_m : m \geq n\}$$

[This exists so long as x_n is bounded above]

$\overline{x}_n$ tells you the least upper bound of the terms of the sequence after and including the n^{th} term. Since this depends on n, it is itself a sequence.

Similarly:

$$\underline{x}_n \coloneqq \inf\{x_m : m \geq n\}$$

This is the greatest lower bound of all the terms of the sequence after and including the nth term.

Finally, we define **limit superior**:

$$\limsup x_n = \lim \overline{x}_n$$

Assuming this limit exists. Remember, $\overline{x}_n$ is itself a sequence so this makes sense.

And **limit inferior**:

$$\liminf x_n = \lim \underline{x}_n$$

There are 2 things we will do in this section:

- Show that every bounded sequence has a limit superior and limit inferior.
- Demonstrate how you can find them.

Theorem 2.15: a_n is bounded above $\Rightarrow \limsup a_n$ exists

pf:

First check that we can define $\overline{x}_n \coloneqq \sup\{a_m : m \geq n\}$.

First, I will let

$$S_n = \{a_m : m \geq n\}$$

$a_n \in S_n$ and so S_n is non-empty, also a_n is bounded above $\Rightarrow S_n$ is bounded above.

Hence by the axiom of completeness, $\sup S_n = \overline{x}_n$ exists.

Next, I need to show that the sequence $\overline{x}_n$ converges.

Intuitively, it seems clear that $\overline{x}_n$ will be bounded and also decreasing, so we will prove these two facts and use monotone convergence.

First to show that it is bounded:

I know that a_n is bounded above and below, hence there exist some L, U such that

$l \leq a_n \leq u$ for every n.

$$\Rightarrow \text{for every } a_m \in S_n, \qquad l \leq a_m \leq u$$

$$\Rightarrow \overline{x}_n = \sup S_n \in [l, u]$$

Hence $\overline{x}_n$ is bounded.

To show that it is decreasing, I need to show that $i \leq j \Rightarrow \overline{x}_i \geq \overline{x}_j$

$$\text{Let } i \leq j$$

$$\text{Let } a_m \in S_j, \text{then } m \geq j \geq i \text{ so } a_m \in S_i$$

$$\text{Hence } S_j \subseteq S_i \ \left(S_j \text{ is a subset of } S_i\right)$$

$$\overline{x}_i \text{ is an upper bound for } S_i \text{ so is also an upper bound for } S_j$$

$$\text{so } \overline{x}_j \ \left(\text{which is} \leq \text{all upper bounds of } S_j\right) \leq \overline{x}_i$$

$$\overline{x}_i \geq \overline{x}_j$$

Hence $\overline{x}_n$ is bounded and decreasing, so by monotone convergence:

$$\limsup x_n \ \text{ exists.}$$

The proof given here uses a lot of sets and can be difficult to follow. If you are struggling to follow it, take a step back and think of the bigger picture. Why do we *expect* the sequence to be bounded? Why do we *expect* it to be decreasing? This is a type of thinking which we need all the time in Analysis. The ability to takes something that seems obvious, and put into a more rigorous form,.

The proof for lim inf is practically identical, I leave this to you if you wish.

Some very important (slightly obvious) facts are the following:

- For any convergent subsequence a_{n_j}, $\liminf a_n \leq \lim a_{n_j} \leq \limsup a_n$ (2.16)

- There always exists a subsequence a_{n_j} with $\lim a_{n_j} = \limsup a_n$ (2.17a)
- There always exists a subsequence a_{n_j} with $\lim a_{n_j} = \liminf a_n$ (2.17b)
- a_n is convergent iff $\liminf a_n = \limsup a_n$, also $\lim a_n = \limsup a_n = \liminf a_n$ (2.18)

Theorem 2.16:

pf:

Let a_n be a bounded sequence, Let a_{n_j} be a convergent subsequence of a_n

By definition:

$$\limsup a_n = \lim \overline{a}_n, \qquad \liminf a_n = \lim \underline{a}_n$$

$$\overline{a}_{n_j} = \sup\{a_m : m \geq n_j\}$$

Notice that $a_{n_j} \in \{a_m : m \geq n_j\}$ so $a_{n_j} \leq \overline{a}_{n_j}$.

Similarly, $\underline{a}_{n_j} \leq a_{n_j}$, so we have $\underline{a}_{n_j} \leq a_{n_j} \leq \overline{a}_{n_j}$

We know (since a subsequence converges to the same limit as the original sequence):

$$\lim \overline{a}_{n_j} = \lim \overline{a}_n = \limsup a_n, \qquad \lim \underline{a}_{n_j} = \lim \underline{a}_n = \liminf a_n$$

Hence (by squeezing theorem):

$$\lim \underline{a}_{n_j} \leq \lim a_{n_j} \leq \lim \overline{a}_{n_j}$$

$$\liminf a_n \leq \lim a_{n_j} \leq \limsup a_n$$

As required

Parts a and b of 2.17 are essentially the same, so I will only do a.

We know that $\limsup a_n$ is defined as the limit of a sequence $(\overline{a}_n)$ so you may be wondering why this result is not trivial. The reason is because $\overline{a}_n$ is not (necessarily) a subsequence of a_n.

The idea of the proof will be that, since each $\overline{a}_n$ is the least upper bound of a set of terms a_m, there must be some terms which get (arbitrarily) close to $\overline{a}_n$. We will find a term within $\frac{1}{k}$ of $\overline{a}_n$ and this term will be our a_{n_k} (the kth term of our subsequence a_{n_j}). The subsequence (hopefully) then converges to $\overline{a}_n$ which itself is a sequence which converges to $\limsup a_n$. Taken together, we are hoping for a_{n_j} to converge to $\limsup a_n$.

$$\text{Theorem 2.17(a):}$$

pf:

$$\text{Let } a_n \text{ be a bounded sequence. Let } L = \lim\sup a_n$$

The goal is to find a subsequence a_{n_j} whose limit is L.

$$\overline{a}_1 = \sup\{a_m : m \geq 1\}$$

$\overline{a}_1$ is an upper bound for this set and $\overline{a}_1 - 1$ is not an upper bound for this set (as $\overline{a}_n$ is a supremum).

$$\overline{a}_1 \text{ is an upper bound, } \overline{a}_1 - 1 \text{ is not} \Rightarrow \exists_{n_1 \geq 1} \text{ such that } \overline{a}_1 > a_{n_1} > \overline{a}_1 - 1$$

$$\text{Consider } \overline{a}_{n_1} = \sup\{a_m : m \geq n_1\}$$

$$\overline{a}_{n_1} \text{ is an upper bound, } \overline{a}_{n_1} - \frac{1}{2} \text{ is not,} \qquad \text{so } \exists_{n_2 \geq n_1} \text{ such that } \overline{a}_{n_1} > a_{n_2} > \overline{a}_{n_1} - \frac{1}{2}$$

Continuing this pattern,

$$\text{Consider } \overline{a}_{n_k} = \sup\{a_m : m \geq n_k\}$$

$$\overline{a}_{n_k} \text{ is an upper bound, } \overline{a}_{n_k} - \frac{1}{k} \text{ is not,} \qquad \text{so } \exists_{n_{k+1} \geq n_k} \text{ s.t.,} \qquad \overline{a}_{n_k} > a_{n_{k+1}} > \overline{a}_{n_k} - \frac{1}{k}$$

We form this subsequence a_{n_j}, which is such that $\overline{a}_{n_j} \geq a_{n_{j+1}} \geq \overline{a}_{n_j} - \frac{1}{j}$ for every j.

$$\lim \overline{a}_{n_j} = \lim \overline{a_n} = \lim\sup a_n = L$$

$$\lim\left(\overline{a}_{n_j} - \frac{1}{j}\right) = \lim \overline{a}_{n_j} - 0 = \lim \overline{a}_n = \lim\sup a_n = L$$

$$\text{Hence (by squeezing): } \lim a_{n_{j+1}} = \lim a_{n_j} = L$$

And we are done. This was probably one of the hardest proof so far so don't worry if you need to read it a few more times to understand why each step works, and more importantly what the big idea of this proof was. If you want to test your understanding, try part b without looking back at this proof. (Find a subsequence with limit $\lim\inf a_n$).

Finally, Theorem 2.18.

$$\text{Theorem 2.18: } a_n \text{ is convergent iff } \lim\inf a_n = \lim\sup a_n$$
$$\text{Also we then have } \lim a_n = \lim\sup a_n = \lim\inf a_n$$

Recall that iff means (if and only if) so we need to prove this in both directions.

pf:

The "⇒" direction is easy.

$$\text{Let } a_n \text{ be a convergent sequence, with } \lim a_n = L$$

$$\text{convergent} \Rightarrow \text{bounded}$$

$$\text{By Theorem 2.17 there must be a subsequence } a_{n_j} \text{ with } \lim a_{n_j} = \limsup a_n$$

$$\text{Since } a_n \text{ is convergent to } L, \text{ every subsequence is convergent to } L, \text{ meaning:}$$

$$L = \lim a_{n_j} = \limsup a_n$$

$$\text{Similarly, there is a subsubsequence with limit } \liminf a_n \text{ hence } L = \liminf a_n$$

$$\text{so we have } \liminf a_n = \limsup a_n = L$$

$$\text{as required.}$$

The "⇐" direction is much harder.

$$\text{Let } L \coloneqq \liminf a_n = \limsup a_n$$

We need to first show that a_n is convergent (specifically to L). We have a limit candidate so maybe we can use the ε-N definition.

$$\text{Let } \varepsilon > 0$$

$$\text{Choose } N \cdots$$

I will come back later once we have a better idea of how to choose N.

$$\text{Let } n \geq N$$

We want $|a_n - L| < \varepsilon$.

The key here is of course that lim sup = lim inf.

If I let $S_n = \{x_m : m \geq n\}$, then $\liminf a_n = \lim(\inf S_n) = L$ and

$\limsup a_n = \lim(\sup S_n) = L$.

Then we can make $|\inf S_n - L|$ and $|\sup S_n - L|$ as small as we like.

Now we choose N such that $\forall_{n \geq N}$, $|\inf S_n - L| < \varepsilon$ and $|\sup S_n - L| < \varepsilon$.

This can be written as:

$$L - \varepsilon < \inf S_n < L + \varepsilon \text{ and } L - \varepsilon < \sup S_n < L + \varepsilon$$

Also, since $a_n \in S_n$, we know that $\inf S_n \leq a_n \leq \sup S_n$. We can then use the above inequalities:

$$L - \varepsilon < \inf S_n \leq a_n \leq \sup S_n < L + \varepsilon$$

$$\Rightarrow L - \varepsilon < a_n < L + \varepsilon$$

$$\Rightarrow |a_n - L| < \varepsilon$$

As required, so we are done.

We can use these theorems to find the lim sup and lim inf of a sequence by considering its subsequences:

$$a_n = \begin{cases} 1 & n = 3k \text{ for some } k \in \mathbb{N} \\ 2 - \dfrac{1}{n^2} & n = 3k - 2 \text{ for some } k \in \mathbb{N} \\ \dfrac{(-1)^n}{n} & n = 3k - 1 \text{ for some } k \in \mathbb{N} \end{cases}$$

Find $\liminf a_n$ and $\limsup a_n$.

We need to split this sequence into subsequences.

When $n = 3k$ (is a multiple of 3), we have 1.

When $n = 3k - 2$ (is 2 less than a multiple of 3) we have $2 - \frac{1}{n^2} = 2 - \frac{1}{(3k-2)^2}$

When $n = 3k - 1$ (is 1 less than a multiple of 3) we have $\frac{(-1)^n}{n} = \frac{(-1)^{3k-1}}{3k-1}$

If we find the limit of all 3 of these subsequences, we can find the limits superior and inferior.

$$\text{First sequence: } \lim 1 = 1 \text{ (constant sequence)}$$

$$\text{Second sequence: } \lim\left(2 - \frac{1}{(3k-2)^2}\right) = 2 \text{ (since } (3k-2)^2 \to \infty)$$

$$\text{Third sequence: } \left|\frac{(-1)^{3k-1}}{3k-1}\right| \leq \frac{1}{3k-1} \text{ and } \lim \frac{1}{3k-1} = 0$$

$$\text{so } \lim \frac{(-1)^{3k-1}}{3k-1} = 0 \text{ (by squeezing theorem)}$$

The limits are 1,2,0. Min is 0, max is 2, so

$$\limsup a_n = 2, \qquad \liminf a_n = 0$$

Infinite Series

Definition and Basic Examples

A sequence is an infinite list of terms. An infinite series is (roughly speaking) what you get when you add those terms together.

The notion of adding infinitely many things together is an intuitive one, but we must still define it in a rigorous way.

Sigma notation for sums was introduced earlier, but as a reminder,

$$\text{for } a, b \in \mathbb{Z} \text{ with } a \leq b:$$

$$\sum_{k=a}^{a} f(k) = f(a), \qquad \sum_{k=a}^{b+1} f(k) = \sum_{k=a}^{b} f(k) + f(b+1)$$

Or less formally:

$$\sum_{k=a}^{b} f(k) = f(a) + f(a+1) + \cdots + f(b-1) + f(b)$$

This chapter is all about infinite series. We define an infinite series by using the ideas built up in the previous chapter on sequences.

We define the sequence,

$$s_n := \sum_{k=1}^{n} a_k$$

We call this the nth partial sum. We then define the infinite series by:

$$\sum_{k=1}^{\infty} a_k := \lim s_n$$

assuming s_n converges. If it does we say the series converges, otherwise we say the series diverges. An infinite series is the limit of the sequence of partial sums. In some situations we will be lucky enough that we can write s_n explicitly and then determine whether the series converges, and perhaps even evaluate it. Unfortunately, most of the time evaluating these infinite series is difficult, and so our primary focus will be simply on determining whether they converge. There are a few special cases though in which we can determine the exact limit.

A first, familiar example is that of the geometric series. This is any series of the form:

$$\sum_{k=1}^{\infty} ar^{k-1}$$

We call this a geometric series with first term a and a common ratio of r.

If we can evaluate the partial sums exactly and consider the limit, then we might be able to say something about its convergence.

$$\text{Let } s_n = a + ar^2 + \cdots + ar^{n-1}$$

The key idea is that multiplying by r gives a lot of the same terms as we already have.

$$rs_n = ar + ar^2 + \cdots + ar^n$$

Almost all the terms are the same and so if we subtract the second from the first:

$$s_n - rs_n = a - ar^n$$

since everything else cancels out. Finally, we solve for s_n

$$s_n(1-r) = a(1-r^n)$$

$$s_n = \frac{a(1-r^n)}{1-r}$$

If you wanted to be more formal than I have, you could prove this formula works by induction.

r is the common ratio, a is the first term and n is the number of terms. Also, clearly this formula doesn't work for $r = 1$ (division by zero), we will need to deal with that separately.

We can then use this to find:

$$\sum_{k=1}^{\infty} ar^{k-1} \coloneqq \lim s_n$$

$$\lim s_n = \frac{a}{1-r}(1 - \lim r^n)$$

This is a geometric sequence, which we have met before. It's limit is 0 for $|r| < 1$ and it diverges for $|r| > 1$ and for $r = -1$.

This means, for $|r| > 1$ and $r = -1$, our infinite series diverges and for $|r| < 1$ our infinite series converges to $\frac{a}{1-r}$.

For $r = 1$, we have

$$s_n = \sum_{k=1}^{n} a = an$$

Then $s_n = an$ is unbounded (if $a \neq 0$) and hence diverges.

To conclude,

$$\sum_{k=1}^{\infty} ar^{k-1}, \qquad a \neq 0$$

Will converge to $\frac{a}{1-r}$ for $|r| < 1$ and will diverge otherwise.

If $a = 0$ then we have $\sum_{k=1}^{\infty} 0 = 0$.

There is another scenario in which we can also determine the exact value of an infinite series. We will use the **method of differences** which only works in a select number of cases. The method of difference is a fancy way of saying "most of the terms cancel out". Consider:

$$\sum_{k=1}^{n} \big(f(k) - f(k-1)\big)$$

Writing the terms out explicitly:

$$f(1) - f(0) + f(2) - f(1) + \cdots + f(n) - f(n-1)$$

All terms except the second and second to last cancel, so we end up with $f(n) - f(0)$. The method of differences can show up in slightly different forms as well,

$$\sum_{k=1}^{n} \big(f(k) - f(k+1)\big) = f(1) - f(2) + f(2) - f(3) + \cdots + f(n) - f(n+1)$$

$$= f(1) - f(n+1)$$

So be careful to make sure that you are cancelling the right things. If you want to be really formal, you can prove the results of the method of differences by using induction. Once we have an explicit form for s_n, we might be able to say whether it converges, and what it converges to.

$$\text{E.g.,} \qquad \text{Evaluate } \sum_{k=1}^{\infty} \left(\frac{1}{k} - \frac{1}{k+1}\right)$$

$$\text{Let } s_n \coloneqq \sum_{k=1}^{n} \left(\frac{1}{k} - \frac{1}{k+1}\right)$$

Using the method of differences,

$$s_n = 1 - \frac{1}{2} + \frac{1}{2} - \frac{1}{3} + \frac{1}{3} - \frac{1}{4} + \cdots + \frac{1}{n} - \frac{1}{n+1}$$

$$= 1 - \frac{1}{n+1}$$

$$\text{Then } \sum_{k=1}^{\infty}\left(\frac{1}{k} - \frac{1}{k+1}\right) \coloneqq \lim\left(1 - \frac{1}{n+1}\right) = 1$$

Before finishing this section, I will prove one important, hopefully obvious, fact about infinite series.

$$\text{Theorem 3.1: if } \sum_{k=1}^{\infty} a_k \text{ converges, then } \lim a_n = 0$$

This seems obvious, because for the sum to converge, it seems that we would need the terms to get smaller and smaller.

The proof is surprisingly simple.

pf:

$$\text{Let } s_n = \sum_{k=1}^{n} a_k$$

$$a_n = s_n - s_{n-1} \text{ (for } n \geq 2)$$

$$\lim(s_n - s_{n-1}) = \lim s_n - \lim s_{n-1} = L - L = 0$$

$$\text{Hence } \lim a_n = 0$$

Tests for Convergence

Unfortunately, we will not be so lucky as to be able to directly evaluate most infinite series. We will however develop a number of ways to determine whether certain infinite series converge. We will cover:

- Comparison Test
- Cauchy Condensation Test
- p-series Test
- Alternating Series Test
- Ratio Test
- Root Test

As you can see we have a lot to get through, so lets get started.

Comparison Test

$$\text{Theorem 3.2: Comparison Test:}$$

$$\text{Let } 0 \leq a_k \leq b_k \text{ for every } k$$

$$\text{If } \sum_{k=1}^{\infty} b_k \text{ converges then } \sum_{k=1}^{\infty} a_k \text{ converges (a)}$$

$$\text{If } \sum_{k=1}^{\infty} a_k \text{ diverges then } \sum_{k=1}^{\infty} b_k \text{ diverges (b)}$$

The statement seems rather obvious, but this is probably the most useful of the convergence tests and so we ought to prove it.

pf:

Part (a):

$$\text{Let } 0 \leq a_k \leq b_k \text{ for every } k$$

$$\text{suppose } \sum_{k=1}^{\infty} b_k \text{ converges}$$

$$\text{This just means that } s_n := \sum_{k=1}^{\infty} b_k \text{ is a convergent sequence}$$

$$\Rightarrow s_n \text{ is bounded, let } M \geq s_n \text{ for every } n$$

$$\text{Let } t_n = \sum_{k=1}^{n} a_k$$

$$\text{Then } t_n \leq s_n \text{ for every } n \text{ (exercise, prove this by induction!)}$$

$$\text{then } t_n \leq s_n \leq M \text{ for every } n \text{ so } t_n \text{ is bounded above}$$

$$t_{n+1} = t_n + a_{n+1} \geq t_n \text{ (since } a_{n+1} \geq 0)$$

$$\text{so } t_n \text{ is monotone increasing.}$$

$$\text{By monotone convergence, } t_n \text{ converges.}$$

$$\text{i.e., } \sum_{k=1}^{\infty} a_k \text{ converges.}$$

Notice that part (b) is the contrapositive of part (a). Recall from the first chapter:

$$\text{Contrapositive: } A \Rightarrow B \text{ is the same as (not } B) \Rightarrow \text{(not } A)$$

Meaning part (b) is equivalent to part (a), and we are done.

For example:

$$\text{Does } \sum_{k=1}^{\infty} \frac{1}{k^2} \text{ converge?}$$

Choosing the right thing to compare this to is tricky. First, we suspect that it will converge (the terms go to zero very quickly). We want to find a sequence b_k such that $0 \leq \frac{1}{k^2} \leq b_k$ for every k, and such that $\sum_{k=1}^{\infty} b_k$ converges.

The choice we will use is:

$$\frac{1}{k^2} \leq \frac{1}{k(k-1)}$$

The way I think about problems like this, is $\frac{1}{k^2} \leq$ (something). I want that (something) to be $\geq \frac{1}{k^2}$. I can make $\frac{1}{k^2}$ bigger by increasing the numerator (top), or by decreasing the denominator (bottom). Here I have decreased the denominator by replacing k by $k-1$.

Now all that remains to be shown is that

$$\sum_{k=2}^{\infty} \frac{1}{k(k-1)} \text{ converges}$$

I am starting at 2 because the $k = 1$ term would be dividing by zero. If this sum converges, then (by the comparison test) $\sum_{k=2}^{\infty} \frac{1}{k^2}$ converges and then the original series converges.

The trick is to use partial fractions.

$$\text{Let } \frac{1}{k(k-1)} = \frac{A}{k} + \frac{B}{k-1}$$

(I haven't justified that we can write it in this form, I will justify this in a moment).

Multiply both sides by $k(k-1)$

$$1 = A(k-1) + Bk$$

Let $k = 1$, $1 = B \Rightarrow B = 1$

Let $k = 0$, $1 = -A \Rightarrow A = -1$

This gives $\frac{1}{k(k-1)} = \frac{1}{k-1} - \frac{1}{k}$

I have not fully justified the steps here, but you can verify that the final result is true, by getting a common denominator:

$$\frac{1}{k-1} - \frac{1}{k} = \frac{k}{k(k-1)} - \frac{k-1}{k(k-1)} = \frac{1}{k(k-1)}$$

Which justifies the result.

Finally, we can evaluate

$$\sum_{k=2}^{\infty} \frac{1}{k(k-1)} = \sum_{k=2}^{\infty} \left(\frac{1}{k-1} - \frac{1}{k}\right)$$

by using the method of differences. It comes out to 1. Importantly it converges. This means that (by the comparison test):

$$\sum_{k=2}^{\infty} \frac{1}{k^2} \text{ converges}$$

and in fact, is ≤ 1. Finally,

$$\sum_{k=1}^{\infty} \frac{1}{k^2} \text{ converges and is } \leq 2$$

Finding the exact value of this sum is a difficult problem known as the Basel problem, which was solved by Leonhard Euler in 1734. The answer turns out to be $\frac{\pi^2}{6}$. I will not be proving this result in this course, but I might do so in future courses.

Cauchy-Condensation Test

$$\text{We have seen that if } \sum_{k=1}^{\infty} a_k \text{ converges, then } a_n \to 0$$

$$\text{Is the converse True?}$$

The answer is no, and one counter-example is perhaps the most simple example you could come up with.

Let $a_n = \frac{1}{n}$. We know that $a_n \to 0$. It is not too difficult to show that $\sum \frac{1}{k}$ diverges. We call this series, the harmonic series.

$$\text{Theorem 3.3 (Harmonic Series):}$$

$$\sum_{k=1}^{\infty} \frac{1}{k} \text{ diverges.}$$

The idea is as follows:

$$\frac{1}{2} \geq \frac{1}{2}, \qquad \frac{1}{3} + \frac{1}{4} \geq \frac{1}{4} + \frac{1}{4} = \frac{1}{2}, \qquad \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} \geq \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{1}{2}, \cdots$$

And hence

$$1 + \left(\frac{1}{2}\right) + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) + \cdots \geq 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \cdots$$

$$= 1 + \sum_{k=1}^{\infty} \frac{1}{2} \text{ which diverges}$$

$$\text{So by comparison: } \sum_{k=1}^{\infty} \frac{1}{k} \text{ diverges.}$$

This is certainly convincing, but there are certain hand-wavy elements, so I will now make it more rigorous by using an induction argument:

$$\text{Let } s_n = \sum_{k=1}^{n} \frac{1}{k}$$

$$\text{Consider } s_{2^n} = \sum_{k=1}^{2^n} \frac{1}{k}$$

We want to show that $s_{2^n} \geq 1 + \frac{n}{2}$ and so is unbounded. We can do this by induction.

$$\text{Basis: } n = 0:$$

$$\text{LHS} = s_{2^0} = s_1 = 1$$

$$\text{RHS} = 1 + \frac{0}{2} = 1 = \text{LHS}$$

$$\text{Assume true for some } m$$

$$s_{2^m} \geq 1 + \frac{m}{2}$$

$$\text{Induction: show that it must be true for } m + 1$$

$$s_{2^{m+1}} = \sum_{k=1}^{2^{m+1}} \frac{1}{k} = \sum_{k=1}^{2^m} \frac{1}{k} + \sum_{k=2^m+1}^{2^{m+1}} \frac{1}{k} = s_{2^m} + \sum_{k=2^m+1}^{2^{m+1}} \frac{1}{k}$$

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

I now need only to put a lower bound on this second sum. Notice that for every k between $2^m + 1$ and 2^{m+1} we have $0 < k \leq 2^{m+1}$ which means $\frac{1}{k} \geq \frac{1}{2^{m+1}}$. This means:

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

Which is a sequence of constant terms (constant with respect to k). There are 2^m terms ($2^{m+1} - 2^m = 2 \cdot 2^m - 2^m = 2^m$), so this sum is

$$\sum_{k=2^m+1}^{2^{m+1}} \frac{1}{2^{m+1}} = \frac{2^m}{2^{m+1}} = \frac{1}{2}$$

Which means, $s_{2^{m+1}} \geq 1 + \frac{m}{2} + \frac{1}{2} = 1 + \frac{m+1}{2}$ which is the statement for $n = m + 1$.

We have proved by induction that $s_{2^n} \geq 1 + \frac{n}{2}$ which is unbounded above, meaning s_{2^n} is also unbouned above and hence diverges (convergent sequences are bounded).

If s_n converged, then every subsequence (including s_{2^n}) would converge. This is not so, and hence s_n diverges.

We are done.

We can generalise the method used here to create the **Cauchy Condensation Test**, which says that, if $a_n \geq 0$ is a monotonic decreasing sequence, then

Theorem 3.4: **Cauchy Condensation Test**

$$\sum_{k=1}^{\infty} a_k \text{ converges} \Leftrightarrow \sum_{k=0}^{\infty} 2^k a_{2^k} \text{ converges.}$$

Applying this to $a_n = \frac{1}{n}$: $\left(\frac{1}{n} \geq 0 \text{ and is monotonic decreasing, so we can use it}\right)$:

$$\sum_{k=0}^{\infty} 2^k a_{2^k} = \sum_{k=0}^{\infty} \frac{2^k}{2^k} = \sum_{k=0}^{\infty} 1 \text{ which diverges, hence the harmonic series diverges.}$$

Sketch of proof:

" $\Rightarrow$ "

$$\text{decreasing means } a_1 \geq a_2 \geq a_3 \geq a_4 \geq \cdots$$

$$\sum_{k=1}^{\infty} a_k = a_1 + a_2 + a_3 + a_4 \cdots \geq a_1 + a_2 + a_4 + a_4 + a_8 + a_8 + a_8 + a_8 + \cdots$$

$$= a_1 + a_2 + 2a_4 + 4a_8 + \cdots = a_1 + \frac{1}{2}(2a_2 + 4a_4 + 8a_8 + \cdots)$$

$$= a_1 + \frac{1}{2}\sum_{k=1}^{\infty} 2^k a_{2^k}$$

So, if $\sum_{k=1}^{\infty} a_k$ converges then $\sum_{k=0}^{\infty} 2^k a_{2^k}$ converges (by comparison test)

" $\Leftarrow$ "

$$\sum_{k=1}^{\infty} a_k = a_1 + a_2 + a_3 + a_4 \cdots \leq a_1 + a_2 + a_2 + a_4 + a_4 + a_4 + a_4 + \cdots$$

$$= a_1 + 2a_2 + 4a_4 + \cdots = \sum_{k=0}^{\infty} 2^k a_{2^k}$$

So, if $\sum_{k=0}^{\infty} 2^k a_{2^k}$ converges then $\sum_{k=1}^{\infty} a_k$ converges (again, by comparison).

p-Series Test

$$\text{We have seen that } \sum_{k=1}^{\infty} \frac{1}{k} \text{ diverges and that } \sum_{k=1}^{\infty} \frac{1}{k^2} \text{ converges}$$

But what about:

$$\sum_{k=1}^{\infty} \frac{1}{k^p}, \qquad \text{for } p \in \mathbb{R}?$$

We haven't yet defined what k^p actually means for real numbers p. The definition of these will rely on the exponential function (which we have not yet defined). I could do this now but so as not to completely break the flow, we will assume certain key properties as given, we will properly define exponentiation later.

We will assume for now that k^x with $k \in \mathbb{N}$ is an increasing function, meaning that $x < y \Rightarrow k^x \le k^y$. Also x^p in an increasing function for $p \ge 0$ and $x > 0$, that is $0 \le x < y \Rightarrow x^p \le y^p$.

We will properly define exponentiation later on.

We can use the comparison test to immediately conclude that for $p \le 1$, $k^p \le k^1 = k$

$$\Rightarrow \frac{1}{k^p} \ge \frac{1}{k}$$

$$\text{So by comparison, since } \sum_{k=1}^{\infty} \frac{1}{k} \text{ diverges, } \sum_{k=1}^{\infty} \frac{1}{k^p} \text{ diverges for } p \le 1$$

$$\text{Similarly, } \sum_{k=1}^{\infty} \frac{1}{k^p} \text{ converges for } p \ge 2$$

The key question then, is what happens when $1 < p < 2$? The typical way of doing this is by using the "integral test", but we will not be defining integration for quite a while, so we will do it without.

Theroem 3.5: ***p*-Series Test**

$$\sum_{k=1}^{\infty} \frac{1}{k^p} \text{ converges iff } p > 1$$

pf:

We have already proved all cases except $1 < p < 2$.

The trick is to split the sum into its even and odd terms. Each term is ≥ 0 so we will attempt to find an upper bound and use monotone convergence.

$$\text{Let } 1 < p < 2:$$

$$\text{Let } s_n = \sum_{k=1}^{n} \frac{1}{k^p}$$

$$s_{2n} = \sum_{k=1}^{2n} \frac{1}{k^p}$$

We consider the subsequence s_{2n} because this ensure that there is the same number of odd and even terms.

$$s_{2n} = \sum_{k=1}^{n} \left(\frac{1}{(2k-1)^p} + \frac{1}{(2k)^p} \right)$$

(This is the odd an even part, if you really wanted to you could use induction to show that this is the same as s_{2n}).

$$s_{2n} = \sum_{k=1}^{n} \frac{1}{(2k-1)^p} + \sum_{k=1}^{n} \frac{1}{(2k)^p}$$

$$= \sum_{k=1}^{n} \frac{1}{(2k-1)^p} + \frac{1}{2^p} \sum_{k=1}^{n} \frac{1}{k^p}$$

$$= \sum_{k=1}^{n} \frac{1}{(2k-1)^p} + \frac{1}{2^p} s_n$$

We factored out a $\frac{1}{2^p}$ in order to get s_{2n} in terms of s_n. We can take this one step further by using comparison, since $\frac{1}{(2k-1)^p} \leq \frac{1}{(2k-2)^p}$ (for $k > 1$).

$$s_{2n} = \sum_{k=1}^{n} \frac{1}{(2k-1)^p} + \frac{1}{2^p} s_n = 1 + \sum_{k=2}^{n} \frac{1}{(2k-1)^p} + \frac{1}{2^p} s_n$$

$$\leq 1 + \sum_{k=2}^{n} \frac{1}{(2k-2)^p} + \frac{1}{2^p} s_n = 1 + \frac{1}{2^p} \sum_{k=2}^{n} \frac{1}{(k-1)^p} + \frac{1}{2^p} s_n$$

$$= 1 + \frac{1}{2^p} \sum_{k=1}^{n-1} \frac{1}{k^p} + \frac{1}{2^p} s_n = 1 + \frac{1}{2^p} s_{n-1} + \frac{1}{2^p} s_n = 1 + \frac{1}{2^p} (s_{n-1} + s_n)$$

$$= 1 + \frac{1}{2^p} \left(s_n - \frac{1}{n^p} + s_n \right) = 1 + \frac{1}{2^p} \left(2s_n - \frac{1}{n^p} \right)$$

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

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

What have we achieved by doing this? Remember we want an upper bound for s_n.

We also know that s_n is an increasing sequence (since each a_k is positive), hence

$$n \leq 2n \Longrightarrow s_n \leq s_{2n} \leq 1 - \frac{1}{(2n)^p} + \frac{s_n}{2^{p-1}}$$

$$s_n \leq 1 - \frac{1}{(2n)^p} + \frac{s_n}{2^{p-1}}$$

$$\Rightarrow s_n\left(1 - \frac{1}{2^{p-1}}\right) \leq 1 - \frac{1}{(2n)^p}$$

$$\Rightarrow s_n(2^p - 2) \leq 2^p - \frac{1}{n^p}$$

$$\Rightarrow s_n(2^p - 2) \leq \frac{(2n)^p - 1}{n^p} \leq \frac{(2n)^p}{n^p} = 2^p$$

$$\Rightarrow s_n \leq \frac{2^p}{2^p - 2} \text{ } (p > 1 \Rightarrow 2^p - 2 > 0, \text{so we can divide by it})$$

Hence, s_n is bounded above by $\frac{2^p}{2^p-2}$, also s_n is increasing, hence by monotone convergence, s_n converges.

Alternating Series Test

The next test we will consider has a rather specific use case.

Theorem 3.6: **Alternating Series Test**

$$\text{If } a_n \geq 0 \text{ is monotone decreasing with } \lim a_n = 0, \text{then } \sum_{k=1}^{\infty}(-1)^k a_k \text{ is convergent.}$$

This seems like a reasonable statement to make, the only thing which might seem strange is that we require a_n to be monotone decreasing, it will become apparent during the proof why we need this condition.

pf:

The idea is that the odd partial sums take us below the limit, the even ones takes us above, but we are always getting closer to that limit (due to monotone decreasing).

$$\text{Let } s_n = \sum_{k=1}^{n}(-1)^k a_k$$

I will consider separately the odd and even subsequences. I seems that s_{2n} (the partial sums above the limit) should be decreasing and bounded below. Let us prove this:

$$s_{2n} = \sum_{k=1}^{2n}(-1)^k a_k = \sum_{k=1}^{n}((-1)^{2k-1}a_{2k-1} + (-1)^{2k}a_{2k})$$

$$= \sum_{k=1}^{n} (a_{2k} - a_{2k-1})$$

Where we have split the sum into odd and even parts. Since a_k is decreasing, $a_{2k} \leq a_{2k-1} \Rightarrow a_{2k} - a_{2k-1} \leq 0$, so the subsequence s_{2n} is decreasing. We can similarly show that s_{2n-1} is increasing:

$$s_{2n-1} = \sum_{k=1}^{2n-1} (-1)^k a_k = -a_1 + \sum_{k=2}^{2n-1} (-1)^k a_k$$

$$= -a_1 - \sum_{k=1}^{2n-2} (-1)^k a_{k+1}$$

$$= -a_1 - \sum_{k=1}^{n-1} ((-1)^{2k-1} a_{2k-1+1} + (-1)^{2k} a_{2k+1})$$

$$= -a_1 + \sum_{k=1}^{n-1} (a_{2k} - a_{2k+1})$$

$$a_{2k+1} \leq a_{2k} \Rightarrow a_{2k} - a_{2k+1} \geq 0 \text{ so } s_{2n-1} \text{ is increasing}$$

Next, we want to show that s_{2n} is bounded below (we suspect by s_1).

$$s_{2n} = s_{2n-1} + (-1)^{2n} a_{2n} = s_{2n-1} + a_{2n}$$

$$\geq s_{2n-1} \geq s_1 \text{ (since } s_{2n-1} \text{ is increasing)}$$

Similarly, we can show that $s_{2n-1} \leq s_2$

Hence both s_{2n} and s_{2n-1} converge.

$$\text{Let} \lim s_{2n} = L_2, \qquad \lim s_{2n-1} = L_2$$

$$s_{2n} = s_{2n-1} + a_{2n}$$

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

$$L_2 = L_1 + 0 \Rightarrow L_1 = L_2 \text{ (we were given that} \lim a_n = 0)$$

The even and odd subsequences both converge to the same limit. Call this common limit L. Since the odd and even subsequences converge to L, we must have $\lim a_n = L$.

Absolute Convergence

Before moving on to the final 2 tests, I will introduce the idea of **absolute convergence**.

A sum $\sum a_k$ **converges absolutely** iff $\sum |a_k|$ converges. If a sum has all positive terms, then absolute convergence is the same a regular convergence. It is not too difficult to see that if a sum converges absolutely, it converges regularly. An actual proof:

$$\text{Suppose } \sum_{k=1}^{\infty} |a_k| \text{ converges}$$

$$\text{Then } 2\sum_{k=1}^{\infty} |a_k| = \sum_{k=1}^{\infty} 2|a_k| \text{ converges}$$

$$2|a_k| = |a_k| + |a_k| \geq |a_k| + a_k,$$

$$\text{and } -a_k \leq |a_k| \Rightarrow a_k + |a_k| \geq 0$$

$$\text{so we can use the comparison test to say:}$$

$$0 \leq |a_k| + a_k \leq 2|a_k| \text{ and } \sum_{k=1}^{\infty} 2|a_k| \text{ converges} \Rightarrow \sum_{k=1}^{\infty} (|a_k| + a_k) \text{ converges.}$$

$$\text{Let } s_n = \sum_{k=1}^{n} |a_k|, \qquad t_n = \sum_{k=1}^{n} (|a_k| + a_k)$$

$$\text{then } \sum_{k=1}^{n} a_k = t_n - s_n$$

$$t_n \text{ and } s_n \text{ converge, hence } \sum_{k=1}^{\infty} a_k \text{ converges.}$$

Another quick definition before moving on:

A sequence a_n **conditionally converges** iff it converges but not absolutely.

Ratio Test

We understand geometric series well. We know when they do and don't converge, and if they do converge we know what they converge to. Even if a series is not exactly geometric, if it becomes approximately geometric as you progress through the terms, we might be able to come up with a new test for convergence.

A geometric series has a common ratio $r = \frac{a_{n+1}}{a_n}$ for every n. When I say that an infinite series is "approximately geometric" what I really mean is that when n is large, we have an approximately common ratio. That is $\lim \frac{a_{n+1}}{a_n}$ exists. Then if $\left|\lim \frac{a_{n+1}}{a_n}\right| = \lim \left|\frac{a_{n+1}}{a_n}\right| < 1$ we may suspect convergence (this limit is the approximate common ratio for large n). We might also suspect the series to diverge if this limit is > 1.

We have already 2 infinite series where this limit is 1, ($\sum \frac{1}{k}$ which diverges and $\sum \frac{1}{k^2}$ which converges) so clearly if the limit is 1, this tells us nothing.

Theorem 3.7: **Ratio Test**

$$\text{Let } r = \lim \left|\frac{a_{n+1}}{a_n}\right|$$

$$\text{If } r < 1 \text{ then } \sum_{k=1}^{\infty} a_k \text{ converges absolutely (a)}$$

$$\text{If } r > 1 \text{ then } \sum_{k=1}^{\infty} a_k \text{ diverges (b)}$$

pf:

Part (a)

The idea is that the series is approximately geometric as k gets large, so we will attempt to compare against a geometric series. We want convergence, so we want $a_k \leq$ some geometric series with common ratio < 1.

$$\text{Let } r = \lim \left|\frac{a_{n+1}}{a_n}\right| < 1$$

Using ε-N:

$$\text{for every } \varepsilon > 0, \quad \exists_{N \in \mathbb{N}} \text{ such that}, \quad \forall_{n \geq N}:$$

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

$$\Rightarrow r - \varepsilon < \frac{|a_{n+1}|}{|a_n|} < r + \varepsilon$$

$$\Rightarrow |a_{n+1}| < (r + \varepsilon)|a_n|$$

$$\text{let } j \in \mathbb{N}, \text{meaning } N + j > N$$

$$0 \leq \left|a_{N+j}\right| < (r+\varepsilon)\left|a_{N+j-1}\right| < (r+\varepsilon)^2\left|a_{|N+j-2}\right| < \cdots < (r+\varepsilon)^j|a_N|$$

This starting to look like a geometric series. We want convergence so we want $(r + \varepsilon) < 1$. We have $r < 1$ and the above works $\forall_{\varepsilon > 0}$ so we can choose any $\varepsilon > 0$ we like.

$$\text{Let } \varepsilon = \frac{1-r}{2} > 0 \text{ and then}$$

$$r + \varepsilon = \frac{r+1}{2} < 1.$$

As a geometric series:

$$\sum_{j=1}^{\infty}(r+\varepsilon)^j|a_N| \text{ converges}$$

By the comparison test:

$$\sum_{j=1}^{\infty}\left|a_{N+j}\right| \text{ converges}$$

$$\text{but } \sum_{k=1}^{\infty}|a_k| = \sum_{k=1}^{N}|a_k| + \sum_{k=N+1}^{\infty}|a_k| = \sum_{k=1}^{N}|a_k| + \sum_{j=1}^{\infty}\left|a_{N+j}\right|$$

The first series is finite, the second converges, hence $\sum|a_k|$ converges, i.e., $\sum a_k$ converges absolutely.

Part (b)

We want to compare against a geometric series which diverges.

$$\text{Let } r = \lim\left|\frac{a_{n+1}}{a_n}\right| > 1$$

Following a similar idea to (a):

$$\text{let } \varepsilon = \frac{r-1}{2} > 0 \text{ then } r - \varepsilon = \frac{r+1}{2} > 1$$

There exists an $n \in \mathbb{N}$ such that $\forall_{n \geq N}, \left|\left|\frac{a_{n+1}}{a_n}\right| - r\right| < \varepsilon$.

$$r - \varepsilon < \frac{|a_{N+1}|}{|a_N|} < r + \varepsilon$$

$$|a_{N+1}| > (r - \varepsilon)|a_N|$$

Let $j \in \mathbb{N},\ N + j > N$

$$|a_{N+j}| > (r - \varepsilon)|a_{N+j-1}| > (r - \varepsilon)^2|a_{N+j-2}| > \cdots > (r - \varepsilon)^j|a_N| \geq 0$$

I know that:

$$\lim (r - \varepsilon)^j |a_N| = \infty \text{ (since } r - \varepsilon > 1)$$

Hence $\lim|a_{N+j}| = \infty$, which means a_{N+j} does not converge to 0.

$$\Rightarrow \sum_{j=1}^{\infty} a_{N+j} = \sum_{k=N+1}^{\infty} a_k \text{ diverges}$$

$$\Rightarrow \sum_{k=1}^{\infty} a_k \text{ diverges}$$

We can actually say a little bit more about the ratio test in the case where the limit does not exist.

Even if $\lim \left|\frac{a_{k+1}}{a_k}\right|$ does not exist, we can still guarantee absolute convergence if $\left|\frac{a_{k+1}}{a_k}\right|$ stays below 1 and does not get arbitarrily close to it. Precisely, if there exists a $q < 1$ such that $\left|\frac{a_{k+1}}{a_k}\right| \leq q < 1$ for all sufficiently large k (i.e., $\left|\frac{a_{k+1}}{a_k}\right|$ *may* get arbitrarily close to q but it will always have a distance of at least $(1 - q)$ from 1 [it does not get arbitrarily close to 1]).

Similarly, even if $\left|\frac{a_{k+1}}{a_k}\right|$ does not converge to a limit > 1, if $\left|\frac{a_{k+1}}{a_k}\right| \geq 1$ for all sufficiently large k, then we have divergence.

Theorem 3.7(strong version): **A Stronger Ratio Test**

If there exists a $q < 1$ and $N \in \mathbb{N}$ such that $\left|\frac{a_{k+1}}{a_k}\right| \leq q < 1\ \forall_{k \geq N}$

$$\text{Then } \sum_{k=1}^{\infty} a_k \text{ converges absolutely}$$

If there exists $N \in \mathbb{N}$ such that $\left|\frac{a_{k+1}}{a_k}\right| \geq 1\ \forall_{k \geq N}$

$$\text{Then } \sum_{k=1}^{\infty} a_k \text{ diverges}$$

The new proof is very similar in spirit to the original, so I give an outline of the steps.

pf:

Suppose there is some N so that $k \geq N \Rightarrow \left|\frac{a_{k+1}}{a_k}\right| \leq q < 1 \Rightarrow |a_{k+1}| \leq q|a_k|$ (for $k \geq N$).

$|a_{k+1}| \leq q|a_k| \leq q^2|a_{k-1}| \leq \cdots \leq q^{k-N+1}|a_N|$.

We know that

$$\sum_{k=1}^{\infty} q^{k-N+1}|a_N| = |a_N|q^{1-N} \sum_{k=1}^{\infty} q^k$$

which converges since $0 \leq q < 1$. Hence by comparison $\sum_{k=1}^{\infty}|a_{k+1}|$ converges, $\sum_{k=1}^{\infty} a_k$ converges absolutely.

For the divergence case, let N be such that $k \geq N \Rightarrow \left|\frac{a_{k+1}}{a_k}\right| \geq 1 \Rightarrow |a_{k+1}| \geq |a_k|$.

$$|a_{k+1}| \geq \cdots \geq |a_{N+1}|$$

We know that $|a_{N+1}| > 0$ (since $\left|\frac{a_{N+1}}{a_N}\right| \geq 1 > 0$).

Hence $|a_{k+1}| \geq |a_{N+1}| > 0$ for every $k \geq N$, meaning $\lim|a_{k+1}| = \lim|a_k| \geq |a_{N+1}| > 0$.

In particular $\lim|a_k| \neq 0$

$$\lim|a_k| \neq 0 \Rightarrow \lim a_k \neq 0 \Rightarrow \sum_{k=1}^{\infty} a_k \text{ diverges}$$

We are done!

We could write the new criteria more concisely by writing them in terms of lim sup and lim inf:

$$\text{Absolute convergence if} \limsup x_n < 1$$

$$\text{Divergence if} \liminf x_n > 1$$

If $\liminf x_n \leq 1 \leq \limsup x_n$ (which includes the case $\lim x_n = 1$) the ratio test tells us nothing.

Root Test

The final test we will consider is the **Root Test**. It is very similar to the ratio test. The idea is again to suppose that the sequence is approximately geometric for large k. Then by taking the kth root, we get the common ratio $(r^k)^{\frac{1}{k}} = r$. If this is less than 1, we have convergence, if it is > 1 we have divergence.

$$\text{Theorem 3.8: } \textbf{Root Test}$$

$$\text{Let } r = \lim |a_k|^{\frac{1}{k}}$$

$$r < 1 \Rightarrow \sum_{k=1}^{\infty} a_k \text{ converges absolutely}$$

$$r > 1 \Rightarrow \sum_{k=1}^{\infty} a_k \text{ diverges}$$

pf:

We use essentially the same idea as before, in fact this proof is a bit easier.

$$\text{Let } r = \lim |a_k|^{\frac{1}{k}}$$

$$\text{for } r < 1:$$

$$\text{Let } \varepsilon = \frac{1-r}{2} \text{ then } r + \varepsilon < 1$$

$$\exists_{N \in \mathbb{N}} \text{ s.t.,} \qquad \forall_{n \geq N}, \qquad \left| |a_n|^{\frac{1}{n}} - r \right| < \varepsilon$$

$$\Rightarrow |a_n|^{\frac{1}{n}} < r + \varepsilon$$

$$\Rightarrow 0 \leq |a_n| < (r + \varepsilon)^n$$

Which immediately gets us our geometric series. By comparison,

$$r + \varepsilon < 1 \Rightarrow \sum_{n=1}^{\infty} (r+\varepsilon)^n \text{ converges} \Rightarrow \sum_{n=1}^{\infty} |a_n| \text{ converges}$$

$$\sum_{n=1}^{\infty} a_n \text{ converges absolutely.}$$

$$\text{for } r > 1:$$

$$\text{Let } \varepsilon = \frac{r-1}{2} > 0 \Rightarrow r - \varepsilon > 1$$

$$\exists_{N\in\mathbb{N}} \text{ s.t.}, \qquad \forall_{n\geq N}$$

$$\left||a_n|^{\frac{1}{n}} - r\right| < \varepsilon$$

$$\Rightarrow r - \varepsilon < |a_n|^{\frac{1}{n}} \Rightarrow |a_n| > (r-\varepsilon)^n$$

$$r - \varepsilon > 1 \Rightarrow \lim(r-\varepsilon)^n = \infty$$

$$\Rightarrow \lim|a_n| \text{ diverges}$$

$$\Rightarrow a_n \text{ does not converge to } 0$$

$$\Rightarrow \sum_{k=1}^{\infty} a_k \text{ diverges}$$

In the case where $\lim|a_k|^{\frac{1}{k}}$ does not exist we can use a similar idea as with the ratio test.

The strong versions of the ratio and root test are not all that useful for determining whether series converge, since in most cases the limit does exist. However the strong version of the root test will be incredibly useful for our purposes, since it always applies (excpet for $r = 1$).

Theorem 3.7 (strong version): **A Stronger Root Test**:

$$\text{If } r = \limsup|a_k|^{\frac{1}{k}} \text{ exists, then}$$

$$r < 1 \Rightarrow \sum_{k=1}^{\infty} a_k \text{ converges absolutely.}$$

$$r > 1 \Rightarrow \sum_{k=1}^{\infty} a_k \text{ diverges}$$

$$\text{If } \limsup|a_k|^{\frac{1}{k}} \text{ does not exist, } \sum_{k=1}^{\infty} a_k \text{ diverges}$$

This covers all possible cases, except for when the lim sup exists and equals 1, this case is still inconclusive.

The proof is similar to the original root test, using comparison with a geometric series.

pf:

Assume lim sup exists, call it r.

If $r < 1$, then let $q = \frac{r+1}{2}$ (half way between r and 1, so that $r < q < 1$). Then take $\varepsilon = \frac{1-r}{2} > 0$. Then $q = r + \varepsilon < 1$. By definition of lim sup, there exists an N such that $|a_k|^{\frac{1}{k}} < r + \varepsilon = q < 1$ for all $k \geq N$.

So we have that $|a_k|^{\frac{1}{k}} \leq q < 1$ for all $k \geq N$. (Similar to what we had for the stronger ratio test).

Then $|a_k| \leq q^k$. We know that $\sum_{k=1}^{\infty} q^k$ converges (geometric series with $0 \leq q < 1$).

Then by comparison, $\sum_{k=1}^{\infty} |a_k|$ converges, meaning $\sum_{k=1}^{\infty} a_k$ converges absolutely.

If $r > 1$, then let $q = \frac{r+1}{2}$ (again, half way between 1 and r, so that $1 < q < r$). Then take $\varepsilon = \frac{r-1}{2}$. Then $q = r - \varepsilon > 1$. By definition of lim sup, we now have $1 < q \leq |a_k|^{\frac{1}{k}}$ for infinitely many k (there are infinitely many terms which are close to r, though there may also be infinitely many terms which are not close to r). This becomes $|a_k| \geq q^k$ with $q > 1$.

We use these infinitely many terms to define the subsequence a_{k_j} such that $\left|a_{k_j}\right| \geq q^{k_j}$ for every j. Then, each $\left|a_{k_j}\right| \geq q^{k_j}$ with $q > 1$, meaning $\lim \left|a_{k_j}\right| \neq 0$, so $\lim a_{k_j} \neq 0$ (or does not exist), hence $\lim a_k \neq 0$ (or does not exist).

$$\text{Therefore, } \sum_{k=1}^{\infty} a_k \text{ diverges}$$

Finally, if lim sup does not exist, then the sequence $|a_k|^{\frac{1}{k}}$ is unbounded above.

Then there exists a subsequence $\left|a_{k_j}\right|^{\frac{1}{k_j}} \geq 2$ for every $j \in \mathbb{N}$, then $\left|a_{k_j}\right| \geq 2^{k_j}$. $2^{k_j} \to \infty$ so $\left|a_{k_j}\right| \to \infty$ so a_k does not $\to 0$ so $\sum_{k=1}^{\infty} a_k$ diverges.

That is all for tests for convergence. You can practice using them with the questions on the problem sheet.

Rearranging Infinite Series

We have dealt a lot with infinite series, and we have been careful not to do such things as rearranging the terms of the series. In this section we will see under which circumstances infinite series may and may not be rearranged, which is a very natural question to ask.

It might seem as though we should always be able to rearrange infininte series because, after all, they are just adding infinitely many things together, and order of addition does not matter. It is important to remember that infinite series are not just a sum of infinitely many things, but are the limit of a process of adding finitely many things in a particular order. Changing the order of terms being added changes this process and so it may change the limit.

Make sure you remember the difference between absolute and conditional convergence for this section because they are key. As a reminder:

Absolute Convergence: $\sum a_k$ converges and $\sum |a_k|$ converges.

Conditional Convergence: $\sum a_k$ converges but $\sum |a_k|$ diverges.

First I need to define a rearrangement of a series.

Let $\sum_{k=1}^{\infty} a_k$ be an infinite series, and let $\sigma(k): \mathbb{N} \to \mathbb{N}$ be a bijective function. (σ maps every natural number to one natural number. Every natural number is mapped to exactly once.) $\sigma(1), \sigma(2), \cdots$ is just some rearrangement of the natural numbers.

$$\text{Let } b_k - a_{\sigma(k)}$$

Then we call b_k a $\mathbf{Rearrangement}$ of a_k. We call $\sum_{k=1}^{\infty} b_k$ a rearrangement of $\sum_{k=1}^{\infty} a_k$.

We will start with the case of conditional convergence. It turns out (and isn't too difficult to understand why) that not only can rearranging change the limit, but in fact we can rearrange our series such that it converges to any given limit. I will fill in the details in a moment, but the idea is as follows:

Let $\sum_{k=1}^{\infty} a_k$ be conditionally convergent. The sum consists of infinitely many positive and negative terms. If $L > 0$, start adding positive terms until you first exceed L. Then start adding the negative terms until you first get below L. Then add positive terms until you exceed L again, continue this process to constrct a sequence of partial sums which alternates between going above and below L. Each time we add or subtract you can think of it as moving a distance $|a_k|$. If this distance, $|a_k| \to 0$, then it seems that the sum must converge, and it seems that it will converge to L.

This is called the **Riemann Rearrangement Theorem**, and of course we need to fill in the details.

To prove this theorem (and the one after this) we will first prove a related Lemma,

Lemma 3.9:

$$\sum_{k=1}^{\infty} |a_k| \text{ converges} \Leftrightarrow \sum_{k=1}^{\infty} a_k^- \text{ and } \sum_{k=1}^{\infty} a_k^+ \text{ are both convergent.}$$

$$\text{Where } a_k^- = \begin{cases} a_k & a_k < 0 \\ 0 & \text{otherwise} \end{cases} \text{ and } a_k^+ = \begin{cases} a_k & a_k \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

(Absolute convergence if and only iff the sum of negative and sum of negative terms both converge).

pf:

We are essentially splitting our sequence into the negative and positive terms.

$$\text{We have } |a_k| = a_k^+ - a_k^-$$

We can prove this by cases:

$$\text{Case 1: } a_k < 0 \Rightarrow |a_k| = -a_k \text{ and } a_k^+ - a_k^- = 0 - a_k = -a_k$$

$$\text{Case 2: } a_k \geq 0 \Rightarrow |a_k| = a_k \text{ and } a_k^+ - a_k^- = a_k - 0 = a_k$$

" $\Leftarrow$ "

$$\text{If } \sum_{k=1}^{\infty} a_k^- \text{ and } \sum_{k=1}^{\infty} a_k^+ \text{ are convergent,}$$

Meaning that $t_n^- := \sum_{k=1}^{n} a_k^-$ and $t_n^+ := \sum_{k=1}^{n} a_k^+$ are both convergent sequences,

$$\text{Then } t_n^+ - t_n^- = \sum_{k=1}^{n} (a_k^+ - a_k^-) = \sum_{k=1}^{\infty} |a_k| \text{ is convergent.}$$

" $\Rightarrow$ "

Note that $0 \leq -a_k^- \leq |a_k|$ and $0 \leq a_k^+ \leq |a_k|$, so by the comparison test,

$$\text{If } \sum_{k=1}^{\infty} |a_k| \text{ converges,} \qquad \text{then } \sum_{k=1}^{\infty} a_k^+, \sum_{k=1}^{\infty} -a_k^- \text{ converge}$$

Riemann Rearrangement Theorem (3.10):

$$\sum_{k=1}^{\infty} a_k \text{ is conditionally convergent} \implies \forall_{L\in\mathbb{R}}, \quad \exists_{\sigma:\mathbb{N}\to\mathbb{N}} \text{s.t.}, \quad \sum_{k=1}^{\infty} a_{\sigma(k)} = L$$

In plain English, if the series is conditionally convergent, then for any real number L, there exists a rearrangement (represented by a bijective function σ) such that the series of the rearranged terms is equal to the given L.

I think the proof for this theorem is by far the hardest in the course, so don't be ashamed if you choose to skip it.

The argument outlined earlier requires a few things to be true. We need to be able to guarantee that as we continue to add positive terms we will eventually exceed the given L (and when we add the negatives that we eventually get below L). For this we need to show that $\sum a_k^- \to -\infty$ and $\sum a_k^+ \to \infty$.

We also need that $\lim|a_k| = 0$.

This proof then consists of 3 parts:

- Show that $\lim|a_k| = 0$
- Show that $\sum_{k=1}^{\infty} a_k^+ = \infty$ and $\sum_{k=1}^{\infty} a_k^- = -\infty$
- Construct a rearrangement which works.

pf:

The first part is easy.

$$\sum_{k=1}^{\infty} a_k \text{ converges} \Rightarrow \lim a_k = 0 \Rightarrow \lim|a_k| = 0$$

For the second part, we will use the previous Lemma (3.9).

$$\sum_{k=1}^{\infty} |a_k| \text{ diverges} \Rightarrow \sum_{k=1}^{\infty} a_k^- \text{ or } \sum_{k=1}^{\infty} a_k^+ \text{ is divergent.}$$

We know that $\sum a_k^-$ is monotone decreasing and so if it is diverent, it must be unbounded below with limit $-\infty$. Similarly, if $\sum a_k^+$ diverges, it must have limit ∞. We need then only show that they both diverge. I know that (from the above) at least one of them diverges.

$$\text{Suppose } \sum_{k=1}^{\infty} a_k^+ \text{ diverges and } \sum_{k=1}^{\infty} a_k^- \text{ converges}$$

We know that $a_k = a_k^+ + a_k^-$ and that $\sum a_k$ converges (due to conditional convergence).

$$\sum_{k=1}^{n} a_k^+ = \sum_{k=1}^{n} (a_k - a_k^-) = \sum_{k=1}^{n} a_k - \sum_{k=1}^{n} a_k^-$$

Since both $\sum_{k=1}^{\infty} a_k$ and $\sum_{k=1}^{\infty} a_k^-$ converge, we must have that $\sum_{k=1}^{n} a_k^+$ converges, which is a contradiction.

We get a similar contradiction if we suppose that $\sum_{k=1}^{\infty} a_k^+$ converges and $\sum_{k=1}^{\infty} a_k^-$ diverges

The only possibility is then that both diverge and hence:

$$\sum_{k=1}^{\infty} a_k^+ = \infty \text{ and } \sum_{k=1}^{\infty} a_k^- = -\infty$$

Also note that this implies that there are infinitely many non-negative and infinitely many negative terms.

For the final part, we need to construct a bijective function σ (representing the rearrangement) such that $\sum_{k=1}^{\infty} a_{\sigma(k)} = L$.

We will define our bijective function constructively, via the method outlined previously:

If $L > 0$, let $\sigma(1)$ be the index of the greatest positive term. If $L \leq 0$, let $\sigma(1)$ be the index of the least negative term (so greatest magnitude).

For $n \geq 2$:

$$\text{If } \sum_{k=1}^{n-1} a_{\sigma(k)} < L \text{ then let } \sigma(n) \text{ be the index of the greastest unused non} - \text{negative term.}$$

$$\text{If } \sum_{k=1}^{n-1} a_{\sigma(k)} \geq L \text{ then let } \sigma(n) \text{ be the index of the least unused negative term.}$$

This means by definition of $\sigma(k)$ that the subsequence (of $a_{\sigma(k)}$) of non-negative terms is decreasing and the subsequence (of $a_{\sigma(k)}$) of negative terms is increasing.

Now to show that $\sum_{k=1}^{\infty} a_{\sigma(k)} = L$:

$$\text{Let } S_n = \sum_{k=1}^{n} a_{\sigma(k)} \text{ (we show that } \lim S_n = L\text{)}$$

$$\text{Let } \varepsilon > 0$$

$$\text{Since } \lim |a_k| = 0, \qquad \exists_{N_1 \in \mathbb{N}} \text{s.t.,} \qquad \forall_{k \geq N_1}, \qquad |a_k| < \varepsilon$$

$$\text{Since } \sigma \text{ is bijective,} \quad \exists_{N_2 \geq N_1} \text{ s.t.,} \quad \forall_{k \geq N_2}, \quad \sigma(k) \geq N_1$$

$$\text{Then since } N_2 \geq N_2 \Rightarrow \sigma(N_2) \geq N_1 \Rightarrow \left|a_{\sigma(N_2)}\right| < \varepsilon$$

$$\text{There exists an } N \geq N_2 \text{ such that } S_{N-1} < L \leq S_N$$

(since this happens infinitely many times)

$$\text{Then let } n \geq N$$

Consider the subsequence of S_n which has $S_{n_j} \geq L$. We next show that this subsequence is bounded by $L \leq S_{n_j} < L + \varepsilon$.

Let $S_{n_j} \geq L$:

If $S_{n_j-1} < L$ then $S_{n_j} = S_{n_j-1} + a_{\sigma(n_j)} = S_{n_j-1} + \left|a_{\sigma(n_j)}\right| < L + \left|a_{\sigma(n_j)}\right| < L + \varepsilon$

If $S_{n_j-1} \geq L$ then $S_{n_j} = S_{n_j-1} + a_{\sigma(n_j)} = S_{n_j-1} - \left|a_{\sigma(n_j)}\right| < S_{n_j-1}$.

In this case if $S_{n_j-2} < L$, we then have $S_{n_j-1} < L + \varepsilon \Rightarrow S_{n_j} < S_{n_j-1} < L + \varepsilon$.

If $S_{n_j-2} \geq L$, we keep going until eventually we get something like $S_{n_j} < S_{n_j-1} < \cdots < S_{n_j-M} < L + \varepsilon \Rightarrow S_{n_j} < L + \varepsilon$ (which will definitely happen eventually since in the worst case we will reach $S_N < S_{N-1}$ and $S_{N-1} < L < L + \varepsilon$).

We now have (since $L - \varepsilon < L$), that $L - \varepsilon < S_{n_j} < L + \varepsilon \Rightarrow \left|S_{n_j} - L\right| < \varepsilon$ meaning this subsequence of terms $\geq L$ converges to L.

Through a symmetric argument, we argue that the subsequence of terms $S_{n_j} < L$ converges to L.

Since these subsequences partition S_n, we have that $\lim S_n = L$, which was what we wanted.

This is probably the hardest proof in this entire course.

We have just seen that you cannot rearrange series which do not converge absolutely and expect the same limit. The idea for Riemann Rearrangement was entirely dependent upon the series being *conditionally* convergent, so perhaps if a series is absolutely convergence, we can rearrange sums without the limit changing. It turns out that this is indeed the case.

The typical proof for this can be difficult to follow, but I have an easier way. I will start with a simple special case which is (relatively) easy to prove. We will then be able to use this special case to prove the general case.

$$\text{Theorem 3.11: If } s = \sum_{k=1}^{\infty} a_k \text{ converges, } a_k \geq 0 \text{ for every } k$$

$$\text{and } a_{\sigma(k)} \text{ is a rearrangement of } a_k,$$

$$\text{Then } \sum_{k=1}^{\infty} a_{\sigma(k)} = s \text{ and converges absolutely}$$

This is a specil case because $|a_k| = a_k$ and so the sum $\sum a_k$ converges absolutely.

pf:

$$\text{Let } s_n = \sum_{k=1}^{n} a_k\,, \qquad t_n = \sum_{k=1}^{n} a_{\sigma(k)}\,, \qquad \text{fix } n$$

$$\text{Let } N = \max\{\sigma(1), \sigma(2), \cdots, \sigma(n)\}$$

Then the first N terms of a_k must contain $a_{\sigma(1)}, \cdots, a_{\sigma(n)}$ (it may contain other terms as well, all of which are positive).

$$\Rightarrow 0 \leq \sum_{k=1}^{n} a_{\sigma(k)} \leq \sum_{k=1}^{N} a_k$$

$$(0 \leq t_n \leq s_N)$$

[The second sum contains all terms of the first sum and potentially more].

$$\lim s_n = s \text{ and } s_n \text{ is increasing, hence } s \text{ is an upper bound for } s_n$$

$$\Rightarrow 0 \leq t_n \leq s_N \leq s$$

$$(\text{critically } 0 \leq t_n \leq s, \qquad t_n \text{ is bounded})$$

Also t_n is increasing so, by monotone convergence $\lim t_n = \sum_{k=1}^{\infty} a_{\sigma(k)}$ converges (again absolutely). Let $t = \sum_{k=1}^{\infty} a_{\sigma(k)}$

We now have $0 \leq t \leq s$.

Let $b_k = a_{\sigma(k)}$, then $b_{\sigma^{-1}(k)} = a_k$ (σ is bijectie so has an inverse).

Finally, we repeat the same argument again, with

$$\text{Let } t_n = \sum_{k=1}^{n} b_k\,, \qquad s_n = \sum_{k=1}^{n} b_{\sigma^{-1}(k)}$$

To show that $0 \leq s \leq t$. Hence $s = t$:

$$\sum_{k=1}^{\infty} a_k = \sum_{k=1}^{\infty} a_{\sigma(k)}$$

We are now ready for the general case:

Theorem 3.12:

$$\text{If } s = \sum_{k=1}^{\infty} a_k \text{ converges absolutely and } a_{\sigma(k)} \text{ is a rearrangement of } a_k,$$

$$\text{Then } \sum_{k=1}^{\infty} a_{\sigma(k)} = s \text{ and converges absolutely}$$

pf:

For this we utilise the special case. We use a_k^+, a_k^- as we defined them earlier.

By Lemma 3.9, $\sum_{k=1}^{\infty} a_k^+$ and $\sum_{k=1}^{\infty} a_k^-$ both converge.

Recall that $a_k = a_k^+ + a_k^-$

$$\text{Note: } \sum_{k=1}^{n} a_k = \sum_{k=1}^{n} a_k^+ + \sum_{k=1}^{n} a_k^- = \sum_{k=1}^{n} a_k^+ - \sum_{k=1}^{n} (-a_k^-)$$

$$\text{Let } s^+ = \sum_{k=1}^{\infty} a_k^+ \text{ and } s^- = \sum_{k=1}^{\infty} (-a_k^-)$$

Then $s = s^+ - s^-$

$$\text{Let } t_n = \sum_{k=1}^{n} a_{\sigma(k)} = \sum_{k=1}^{n} a_{\sigma(k)}^+ - \sum_{k=1}^{n} \left(-a_{\sigma(k)}^-\right)$$

Since $a_k^+ \geq 0$ and $-a_k^- \geq 0$ we can use the special case:

$$\sum_{k=1}^{n} a_{\sigma(k)}^+ \text{ and } \sum_{k=1}^{n} \left(-a_{\sigma(k)}^-\right) \text{ both converge, with limits of}$$

$$\sum_{k=1}^{\infty} a_{\sigma(k)}^+ = s^+ \text{ and } \sum_{k=1}^{\infty} \left(-a_{\sigma(k)}^-\right) = s^-$$

Hence t_n converges absolutely (Lemma 3.9) with

$$\lim t_n = \sum_{k=1}^{\infty} a_{\sigma(k)}^+ - \sum_{k=1}^{\infty} \left(-a_{\sigma(k)}^-\right) = s^+ - s^- = s$$

$$s = \lim t_n$$

$$\sum_{k=1}^{\infty} a_k = \sum_{k=1}^{\infty} a_{\sigma(k)}$$

And we are done!

The difficult part of this section is now over. Good job if you made it through. Next we will look at something called the **Cauchy Product** which gives us a way to write the product of two sums as the sum of a sum.

We define:

$$\text{Def: } \sum_{n=0}^{\infty} \left(\sum_{k=0}^{n} a_k b_{n-k} \right) \text{ is the "Cauchy Product" of } \sum_{k=0}^{\infty} a_k \text{ and } \sum_{k=0}^{\infty} b_k$$

Cauchy Product Theorem (3.13):

$$\text{If } \sum_{k=0}^{\infty} a_k \text{ and } \sum_{k=0}^{\infty} b_k \text{ converge absolutely}$$

$$\text{Then } \left(\sum_{k=0}^{\infty} a_k \right) \times \left(\sum_{k=0}^{\infty} b_k \right) = \sum_{n=0}^{\infty} \left(\sum_{k=0}^{n} a_k b_{n-k} \right)$$

To explain what is going on here, we will arrange the terms of the product in a grid.

$$\text{We expand: } (a_0 + a_1 + a_2 + \cdots)(b_0 + b_1 + b_2 + \cdots)$$

$$= a_0(b_0 + b_1 + b_2 + \cdots)$$

$$+a_1(b_0 + b_1 + b_2 + \cdots)$$

$$+a_2(b_0 + b_1 + b_2 + \cdots)$$

$$\vdots$$

$$= a_0b_0 + a_0b_1 + a_0b_2 + \cdots$$

$$+a_1b_0 + a_1b_1 + a_1b_2 + \cdots$$

$$+a_2b_0 + a_2b_1 + a_2b_2 + \cdots$$

$$\vdots$$

(Expanding infinite series is okay since we are essentially saying that $\left(\sum_{i=0}^{\infty} a_i\right)\left(\sum_{j=0}^{\infty} b_j\right) = \sum_{i=0}^{\infty}\left(a_i \sum_{j=0}^{\infty} b_j\right) = \sum_{i=0}^{\infty}\left(\sum_{j=0}^{\infty}\left(a_i b_j\right)\right)$).

We can arrange these terms in a grid:

$$\begin{matrix} a_0b_0 & a_0b_1 & a_0b_2 & \cdots \\ a_1b_0 & a_1b_1 & a_1b_2 & \cdots \\ a_2b_0 & a_2b_1 & a_2b_2 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{matrix}$$

We then count the term diagonally:

$$\text{Zeroeth diagonal: } (a_0b_0)$$

$$\text{First Diagonal: } (a_0b_1 + a_1b_0)$$

$$\text{Second Diagonal: } (a_0b_2 + a_1b_1 + a_2b_0)$$

$$\vdots$$

$$n\text{th diagonal: } \sum_{k=0}^{n} a_k b_{n-k}$$

Adding all the diagonals together:

$$\sum_{n=0}^{\infty} \sum_{k=0}^{n} a_k b_{n-k}$$

The Cauchy Product!

Note that this explanation required that we rearrange the terms of the series. This is okay since both series are absolutely convergent.

I won't give any examples of using the Cauchy product here because there will be one particularly important example in the next (and final) section of this chapter.

The Exponential and Logarithm

The exponential and logarithm functions are familiar to every calculus student.

We could have defined these back in the chapter on sequences using $e^x = \lim\left(\left(1+\frac{x}{n}\right)^n\right)$, but in order to do that you need to show that the sequence converges. The proof isn't difficult, but I found it hard to motivate the steps so I have left this until now.

We define:

$$\exp(x) := \sum_{n=1}^{\infty} \frac{x^k}{n!}$$

The reason why I am using $\exp(x)$ instead of the more familiar e^x is so that we don't habitually familiar properties such as $e^x e^y = e^{x+y}$ and $e^0 = 1$. To fully justify the notation e^x we need to prove these properties.

We need to also prove that this **sum converges** or else the definition is meaningless.

Using the **Ratio Test**: $\lim\left|\frac{x^{n+1}}{(n+1)!} \div \frac{x^n}{n!}\right| = \lim\left|\frac{x}{n+1}\right| = 0 < 1$ for every x, hence we have convergence for every x, the definition is fine.

Also define:

$$e := \exp(1) = \sum_{n=1}^{\infty} \frac{1}{n!}$$

We will next prove the familiar properties of this function. Most easily, that $\exp(0) = 1$:

(Note that we define $0^0 = 1$)

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

Perhaps the most important property to justify the notation e^x is that $e^x e^y = e^{x+y}$.

$$\exp(x)\exp(y) = \left(\sum_{k=0}^{\infty} \frac{x^k}{k!}\right)\left(\sum_{k=0}^{\infty} \frac{y^k}{k!}\right)$$

We can apply the **Cauchy Product**:

$$\exp(x)\exp(y) = \sum_{n=0}^{\infty}\left(\sum_{k=0}^{n} \frac{x^k}{k!} \cdot \frac{y^{n-k}}{(n-k)!}\right)$$

We can simplify the inside sum by using the binomial theorem:

$$\sum_{k=0}^{n} \frac{x^k}{k!} \cdot \frac{y^{n-k}}{(n-k)!} = \frac{1}{n!} \sum_{k=0}^{n} \frac{n!}{k!\,(n-k)!} x^k y^{n-k}$$

$$\frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} = \frac{1}{n!} (x+y)^n$$

$$\text{Finally: } \exp(x)\exp(y) = \sum_{n=0}^{\infty} \frac{1}{n!}(x+y)^n = \exp(x+y) \text{ as required.}$$

Very similar to this, we wish to prove that $\exp(-x) = \frac{1}{\exp(x)}$

We can do this quickly by noting that $1 = \exp(0) = \exp(x-x) = \exp(x)\exp(-x) \Longrightarrow \exp(-x) = \frac{1}{\exp(x)}$ as required.

Note that this also means that $e^n = e^{n-1}e$ (we already defined $e = e^1$) and so this definition matches our definition for integer exponents.

Another familiar property is that $e^x > 0$ for every x. For $x \geq 0$ this is easy since $e^x = 1 + \sum_{k=1}^{\infty} \frac{x^k}{k!} \geq 1 > 0$. For $x < 0$ we know that $-x > 0$ and hence $e^{-x} = \frac{1}{e^x} > 0 \Rightarrow e^x > 0$.

I will next show that e^x is strictly increasing, meaning whenever $x < y$ we have $e^x < e^y$.

$$\text{Let } x < y$$

$$\text{Define } h = y - x > 0 \Rightarrow y = x + h$$

$$\text{Then } e^y = e^{x+h} = e^x e^h$$

$$= e^x \left(1 + \sum_{n=1}^{\infty} \frac{h^n}{n!}\right) > e^x \text{ (since } h > 0)$$

We finally wish to define the logarithm as the inverse of the exponential function, but in order to do this we need to know that e^x is bijective.

We are taking the domain to be $\mathbb{R}$ (e^x is defined for all real numbers).

We will take the codomain to be the set of positive real numbers (since $e^x > 0$).

Theorem 3.14:

$e^x: \mathbb{R} \to \mathbb{R}^+$ is bijective

First we show that e^x is injective (each output has at most one input). Precisely, that

$$x \neq y \Rightarrow e^x \neq e^y$$

This is easy since if $x \neq y$ then either $x < y \Rightarrow e^x < e^y$ or $y < x \Rightarrow e^y < e^x$. In either case $e^x \neq e^y$.

Second, we show that e^x is surjective (every output has at leat one input).

$$\text{Let } y > 0$$

I need to show that there exists a real number x such that $y = e^x$. We will use the axiom of completeness.

Suppose for now that $y > 1$. Consider the set $X = \{x: e^x < y\}$. $0 \in X$ hence non-empty.

I know that $e^x \geq 1 + x$ for $x > 0$ (first 2 terms of series definition) hence e^x is unbounded above. Let M be such that $e^M \geq y$ (possible due to unboundness).

Then $x \in X \Rightarrow e^x < y \leq e^M \Rightarrow x < M$ so M is an upper bound for X. We can now use AoC on the set X.

Let $\alpha = \sup X$. I want to show that $y = e^\alpha$. We proceed by contradiction.

$$\text{Suppose } e^\alpha < y$$

We suspect that we will find that α is not an upper bound which will give us a contradiction.

Note that $\left(\alpha + \frac{1}{n}\right) > \alpha$. I want to choose an n such that $\left(\alpha + \frac{1}{n}\right) \in X$, meaning I want $e^{\left(\alpha + \frac{1}{n}\right)} < y$.

$$\Leftrightarrow e^{\alpha + \frac{1}{n}} = e^\alpha e^{\frac{1}{n}} = e^\alpha \sum_{k=0}^{\infty} \frac{1}{k!\, n^k} = e^\alpha + e^\alpha \sum_{k=1}^{\infty} \frac{1}{k!\, n^k} < y$$

$$\Leftrightarrow \sum_{k=1}^{\infty} \frac{1}{k!\, n^k} < e^{-\alpha}(y - e^\alpha)$$

If I can find an n to make this inequality true, then $\alpha + \frac{1}{n} \in X$.

Note that $0 \leq \frac{1}{k!n^k} \leq \frac{1}{n^k}$ so by the comparison test, $\sum_{k=1}^{\infty} \frac{1}{k!n^k}$ converges with limit$\leq \frac{\frac{1}{n}}{1-\frac{1}{n}} = \frac{1}{n-1}$.

Choose n such that $\frac{1}{n-1} < e^{-\alpha}(y - e^{\alpha})$ (archimedes).

$$\text{Then } \sum_{k=1}^{\infty} \frac{1}{k!\, n^k} \leq \frac{1}{n-1} < e^{-\alpha}(y - e^{\alpha})$$

$$\Rightarrow \alpha + \frac{1}{n} \in X$$

Hence α is not an upper bound.

The $e^{\alpha} > y$ case is similar. I leave this as an exercise for you.

Remember we assumed that $y > 1$, this was so that we could ensure that X was non-empty. If $y = 1$ then $e^0 = 1$. If $0 < y < 1$, we can re-cycle the reslult for $y > 1$.

$$\text{Let } 0 < y < 1$$

$$\Rightarrow \frac{1}{y} > 1$$

$$\Rightarrow \exists_{x \in \mathbb{R}} \text{s.t.}, \qquad e^x = \frac{1}{y}$$

$$\Rightarrow y = e^{-x} \text{ so } (-x) \in \mathbb{R} \text{ is our number.}$$

We are now done, and e^x is bijective. We are now allowed to define:

$$\text{def: the } \mathbf{logarithm} \text{ of } x,$$

$$\log x \coloneqq \exp^{-1} x$$

This function has a domain of $\mathbb{R}_+$ (it is defined only for positive real numbers).

You may be used to writing $\ln x$. I may use $\ln$ and $\log$ interchangeably, but mainly $\log$.

We can use the logaithm and exponential function to define arbitrary real powers.

$$\text{Def: } x^{\alpha} \coloneqq e^{\alpha \log x}, \text{for } x > 0$$

We need to check that this gives the same result as the existing definition if α is an integer. Let $x > 0$. $x^0 = e^{0 \log x} = e^0 = 1$ (as expected). Suppose $x^n = e^{n \log x}$ then $x^{n+1} = x^n x = e^{n \log x} e^{\log x} = e^{n \log x + \log x} = e^{(n+1)} \log x$ and $x^{n-1} = \frac{x^n}{x} = \frac{e^{n \log x}}{e^{\log x}} = e^{n \log x - \log x} = e^{(n-1) \log x}$, so by induction $x^n = e^{n \log x}$ for integers n.

Finally, we define:

$$\text{Def: } \log_a x = \frac{\log x}{\log a}$$

Which has the property that $a^{\log_a x} = x$ (prove this).

The proofs of the properties of logarithms and exponentials, are relatively easy and are on the problem sheet.

This definition is only good for $x > 0$ as otherwise the logarithm is undefined.

We can separately define:

$$0^\alpha = \begin{cases} 0 & \alpha \neq 0 \\ 1 & \alpha = 0 \end{cases}$$

And then for $x < 0$ and α rational, let $\alpha = \frac{m}{n}$ where m and n have no common factors except 1. Also we require n be odd to avoid square rooting negative numbers:

$$x^{\frac{m}{n}} \coloneqq \beta \text{ such that } \beta^n = x^m.$$

We do not define x^α for $x < 0$ and α irrational (this would need complex numbers).

This definition with all the edge cases is annoying, and we will not be using it. We only really care about the $x > 0$ case.

The definition becomes much simpler if we allow complex numbers as then we can extend the domain of the logarithm to include negatives. We can then use the $e^{\alpha \log x}$ definition always. We will of course be ignoring this, as this is *real* analysis.

Of course we know that e is $2.718281828\cdots$ (by using calculators or lengthy calculations) but we want to rigorously prove that $2.5 < e < 3$ to give us an idea of how big e is.

$e > 2.5$ is easy since $e = \sum_{k=0}^{\infty} \frac{1}{k!} = 1 + 1 + \frac{1}{2} + \sum_{k=3}^{\infty} \frac{1}{k!} > 1 + 1 + \frac{1}{2} = 2.5$,

$$\text{i.e.,} \qquad e > 2.5$$

$e < 3$ is a bit harder. We can do this by comparing with a geomtric series.

$(n+1)!$ grows faster than 2^n.

$$\text{You can prove by induction that } k! \geq 2^{k-1} \text{ for } k \in \mathbb{N} \text{ (problem sheet Q13)}$$

Meaning $\frac{1}{k!} \leq \frac{1}{2^{k-1}} = \left(\frac{1}{2}\right)^{k-1}$ (equality when $k = 1,2,$ otherwise strict inequality).

$$e = \sum_{k=0}^{\infty} \frac{1}{k!} = 1 + \sum_{k=1}^{\infty} \frac{1}{k!} < 1 + \sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^{k-1} = 1 + \frac{1}{1 - \frac{1}{2}} = 1 + 2 = 3$$

$$\text{i. e.,} \qquad e < 3$$

so $2.5 < e < 3$.

Limits of Functions and Continuity

The Limit of a Function

This section is one of the most important, as it makes formal many of the ideas from caluculus, namely limits of functions and continuity. We start by defining the limit of a function.

We write $\lim_{x\to c} f(x)$ to denote the limit of $f(x)$ as $x \to c$.

We already defined limits of sequences, and so we expect there to be some similarities here. The idea of the limit of a sequence was that for any positive ε you could find terms big enough that they were all within a distance ε from L. Here we will use the same idea, but instead of n being "large", we need x being "close to" c. By n being large, we meant $n \geq N$ (where N is some natural number). By close to c, we mean x has is within a distance of δ from c (where δ is some positive real number).

We can write "x is within a distance of δ from c" as $|x - c| < \delta$.

We also don't care about what happens *at* the point c, only near it, so we can exclude $x = c$ and write $0 < |x - c| < \delta$.

We then have the following definition:

$$\text{Def: } \lim_{x\to c} f(x) = L \text{ means:}$$

$$\forall_{\varepsilon>0}, \qquad \exists_{\delta>0} \text{ s.t.}, \quad 0 < |x - c| < \delta \Rightarrow |f(x) - L| < \varepsilon$$

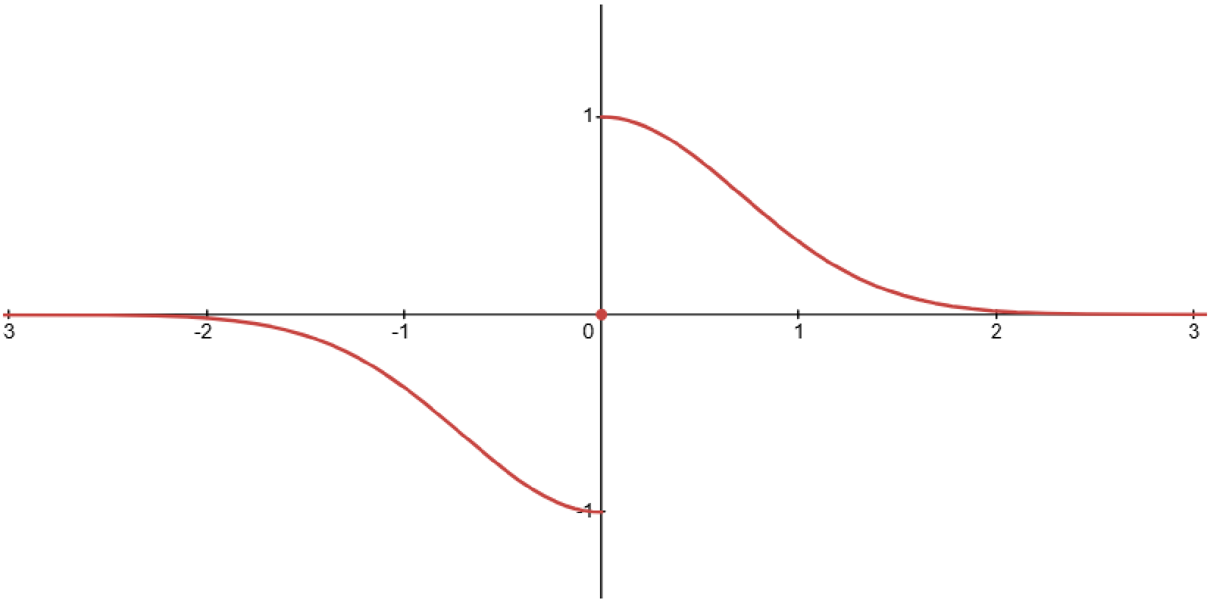
This is called the ε-δ definition.

In plain english, this is saying:

For any positive ε, there is some positive δ such that, for every real number x which is within a distance δ from c (but not equal to c), then $f(x)$ is within a distance ε from L.

In even more simple terms, when x is close to c, $f(x)$ is close to L.

If we have a function defined on an ineterval, (a, b),$(a, b]$,$[a, b)$, or $[a, b]$ then this definition does not allow a limit to exist at the endpoints a and b. We may also have a situation where the function seems to approach one value from the left and another from the right. E.g., here:



The limit as $x \to 0$ from the left appears to be -1, but from the right is 1. We hence define the notion of left and right limits. We write "the limit as $x \to c$ from the left of $f(x)$" as $\lim_{x \to c^-} f(x)$, and likewise the limit from the right as $\lim_{x \to c^+} f(x)$.

We define the left limit using the ε-δ definition above, but with $x < c$.

$$\text{Def: } \lim_{x \to c^-} f(x) = L \text{ means:}$$

$$\forall_{\varepsilon>0}, \qquad \exists_{\delta>0} \text{ s.t.}, \quad (x < c \text{ and } 0 < c - x < \delta) \Rightarrow |f(x) - L| < \varepsilon$$

Similarly:

$$\text{Def: } \lim_{x \to c^+} f(x) = L \text{ means:}$$

$$\forall_{\varepsilon>0}, \qquad \exists_{\delta>0} \text{ s.t.}, \qquad (x > c \text{ and } 0 < x - c < \delta) \Rightarrow |f(x) - L| < \varepsilon$$

It seems clear (from pictures) that the limt exists if and only if the left and right limits both exist and are equal, in which case the limit is equal to the left and right limits. The proof is straightforward.

Theorem 4.1:

$$\lim_{x \to c} f(x) \text{ exists} \Leftrightarrow \lim_{x \to c^-} f(x) \text{ and } \lim_{x \to c^+} f(x) \text{ both exist and are equal,}$$

$$\text{and also } \lim_{x \to c} f(x) = \lim_{x \to c^-} f(x) = \lim_{x \to c^+} f(x)$$

pf:

The "⇒" direction is rather trivial, but I encourage you to try it yourself anyway (problem sheet).

For the "⇐" direction:

$$\text{Suppose } \lim_{x\to c^+} f(x) = \lim_{x\to c^-} f(x) = L$$

$$\text{Let } \varepsilon > 0$$

$$\text{There exists } \delta_1 > 0 \text{ such that if } x < c \text{ and } 0 < c - x < \delta_1 \text{ then } |f(x) - L| < \varepsilon$$

$$\text{There exists } \delta_2 > 0 \text{ such that if } x > c \text{ and } 0 < x - c < \delta_2 \text{ then } |f(x) - L| < \varepsilon$$

$$\text{Choose } \delta = \min(\delta_1, \delta_2)$$

$$\text{Now let } x \text{ be a real number with } 0 < |x - c| < \delta$$

$$\Rightarrow x \neq c$$

$$\text{Either, } x < c \Rightarrow 0 < |x - c| = c - x < \delta \leq \delta_1 \Rightarrow |f(x) - L| < \varepsilon$$

$$\text{Or } x > c \Rightarrow 0 < |x - c| = x - c < \delta \leq \delta_2 \Rightarrow |f(x) - L| < \varepsilon$$

$$\text{Either way,} \qquad 0 < |x - c| < \delta \Rightarrow |f(x) - L| < \varepsilon$$

$$\text{Hence } \lim_{x\to c} f(x) = L$$

A standard exercise is to prove the following results:

$$\lim_{x\to c} f(x) + \lim_{x\to c} g(x) = \lim_{x\to c}\big(f(x) + g(x)\big)$$

$$\lim_{x\to c}\big(f(x)g(x)\big) = \lim_{x\to c} f(x) \lim_{x\to c} g(x)$$

$$\lim_{x\to c} \frac{f(x)}{g(x)} = \frac{\lim_{x\to c} f(x)}{\lim_{x\to c} g(x)}$$

$$\left[\text{the last one assuming } \lim_{x\to c} g(x) \neq 0 \text{ and } g(x) \neq 0 \text{ sufficiently close to } c\right]$$

from the ε-δ definition. I leave this as an exercise to you, but I will show you a lazier way to prove these same results in the next section.

Also left as an exercise is that limits are unique.

Finally, we define limits as $x \to \pm\infty$ in essentially the same way as how we defined the limit of a sequence.

$$\lim_{x\to\infty} f(x) = L \text{ means:}$$

$$\forall_{\varepsilon>0}, \qquad \exists_{M\in\mathbb{R}} \text{ s.t.,} \qquad \forall_{x\geq M}, \qquad |f(x) - L| < \varepsilon$$

Which reads, for every positive ε, there is some real number M, so that every number larger than M is close to L (within a distance ε).

Similarly:

$$\lim_{x \to -\infty} f(x) = L \text{ means:}$$

$$\forall_{\varepsilon>0}, \qquad \exists_{M\in\mathbb{R}}\, s.t., \qquad \forall_{x\leq M}, \qquad |f(x) - L| < \varepsilon$$

Limits of Functions as Limits of Sequences

Since we have already done so much work on limits of sequences, it would be a shame to start from scratch. We can formulate limis of functions in terms of limits of sequences which will allow to to recycle those old results.

If we consider a sequence x_n with $\lim x_n = c$, (meaning x_n gets closer to c) it seems that $f(x_n)$ gets close to L. Notice that $f(x_n)$ is itself a sequence, and we are proposing that $\lim f(x_n) = L$. Also, since we don't care about what happens at the point $x = c$ (f might even be undefined there) we stipulate that $x_n \neq c$ for any n.

$$\text{Theorem 4.1:}$$

$$\lim_{x \to c} f(x) = L \text{ iff}$$

$$\lim f(x_n) = L \text{ for every sequence } x_n \neq c \text{ with } \lim x_n = c$$

This is an "if and only if" statement, meaning we need to prove both directions.

pf:

$" \Rightarrow "$

$$\text{Let} \lim_{x \to c} f(x) = L, \qquad \text{Let } x_n \text{ be any sequence with } \lim x_n = c \text{ and } x_n \neq c$$

I want to show $\lim f(x_n) = L$. Let $\varepsilon > 0$

$$\text{We know there exists } \delta > 0 \text{ so that } 0 < |x - c| < \delta \Rightarrow |f(x) - L| < \varepsilon$$

We also know there exists some $N \in \mathbb{N}$ so that for each $n \geq N$, $0 < |x_n - c| < \delta$

(since $\lim x_n = c$ and $x_n \neq c$)

Putting this all together:

$$\text{Let } n \geq N \Rightarrow 0 < |x_n - c| < \delta \Rightarrow |f(x_n) - L| < \varepsilon$$

$$\text{Hence} \lim f(x_n) = L$$

$" \Leftarrow "$

Proving the left direction is a little tricky, but we can make it easier by using the contrapoisitive. This means instead of showing that $B \Rightarrow A$ we show that $\neg A \Rightarrow \neg B$.

$$\text{Suppose } \lim_{x \to c} f(x) \neq L \text{ (or that the limit does not exist).}$$

We want to show that there exists a sequence x_n with $x_n \neq c$ and $\lim x_n = c$ such that $\lim f(x_n) \neq L$.

Using $\lim_{x \to c} f(x) \neq L$ (or doesn't exist), we use the negation of the ε-δ definition.

$$\exists_{\varepsilon > 0} \text{ s.t.,} \qquad \forall_{\delta > 0}, \qquad \exists_x \text{ s.t.,} \qquad 0 < |x - c| < \delta \text{ but } |f(x) - L| \geq \varepsilon$$

We are given this $\varepsilon > 0$, and we can choose any $\delta > 0$. We choose $\delta_n = \frac{1}{n}$ because then

There exists x_n such that $0 < |x_n - c| < \frac{1}{n}$ but $|f(x_n) - L| \geq \varepsilon$.

We repeat this process for $n = 1,2,3 \cdots$ to create a sequence x_n such that $0 < |x_n - c| < \frac{1}{n}$ for every n, and $|f(x_n) - L| \geq \varepsilon$ for every n.

By squeezing, $|x_n - c| < \frac{1}{n} \Rightarrow \lim x_n = c$. Also $0 < |x_n - c|$ means $x_n \neq c$.

We now have an $\varepsilon > 0$ so that for any $N \in \mathbb{N}$, there exists an $n \geq N$ (any $n \geq N$ will do) so that $|f(x_n) - L| \geq \varepsilon$ (since this is true for every n), meaning $\lim f(x_n) \neq L$ (or does not exist) as required.

The reason for proving the above theorem is that it allows us to skip the usual work of proving all sorts of results directly from the ε-δ definition. For example:

Let x_n be a sequence with $\lim x_n = c$ and $x_n \neq c$.

$$\text{So then } \lim_{x \to c} f(x) \lim_{x \to c} g(x) = \lim f(x_n) \lim g(x_n) = \lim f(x_n) g(x_n) = \lim_{x \to c} f(x) g(x)$$

Where we have use the fact that $\lim a_n \lim b_n = \lim a_n b_n$.

We can then also prove the similar results for $\lim_{x \to c} \big(f(x) + g(x)\big)$ and $\lim_{x \to c} \frac{f(x)}{g(x)}$.

Continuity

An informal notion of continuity is that you can draw the graph of the function without taking your pen of the page. Alternatively, if x is close to c, then $f(x)$ is close to $f(c)$. How do we make this idea precise? We use the notion of limits.

$$\text{Def: We say } f \text{ is continuous at } c \text{ iff } \lim_{x\to c} f(x) = f(c)$$

$$\text{Def: We say } f \text{ is continuous iff it is continuous at every point in its domain}$$

As a small caveat to this, if f is defined on a closed interval $[a, b]$, we require only the right-hand limit for the left endpoint and only the left-hand limit for the right endpoint.

We can use the ε-δ definition of the limit to write the definition of continuity in terms of ε's and δ's.

Let f be a function with domain $X \subseteq \mathbb{R}$

$$f(x) \text{ is continuous means:}$$

$$\forall_{c\in X}, \qquad f \text{ is continuous at } c$$

Then using the ε-δ definition setting the limit $L = f(c)$:

$$\forall_{c\in X}, \qquad \forall_{\varepsilon>0}, \qquad \exists_{\delta>0}\text{s.t.,} \qquad 0 < |x-c| < \delta \Rightarrow |f(x) - f(c)| < \varepsilon$$

One final note is that when it comes to continuity, we actually do care about what happens at $x = c$. Clearly at $x = c$ we have both $|x - c| = 0 < \delta$ and we have $|f(x) - f(c)| = 0 < \varepsilon$ which means we can ignore the $0 <$ part:

$$\forall_{c\in X}, \qquad \forall_{\varepsilon>0}, \qquad \exists_{\delta>0}\text{s.t.,} \qquad |x-c| < \delta \Rightarrow |f(x) - f(c)| < \varepsilon$$

(For non-endpoints. If we have endpoints of a closed interval, we use the one-sided limits).

We can use this definition to show that certain functions are continuous.

$$\text{e.g., Show that } f(x) = k \text{ is continuous (where } k \text{ is some constant)}$$

Here the domain is $\mathbb{R}$. The inequality $|f(x) - f(c)| < \varepsilon$ becomes $|k - k| = 0 < \varepsilon$ which is always true since $\varepsilon > 0$ so we can choose any $\delta > 0$ and it will work (e.g., $\delta = 1$).

pf:

$$\text{Let } c \in \mathbb{R}, \qquad \text{Let } \varepsilon > 0$$

$$\text{Choose } \delta = 1$$

$$\text{Suppose } |x - c| < \delta$$

$$\text{Then } |f(x) - f(c)| = |k - k| = 0 < \varepsilon$$

Another simple example;

$$\text{e.g., Show that } f(x) = x \text{ is continuous.}$$

Here the domain is again $\mathbb{R}$. The inequality $|f(x) - f(c)| < \varepsilon$ becomes $|x - c| < \varepsilon$. Since we will assume the $|x - c| < \delta$, setting $\delta = \varepsilon$ will work.

pf:

$$\text{Let } c \in \mathbb{R}, \qquad \text{Let } \varepsilon > 0$$

$$\text{Choose } \delta = \varepsilon$$

$$\text{Suppose } |x - c| < \delta$$

$$|f(x) - f(c)| = |x - c| < \delta = \varepsilon$$

A more difficult example;

$$\text{e.g., Show that } f(x) = \sqrt{x} \text{ is continuous.}$$

The domain of the $\sqrt{}$ function is $[0, \infty)$ (positive real numbers). We first consider the endpoint 0. We want to show that $\lim_{x \to 0^+} \sqrt{x} = \sqrt{0} = 0$. The inequality $|f(x) - f(0)| < \varepsilon$ will become $|\sqrt{x} - 0| = \sqrt{x} < \varepsilon$. We already know that $x - 0 = x < \delta$ so setting $\delta = \varepsilon^2$ should get the job done.

$$\text{Let } c = 0, \qquad \text{Let } \varepsilon > 0$$

$$\text{Choose } \delta = \varepsilon^2 > 0$$

$$\text{Suppose } x - 0 < \delta$$

$$\text{Then } x < \delta = \varepsilon^2 \Rightarrow \sqrt{x} < \varepsilon \text{ (since } \sqrt{x} \text{ and } \varepsilon \text{ are non-negative)}$$

$$|f(x) - f(c)| = |\sqrt{x} - \sqrt{0}| = \sqrt{x} < \varepsilon$$

Next for $c > 0$ we use the regular (two-sided) limit. First the idea:

$|f(x) - f(c) < \varepsilon$ becomes $|\sqrt{x} - \sqrt{c}| < \varepsilon$. We would like to write this in terms of $|x - c|$. We can use difference of squares to write $|x - c| = |\sqrt{x} - \sqrt{c}||\sqrt{x} + \sqrt{c}|$ (which is fine since $x, c > 0$).

We know that $|x - c| = |\sqrt{x} - \sqrt{c}||\sqrt{x} + \sqrt{c}| < \delta$. Also we know that $|a - b| \leq |a + b|$ whenever a and b are non-negative (this seems reasonable but you can prove it rigorously by the triangle inequality). $|a - b| = |a + (-b)| \leq |a| + |-b| = |a| + |b| = a + b = |a + b|$.

In our example, $\sqrt{x}$ and $\sqrt{c}$ are non-negative, so we can apply this fact.

$$|\sqrt{x} - \sqrt{c}| \leq |\sqrt{x} + \sqrt{c}| \Rightarrow |\sqrt{x} - \sqrt{c}|^2 \leq |\sqrt{x} - \sqrt{c}||\sqrt{x} + \sqrt{c}| < \delta$$

So (since we have $\left|\sqrt{x}-\sqrt{c}\right|^2 < \delta$) a wise choice might be $\delta = \varepsilon^2$ (the same as we had before). Putting this all together:

pf:

$$\text{Let } c > 0, \qquad \text{Let } \varepsilon > 0$$

$$\text{Choose } \delta = \varepsilon^2 > 0$$

$$\text{Suppose } |x-c| < \delta$$

$$\text{Then } \left|\sqrt{x}-\sqrt{c}\right|^2 = \left|\sqrt{x}-\sqrt{c}\right|\left|\sqrt{x}-\sqrt{c}\right| \leq \left|\sqrt{x}-\sqrt{c}\right|\left|\sqrt{x}+\sqrt{c}\right| = |x-c| < \delta = \varepsilon^2$$

$$\Rightarrow |f(x)-f(c)| = \left|\sqrt{x}-\sqrt{c}\right| < \varepsilon$$

For one final example, I will demonstrate the continuity of the exponential function, e^x.

First the idea, then a streamlined proof.

The inequality $|f(x)-f(c)| < \varepsilon$ will be $|e^x - e^c| < \varepsilon$. Using properties of the exponential function, this becomes $e^c|e^{x-c}-1| < \varepsilon$ which is useful because e^c is just some constant. Using the series definition, this becomes:

$$e^c\left|\sum_{k=1}^{\infty}\frac{(x-c)^k}{k!}\right|$$

We want to make this $< \varepsilon$. First, I use the triangle inequality on the infinite series (I will justify this in a moment).

$$e^c\left|\sum_{k=1}^{\infty}\frac{(x-c)^k}{k!}\right| \leq e^c\sum_{k=1}^{\infty}\frac{|x-c|^k}{k!}$$

Which is better because it is now in terms of $|x-c| < \delta$.

$$e^c\sum_{k=1}^{\infty}\frac{|x-c|^k}{k!} < e^c\sum_{k=1}^{\infty}\frac{\delta^k}{k!} \text{ (comparison test)}$$

Finally, comparing this to a geometric series

$$e^c\sum_{k=1}^{\infty}\frac{\delta^k}{k!} \leq e^c\sum_{k=1}^{\infty}\delta^k \left(\text{by the comparison test since} \frac{\delta^k}{k!} \leq \delta^k\right)$$

$$= \frac{e^c\delta}{1-\delta} \text{ (Assuming } \delta < 1 \text{ for convergence).}$$

If we can have $\frac{e^c\delta}{1-\delta} = \varepsilon$, we will be done.

$$\frac{e^c\delta}{1-\delta} = \varepsilon \Rightarrow \delta = \frac{\varepsilon}{e^c+\varepsilon} \in (0,1) \text{ (so indeed } \delta < 1)$$

The use of the infinite triangle inequality follows directly from the finite version.

Let $b_n = |\sum_{k=1}^{n} a_k|, c_n = \sum_{k=1}^{n} |a_k|$ (assuming these both converge). Then by the finite triangle inequality, $b_n \leq c_n$. Hence $\lim b_n \leq \lim c_n$ (squeeze theorem) $\Rightarrow |\sum_{k=1}^{\infty} a_k| \leq \sum_{k=1}^{\infty} |a_k|$.

Writing this all into a more streamlined proof:

pf:

$$\text{Let } c \in \mathbb{R}, \qquad \text{Let } \varepsilon > 0$$

$$\text{Choose } \delta = \frac{\varepsilon}{e^c+\varepsilon} > 0$$

$$0 < \varepsilon < e^c + \varepsilon \Rightarrow 0 < \frac{\varepsilon}{e^c+\varepsilon} < 1, \qquad \text{i.e.}, 0 < \delta < 1$$

$$\text{Suppose } |x-c| < \delta$$

$$|f(x)-f(c)| = |e^x - e^c| = e^c|e^{x-c}-1| = e^c\left|\sum_{k=1}^{\infty}\frac{(x-c)^k}{k!}\right| \leq e^c\sum_{k=1}^{\infty}\frac{|x-c|^k}{k!}$$

$$< e^c\sum_{k=1}^{\infty}\frac{\delta^k}{k!} \leq e^c\sum_{k=1}^{\infty}\delta^k = \frac{e^c\delta}{1-\delta} \text{ (converges since } \delta \in (0,1))$$

$$= \frac{e^c\varepsilon}{e^c+\varepsilon} \div \left(1-\frac{\varepsilon}{e^c+\varepsilon}\right) = \frac{e^c\varepsilon}{e^c+\varepsilon} \div \frac{e^c}{e^c+\varepsilon} = \varepsilon$$

$$\text{i.e.,} \qquad |f(x)-f(c)| < \varepsilon$$

Hence e^x is continuous!

The δ that we choose is allowed to depend on ε and the point c. In the first 3 examples, the δ's we chose did not depend on c. In the last example, $\delta = \frac{\varepsilon}{e^c+\varepsilon}$ does depend on c.

We use this distinction to define the notion of **uniform continuity**.

The idea of uniform continuity is that there exists a value of δ which works for every value of c. In other wods, it is possible to choose a δ before being given c. Formally we define this by simply moving the $\forall_{c\in X}$, to after the $\exists_{\delta>0}$ in the definition (we choose δ before we choose c).

$$\forall_{\varepsilon>0}, \qquad \exists_{\delta>0}\text{s.t.}, \qquad \forall_{c\in X}, \qquad |x-c|<\delta \Rightarrow |f(x)-f(c)|<\varepsilon$$

You should NOT memorise this definition. Instead, if you want to show that something is uniformly continuous, you simply need to show that it is continuous, using a δ which does not depend on c. We have already shown that k, x, and $\sqrt{x}$ are **uniformly continuous**. Note that we have not demonstrated that e^x is not uniformly continuous. We found a δ which depends on c, but it might be the case that there is some clever choice of δ which does not. Proving that this is indeed impossible (and hence e^x is not uniformly continuous) is left as an exercise on the problem sheet.

e^x is not uniformly continuous. We know it is continuous though, by using $\delta=\frac{\varepsilon}{\varepsilon+e^c}$ (which depends on c). Notice that if instead we were to consider e^x on some closed interval $[a,b]$ then $a\le c\le b\Rightarrow e^c\le e^b\Rightarrow\frac{\varepsilon}{\varepsilon+e^c}\ge\frac{\varepsilon}{\varepsilon+e^b}$.

So $\delta=\frac{\varepsilon}{\varepsilon+e^b}\le\frac{\varepsilon}{\varepsilon+e^c}$ would work everywhere on this interval. This δ does not depend on c. It does depend on a, meaning the δ you choose depends on the interval, but it is independent of where in the interval you are.

We might then suspect that a continuous function on a bounded interval I is uniformly continuous. This is almost true but it turns out that we also require the interval to be closed.

Theorem 4.2:

If f is continuous on a closed interval, $[a,b]$

Then f is uniformly continuous on $[a,b]$

pf:

We proceed by contradiction. Let f be continuous on $[a,b]$. Suppose f is not uniformly continuous on $[a,b]$.

Using the negation of uniform continuity, f is **not** uniformly continuous iff:

$$\exists_{\varepsilon>0}\text{s.t.}, \qquad \forall_{\delta>0}, \qquad \exists_{x,c\in[a,b]}\text{s.t.}, \qquad |x-c|<\delta \text{ but } |f(x)-f(c)|\ge\varepsilon$$

We are assuming that this is true.

There exists an $\varepsilon > 0$, this ε is given. Then it says, for all $\delta > 0$ so we can choose any δ we like. Let's let $\delta_n = \frac{1}{n}$, then for each δ_n there exists some $x_n, c_n \in [a, b]$ such that $|x_n - c_n| < \delta_n = \frac{1}{n}$ and $|f(x_n) - f(c_n)| \geq \varepsilon$. Do this for each n to form sequences x_n and c_n.

Our goal is to show that $|f(x_n) - f(c_n)| \geq \varepsilon$ will lead to a contradiction (since we suspect that $f(x_n) - f(c_n)$ gets smaller as n gets larger, i.e., that $\lim\big(f(x_n) - f(c_n)\big) = 0$. Since f is continuous it would be easy to prove this if we knew that $\lim x_n = \lim c_n$. Unfortunately, we cannot conclude this, since we don't know that x_n and c_n are convergent. We do however know that they are bounded by $[a, b]$ so we use the Bolzano Weierstrass theorem:

$$\text{There is a convergent subsequence } x_{n_j} \text{ of } x_n$$

$$\text{Let } L = \lim x_{n_j}, \qquad \text{then } L \in [a, b]$$

$$\text{Then } \left|x_{n_j} - c_{n_j}\right| < \frac{1}{n_j} \Rightarrow \lim\left(c_{n_j} - x_{n_j}\right) = 0 \text{ (squeeze theorem)}$$

$$\text{Then } \lim\left(c_{n_j} - x_{n_j}\right) + \lim x_{n_j} = \lim c_{n_j} \ \left(\text{so} \lim c_{n_j} \text{ exists}\right)$$

$$\text{and} \lim c_{n_j} = L$$

Since f is continuous at L, $\lim f\left(x_{n_j}\right) = f\left(\lim x_{n_j}\right) = f(L)$ and $\lim f\left(c_{n_j}\right) = f(L)$.

$$\text{So } \lim\left(f\left(x_{n_j}\right) - f\left(c_{n_j}\right)\right) = \lim f\left(x_{n_j}\right) - \lim f\left(c_{n_j}\right) = f(L) - f(L) = 0$$

$\lim\left(f\left(x_{n_j}\right) - f\left(c_{n_j}\right)\right) = 0 \Rightarrow$there exists a j such that $\left|f\left(x_{n_j}\right) - f\left(c_{n_j}\right)\right| < \varepsilon$.

But by definition $\left|f\left(x_{n_j}\right) - f\left(c_{n_j}\right)\right| \geq \varepsilon$. This is a contradiction and so the assumption (that f is not uniformly continuous) was false. f is uniformly continuous.

And we are done!

The reason we needed that the interval was closed is since if we had an open interval (a, b) we would still have $L \in [a, b]$ (L could be still an endpoint) and the proof relies on continuity of f at L so we need f to be continuous at the endpoints, meaning the proof only works for closed intervals.

There is one final type of continuity which actually simplifies the idea of continuity considerably, (though it is a more strict condition).

When looking at the definition of uniform continuity, the tricky part is showing that $|x-c|<\delta \Rightarrow |f(x)-f(c)|<\varepsilon$.

If we are forunate enough to have a situation where we know that $|f(x)-f(c)| \leq M|x-c|$ for some constant $M>0$ and every x and c, then we may set $\delta = \frac{\varepsilon}{M}$ to get

$$|f(x)-f(c)| \leq M|x-c| < M\delta = \varepsilon$$

We use this idea to define a final type of continuity:

$$\text{Def } f(x) \text{ is } \textbf{Lipschitz Continuous} \text{ iff:}$$

$$\exists_{M>0} \text{ such that: } |f(x)-f(y)| \leq M|x-y|$$

$$\text{for every } x, y \text{ (in the domain of } f)$$

We call $M>0$ the $\textbf{Lipschitz Constant}$. Note that this constant is not necessarily unqiue (there may be more than one which works, in fact if M works, then anything $>M$ would also work).

I have already explained why $\textbf{Lipschitz Continuity} \Rightarrow \textbf{Uniform Continuity} \Rightarrow \textbf{Continuity}$. This definition is nice because it doesn't involve any ε's or δ's. This means if we are fortunate enough to have a Lipschitz continuous function, it is (sometimes) easier to prove than the other types of continuity, and then we get the other two types for free.

The **interpretation** of Lipschitz continuity is (roughly) to do with the slope of a function. $\frac{f(x)-f(y)}{x-y}$ represents the "slope" of a line connecting two point on the function. $\frac{|f(x)-f(y)|}{|x-y|} \leq M$ means that this slope is bounded above, and so the function doesn't become "infinitely steep" in some sense. For this reason we would not expect $\sqrt{x}$ to be Lipschitz continuous (the gradient seems to become vertical close to 1). We can easily prove that k and x are Lipschitz continuous functions.

$$|k-k| = 0 \leq 0|x-y| \text{ So Lipschitz with Lipschitz constant 0.}$$

$$|x-y| \leq 1|x-y| \text{ So Lipschitz with Lipschitz constant 1.}$$

Moving back to regular continuity, we can determine whether most function we deal with are continuous by combining results we already know.

$$\text{If } f(x) \text{ and } g(x) \text{ are continuous then}$$

$$f(x)+g(x) \text{ is continuous}$$

$$f(x)g(x) \text{ is continuous}$$

$$\frac{f(x)}{g(x)} \text{ is continuous (wherever } g(x) \neq 0)$$

These results follow immediately from the results for combining limits (and the original definition of continuity that $\lim_{x\to c} f(x) = f(c)$).

There is however a new way of combining functions which we have not yet considered. Function composition. If $f: B \to C$ and $g: A \to B$ are functions, then $f \circ g: A \to C$ is a function. The arrow notation here, if you need reminding means f takes an element in B as an input and returns an element in C as an output.

The " $\circ$ " notation means: $(f \circ g)(x) = f\big(g(x)\big)$, we call this f composed with g. This passes an element of A into g, getting an element of B out, which is then passed into f, which returns an element of C as the final output.

It would be nice if: when f and g are continuous, then so is $f \circ g$. That is,

$$\lim_{x\to c} f\big(g(x)\big) = f\big(g(c)\big)$$

If we could bring the limit inside of f, we could say that $\lim_{x\to c} f\big(g(x)\big) = f\left(\lim_{x\to c} g(x)\right)$, and then use continuity ($\lim_{x\to c} g(x) = g(c)$) to get $\lim_{x\to c} f\big(g(x)\big) = f\big(g(c)\big)$.

We need to justify the step of bringing the limit inside.

Theorem 4.3 (a):

If $L = \lim_{x\to c} g(x)$ exists and f is continuous at L

Then $\lim_{x\to c} f\big(g(x)\big)$ exists and equals $f\left(\lim_{x\to c} g(x)\right)$

We first ask why we would expect this to be true. "f is continuous at L" roughly means that when u is close to L, $f(u)$ is close to $f(L)$.

When x is close to c, $g(x)$ is close to L, meaning $f\big(g(x)\big)$ is close to $f(L)$.

pf:

$$\text{Let } L = \lim_{x\to c} g(x)$$

I aim to show that $\lim_{x\to c} f\big(g(x)\big) = f(L)$.

$$\text{Let } \varepsilon > 0$$

Using the continuity of f at L:

$$\exists_{\delta_1>0} \text{ such that } 0 < |u - L| < \delta_1 \Rightarrow |f(u) - f(L)| < \varepsilon$$

Then using the limit of $g(x)$,

$$\text{for all } \tilde{\varepsilon} > 0, \qquad \exists_{\delta>0} \text{ such that,} \qquad 0 < |x - c| < \delta \Rightarrow |g(x) - L| < \tilde{\varepsilon}$$

Then set $\tilde{\varepsilon} = \delta_1$ and choose this new $\delta > 0$ as our δ.

$$\text{Suppose } 0 < |x - c| < \delta$$

$$\Rightarrow |g(x) - L| < \delta_1$$

$$\text{Assuming } g(x) \neq L, \text{we have } 0 < |g(x) - L| < \delta_1$$

$$\Rightarrow \big|f\big(g(x)\big) - f(L)\big| < \varepsilon$$

If $g(x) = L$ then $\big|f\big(g(x)\big) - f(L)\big| = |f(L) - f(L)| = 0 < \varepsilon$.

Hence we have found $\delta > 0$ such that $0 < |x - c| < \delta \Rightarrow \big|f\big(g(x)\big) - f(L)\big| < \varepsilon$.

i.e., $\lim_{x\to c} f\big(g(x)\big) = f(L) = f\left(\lim_{x\to c} g(x)\right)$ (so long as f is continuous at L).

This theorem fully justifies the proof:

$$\text{Let } f \text{ be continuous at } g(c), \qquad g \text{ be continuous at } c$$

$$\text{Then } \lim_{x\to c} g(x) = g(c) \text{ and } f \text{ continuous at } g(c)$$

$$\Rightarrow \lim_{x\to c} f\big(g(x)\big) = f\left(\lim_{x\to c} g(x)\right) \ \big(\text{using theorem 4.2(a)}\big)$$

$$= f\big(g(c)\big)$$

$$\text{Hence } f \circ g \text{ is continuous at } c$$

The reason I call this 4.2 (a) is because this only covers limits with $x \to c \in \mathbb{R}$, but we may also wish to take limits with $x \to \infty$ or $x \to -\infty$.

Theorem 4.3 (b):

$$\text{If } L = \lim_{x\to\infty} g(x) \text{ exists and } f \text{ is continuous at } L.$$

$$\text{Then } \lim_{x\to\infty} f\big(g(x)\big) \text{ exists and equals } f\left(\lim_{x\to\infty} g(x)\right)$$

The proof is similar to (a)

pf:

$$\text{Let } L = \lim_{x\to\infty} g(x)$$

$$\text{Let } \varepsilon > 0$$

Continuity of f at L:

$$\exists_{\delta>0} \text{ such that } 0 < |u - L| < \delta \Rightarrow |f(u) - f(L)| < \varepsilon$$

Then using the limit of $g(x)$,

$$\text{for all } \tilde{\varepsilon} > 0, \qquad \exists_M \text{ such that,} \qquad x \geq M \Rightarrow |g(x) - L| < \tilde{\varepsilon}$$

Then set $\tilde{\varepsilon} = \delta$, which gives us a value for M.

$$\text{Suppose } x \geq M$$

$$\Rightarrow |g(x) - L| < \delta$$

$$\text{Assuming } g(x) \neq L, \text{we have } 0 < |g(x) - L| < \delta$$

$$\Rightarrow \left|f\big(g(x)\big) - f(L)\right| < \varepsilon$$

If $g(x) = L$ then $\left|f\big(g(x)\big) - f(L)\right| = |f(L) - f(L)| = 0 < \varepsilon$.

Given any $\varepsilon > 0$, we have found an M such that $x \geq M \Rightarrow \left|f\big(g(x)\big) - f(L)\right| < \varepsilon$.

Meaning, $\lim_{x\to\infty} f\big(g(x)\big) = f(L) = f\left(\lim_{x\to\infty} g(x)\right)$.

Limits by Substitution

One often overlooked theorem in real analysis is that of performing substitutions on limits.

If I have to evaluate $L = \lim_{x \to c} f\big(g(x)\big)$ one may wish to write "Let $u = g(x)$, then $L = \lim_{u \to g(c)} f(u)$ ". We can only do this though, if certain conditions are met, most importantly that g is continuous at c. There are also some weird extra conditions the need for which will become apparent during the proof.

$$\text{Theorem 4.4: (limits by substitution)}$$

$$\lim_{x \to c} f\big(g(x)\big) = \lim_{u \to g(c)} f(u) \text{ if } g \text{ is continuous at } c$$

$$\text{and either } [f \text{ is continuous at } g(c)]$$

$$\text{or } [g(x) \neq g(c) \text{ for all } x \text{ sufficiently close to } c]$$

These two extra conditions will used to deal with an annoying edge-case.

Substitution is often done without justification. This section will consist of a proof of this theorem. The idea (roughly) is that if x is close to c then $g(x)$ is close to $g(c)$ (by continuity of g at c), and so (letting $u = g(x)$), $f(u)$ is close to $\lim_{u \to g(c)} f(u)$.

This will be similar to the proof at the end of the previous section (that $\lim f\big(g(x)\big) = f(\lim g(x))$ when f is continuous).

pf:

$$\text{Let } L = \lim_{u \to g(c)} f(u)$$

$$\text{Let } \varepsilon > 0$$

$$\text{There exists a } \delta_1 > 0 \text{ such that } 0 < |u - g(c)| < \delta_1 \Rightarrow |f(u) - L| < \varepsilon$$

Then using continuity of g at c:

$$\text{There exists a } \delta > 0 \text{ such that } |x - c| < \delta \Rightarrow |g(x) - g(c)| < \delta_1$$

$$\text{Suppose } |x - c| < \delta,$$

$$\text{Then } |g(x) - g(c)| < \delta_1$$

If $g(x) \neq g(c)$ then:

$$0 < |g(x) - g(c)| < \delta_1$$

$$\Rightarrow \big|f\big(g(x)\big) - L\big| < \varepsilon$$

If $g(x) = g(c)$, we deal with this edge case by using the one of the extra conditions.

If f is continuous at $g(c)$ then $L = \lim_{u \to g(c)} f(u) = f\big(g(c)\big)$.

Then $|f(g(x) - L| = |f(g(c) - L| \ = |L - L| = 0 < \varepsilon$ (since $g(x) = g(c)$ and $f\big(g(c)\big) = L$).

If we have the second condition, $g(x) \neq g(c)$ for all x sufficiently close to c, then we can just choose δ such that x is close enough to c so that $g(x) \neq g(c)$, avoiding the edge case.

In all cases we have shown that given $\varepsilon > 0$, we can choose $\delta > 0$ so that $|x - c| < \delta \Rightarrow \left|f\big(g(x)\big) - L\right| < \varepsilon$ (where $L = \lim_{u \to g(c)} f(u)$)

$$\text{i.e.,} \lim_{x \to c} f\big(g(x)\big) = L = \lim_{u \to g(c)} f(u)$$

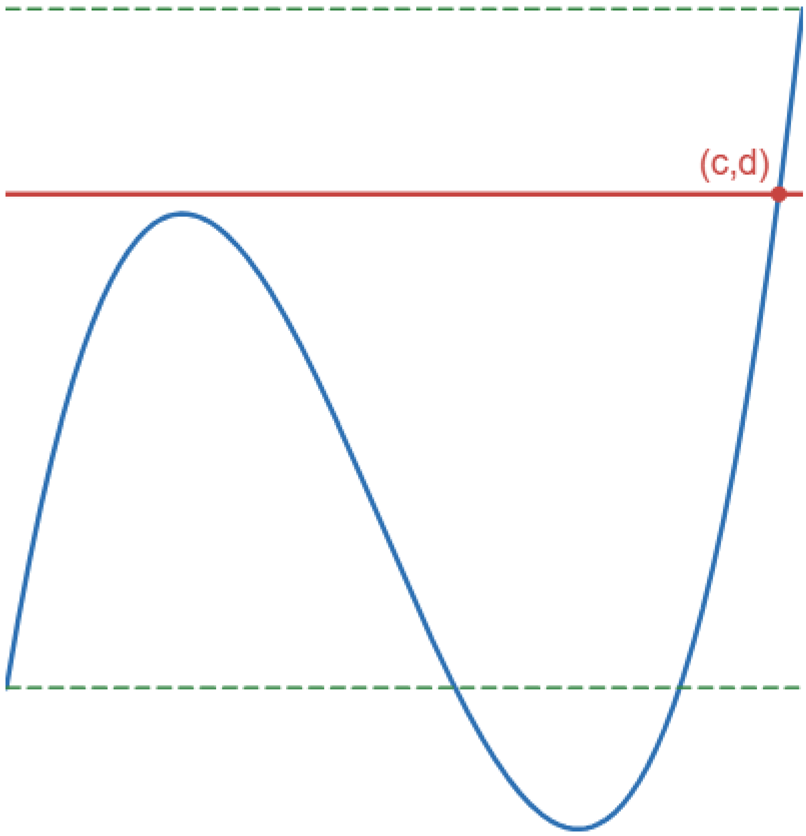
We will be using substitutions later and so it is a good idea to justify their use.

Intermediate and Extreme Value Theorems

This section covers two important (yet very intuitive) theorems.

The **Intermediate Value Theorem** (IVT) says that if a function f is continuous on $[a, b]$ then for any point d between $f(a)$ and $f(b)$, there is a point c on $[a, b]$ such that $f(c) = d$.

Intuitively, the graph of any continuous function must pass through any horizontal line which is vertically between the start ane end points.



Intermediate Value Theorem (IVT) (4.5):

If f is continuous on $[a, b]$ and d is between $f(a)$ and $f(b)$ (inclusive)

Then there exists a $c \in [a, b]$ such that $f(c) = d$

Note: if $f(a) \leq f(b)$ then $d \in [f(a), f(b)]$, if $f(a) \geq f(b)$ then $d \in [f(b), f(a)]$.

We, will need some kind of "candidate" for our value of c. The proof will clearly rely on our function being continuous, so we will be using the definition of continuity. If our function was defined only for rational numbers, there could be "gaps", e.g., take $f(x) = x^2$ on $[0,2]$, then $f(a) = 0$, $f(b) = 4$. Clearly $d = 2 \in [0,4]$ would be a fine choice for d, but then there is no rational number $c \in [0,2]$ such that $f(c) = c^2 = 2$. This suggests that the the "lack of gaps" in the real numbers is essential to this theorem. For this reason, we will be invoking the axiom of completeness. Recall, the AoC says that any non-empty set of real numbers which is bonded above has a supremum (least upper bound).

We will consider the set, $S = \{x \in [a, b],\ f(x) < d\}$

This is the set of all values in the interval such that the graph is below the horizontal line. For the graph shown on the previous page, this the the set of all points with the blue beneath the red (in this particular graph this happens to be all points left of the intersection). I looks like the supremum (call it c) is the x co-ordinate where the curve (blue) meets the horizontal line (red), that is $f(c) = d$). If the curve were to intersect the line multiple times, the supremum would seem to be the x value where it happens last.

Our plan will be to define c as the supremum of this set S and then show that $f(c) = d$.

pf:

$$\text{Case 1: } f(a) = f(b)$$

The only value of d allowed is $d = f(a) = f(b)$ and so $c = a$ or $c = b$ would each work.

$$\text{Case 2: } f(a) < f(b)$$

If $d = f(a)$ then $c = a$ works, if $d = f(b)$ then $c = b$ works. Otherwise $d \in \big(f(a), f(b)\big)$ and $c \in (a, b)$.

$$\text{Let } S = \{x \in [a, b], \qquad f(x) < d\}$$

$f(a) \in S$ so non empty.

$x \in S \Rightarrow x \in [a, b] \Rightarrow x \leq b$ so S is bounded above.

We can then apply the axiom of completeness, the supremum exists.

$$\text{Let } c = \sup S$$

We now have a candidate for $f(c) = d$. To prove this we will use contradiction.

$$\text{Suppose } f(c) < d$$

This is similar to the proof that $\sqrt{2}$ is a real number. One case will contradict c being an upper bound, the other case will contradict no upper bounds being less than c. Which case is this?

Consulting the diagram, this should lead to problems because if $f(c) < d$ (the blue function is below the red line) then c is too far left (too small). We seek to show that c is too small to be an upper bound. We want to find some $x \in S$ with $c < x$. For $x \in S$ we require $x \in [a, b]$ and $f(x) < d$.

Recall the definition of continuity, in particular continuity at our point c:

$$\forall_{\varepsilon>0}, \qquad \exists_{\delta>0} \text{ such that,} \qquad |x-c|<\delta \Rightarrow |f(x)-f(c)|<\varepsilon$$

This is the definition for continuity at an interior point (non-endpoint) which is fine since $c \in (a,b)$.

Taking inspiration from our "$\sqrt{2} \in \mathbb{R}$" proof (from quite a long time ago), we notice that $d-f(c)>0$ (since we are assuming that $f(c)<d$) and we

$$\text{Let } \varepsilon = d - f(c) > 0$$

$$\text{Then there is a } \delta > 0 \text{ such that}$$

$$|x-c|<\delta \Rightarrow |f(x)-f(c)|<\varepsilon\big(=d-f(c)\big)$$

We take this δ as given.

We wish to find an $x>c$ with $a \le x \le b$ and $f(x)<d$ (that is, find an $x>c$ with $x \in S$).

If we find such an x, we have shown that c is not an upper bound and we will have our contradiction.

$$\text{Let } x = c + \frac{\delta}{2}$$

Certainly $x>c$.

Also $|x-c| = \frac{\delta}{2} < \delta$ and so $|f(x)-f(c)|<\varepsilon$.

This also implies that $f(x)$ exists meaning $x \in [a,b]$ (limits require the function to be defined everywhere close to c).

Since we chose $\varepsilon = d - f(c)$, we have $|f(x)-f(c)| < d - f(c)$, which is equivalent to:

$$f(c)-d < f(x)-f(c) < d-f(c)$$

$$\Rightarrow f(x)-f(c) < d - f(c)$$

$$\Rightarrow f(x) < d$$

Which means $x \in S$ so we have a contradiction (as c is not an upper bound).

To complete the proof we assume $f(c)>d$ and reach a contradiction. The proof is similar, but a bit tricker. We need to find an upper bound which is less than c. We again call it x.

Let $\varepsilon = f(c)-d \Rightarrow \exists_{\delta>0}$ such that $|x-c|<\delta \Rightarrow |f(x)-f(c)| < f(c)-d$.

$$\text{Let } x = c - \delta$$

Clearly $x < c$. We want to show that x is an upper bound.

Let $\tilde{x} > x$, we wish to show that $\tilde{x} \notin S$, hence x is an upper bound.

Either $\tilde{x} \in (x, c]$ or $\tilde{x} > c$.

If $\tilde{x} > c \Rightarrow \tilde{x} \notin S$ (since c is an upper bound for S).

If $\tilde{x} \in (x, c] \Rightarrow \tilde{x} \in (c - \delta, c] \Rightarrow \tilde{x} \in (c - \delta, c + \delta) \Rightarrow |\tilde{x} - c| < \delta \Rightarrow |f(\tilde{x}) - f(c)| < \varepsilon$

$= f(c) - d$

Then $d - f(c) < f(\tilde{x}) - f(c) < f(c) - d \Rightarrow d - f(c) < f(\tilde{x}) - f(c) \Rightarrow f(\tilde{x}) > d$

$\Rightarrow \tilde{x} \notin S$ (by definition of S).

Hence x is an upper bound for S and $x < c$. This contradicts $c = \sup S$.

We have now contradicted $f(c) < d$ and $f(c) > d$, and we have victory.

$$f(c) = d \text{ where } c \in [a, b] \text{ as required.}$$

Case 3: $f(b) < f(a)$

This case is symmetric to case 2. You could either go through all the working again (which isn't a bad exercise if you want to check that you understood this proof). We can also use a shortcut.

$$\text{Let } g(x) = -f(x)$$

$f(b) < d < f(a) \Rightarrow -f(a) < -d < -f(b) \Rightarrow g(a) < -d < g(b)$

Using case 2: there exists a $c \in [a, b]$ with $g(c) = -d$. Then $f(c) = -g(c) = d$ and we are done.

That was a lot of work for something which seems obvious, but there are occasionally seemingly obvious things in mathematics which turn out to be false. How do we know that there isn't some very clever edge case where the statement doesn't hold? Just because we can't think of one, doesn't mean it doesn't exist. The only way to be certain is with proof, which we have now provided.

This proof seemed quite similar to the proof that $\sqrt{2}$ is real. This is because IVT is essentially just a generalisation of this fact. Setting $f(x) = x^2$, $a = 0$, $b = 2$, $d = 2$, we get that $c^2 = 2$ and $c \in [0,2]$, meaning $c = \sqrt{2}$ is real.

An immediate consequence of IVT, is that continuous functions map intervals onto intervals. Put precisely, if f is continuous and I is an interval then $J \coloneqq f(I)$ is an interval. This is because if $a, b \in J$ with $a < b$ and $c \in (a, b)$; then $a = f(x_1)$ and $b = f(x_2)$ for some $x_1, x_2 \in I$ meaning (by IVT) there is some $x_3 \in I$ with $f(x_3) = c$, hence $c \in J$.

The final theorem we will prove in this chapter is the **EXTREME Value Theorem** (EVT). Make sure you say the word EXTREME in the most extreme way you can. This theorem says that if a function is continuous on a closed interval (similar to IVT), then the function has a minimum and maximum value (meaning the function is bounded and the "extreme" values are attained).

EXTREME Value Theorem (EVT) (4.6):

If f is continuous on $[a, b]$

then there is some upper bound M, with $M \geq f(x)$ for every x, and $M = f(C)$ for some C

and there is some lower bound m, with $m \leq f(x)$ for every x, and $m = f(c)$ for some c

M clearly seems to be a supremum, so the idea will be to use the axiom of completeness (AoC) again and go from there. We will also do some magic with sequences as that will make our lives a little bit easier.

pf:

$$\text{Let } S = \{f(x), x \in [a, b]\}$$

$f(a) \in S$ so non-empty.

Showing that S is bounded above is slightly harder. We will use contradiction.

Suppose S is not bounded above.

Then 1 is not an upper bound meaning there exists an x with $x > 1$. Call this x_1. 2 is also not an upper bound, so there is some $x_2 > 2$. $n \in \mathbb{N}$ is not an upper bound, so there is some $x_n > n$. We use this to form a sequence x_n.

Next, we know that $x_n \in [a, b]$ and so is bounded. We use Bolzano-Weierstrass to find a subsequence x_{n_j} which is convergent. Call the limit $c \in [a, b]$. We may think of x_{n_j} as being a function of j with domain $\mathbb{N}$. We then can use the continuity of f at c with theorem 4.3 (b) to say that:

$$\lim_{j \to \infty} x_{n_j} \text{ exists} \Rightarrow \lim_{j \to \infty} f\left(x_{n_j}\right) \text{ exists}$$

Since convergent sequences are bounded, $f\left(x_{n_j}\right)$ is bounded.

But we constructed this sequence so that $f\left(x_{n_j}\right) > n_j$ which is unbounded above. This is a contradiction meaning S is bounded above.

We then let $M = \sup S$, then $M \geq f(x)$ for every $x \in [a, b]$. We need to find a C so that $f(C) = M$. We will again do this by using sequences.

By Theorem 2.8, there exists a sequence $y_n \in S$ with $\lim y_n = M$.

$y_n \in S \Rightarrow y_n = f(x_n)$ for some $x_n \in [a, b]$. This gives us some (not necessarily unique, there may be more than one choice) sequence x_n.

We know that $\lim y_n = \lim f(x_n) = M$. If we knew that x_n were convergent we could move the limit inside the function (since f is continuous). x_n is not necessarily convergent, but it is bounded ($x_n \in [a, b]$). The means (by B-W) there is a subsequence x_{n_j} which is convergent.

$$\text{Let } C = \lim x_{n_j} \in [a, b]$$

We now have that $f\left(x_{n_j}\right) = y_{n_j}$ is a subsequence of y_n. We know that $\lim y_n = M$ and so $\lim y_{n_j} = \lim f\left(x_{n_j}\right) = M$.

Since we do know that x_{n_j} converges, we can now bring the limit inside f (since f is continuous).

$$M = \lim f\left(x_{n_j}\right) = f\left(\lim x_{n_j}\right) = f(C), \qquad \text{where } C \in [a, b]$$

We have $f(C) = M$ with $C \in [a, b]$ and so we are done. The upper bound M is attained so we call it a maximum.

For the minimum you could either do the entire thing again with a symmetric argument (not a bad idea to test your understanding) but we can also use a shortcut.

$$\text{Let } g(x) = -f(x)$$

Using what we have proved so far, find c suc that $g(c) = -m$. (where $-m$ is an upper bound for g).

$$-m \geq g(x)\ \forall_x \Rightarrow m \leq f(x)\ \forall_x$$

This means m is a lower bound for f.

Then $-f(c) = g(c) = -m \Rightarrow f(c) = m$. The lower bound m is attainted so we call it a minimum.

Inverse of Continuous Functions

We have seen that if f and g are continuous functions, then so are $f + g, fg, f \circ g$. If additionally, $g \neq 0$, we have $\frac{f}{g}$ is continuous. There is one more result which I omitted because we will need to use the intermediate value theorem (which we did not have at the time) to prove it. If f is bijective (meaning it has an inverse) and continuous, then its inverse is also continuous.

To prove this we first need to prove a lemma.

Since f is cotinuous you can draw its graph without taking your pen off the page. Since it's bijective it can never go back on itself (since then you would have multiple inputs giving the same output). This means f must be monotonic. We will now prove this more rigorously:

Lemma 4.7:

If f is continuous and injective on an interval I

Then f is strictly monotone.

The standard proof for this theorem makes some assumptions which I don't think are usually fully justified. I have modified the proof here as to fully justify these assumptions.

pf:

The trouble with this proof is that monotone might mean increasing or decreasing, and we don't know which it will be. We will assume (for now) that $I = [a, b]$ is a closed interval with $a < b$. To determine whether we should be aiming to show increasing or decreasing:

$$f(a) \neq f(b) \text{ (injectivity)}$$

$$\text{So either } f(a) < f(b) \text{ or } f(b) < f(a)$$

$$\text{Suppose } f(a) < f(b)$$

In this case we want to show that f is monotone increasing. That is, given $x_1 < x_2$ I need to show that $f(x_1) < f(x_2)$. We first show that given $x_1 \in (a, b)$ we must have $f(a) < f(x_1) < f(b)$.

$$\text{Suppose } a < x_1 < b$$

Injectivity tells us that $f(x_1) \neq f(a)$ and $f(x_1) \neq f(b)$.

Suppose $f(x_1) < f(a)$ and aim for contradiction.

Since f is continuous on $[x_1, b]$ and $f(x_1) < f(a) < f(b)$, by the intermediate value theroem, there exists some $x_2 \in [x_1, b]$ such that $f(x_2) = f(a)$. But $a < x_1 \leq x_2 \Rightarrow x_2 \neq a \Rightarrow f(x_2) \neq f(a)$ (injectivity). This is a contradiction (the assumption being that $f(x_1) < f(a)$) hence $f(a) < f(x_1)$.

Similarly, if we suppose $f(x_1) > f(b)$ then $a < x_1 < b$ and $f(a) < f(b) < f(x_1)$. Use IVT to find x_2 such that:

$$a \leq x_2 \leq x_1 \text{ and } f(x_2) = f(b)$$

But $x_2 \leq x_1 < b \Rightarrow x_2 \neq b \Rightarrow f(x_2) \neq f(b)$ (contradiction).

Hence $f(a) < f(x_1) < f(b)$

We have shown that if f is continuous and injective on $[a, b]$ with $a < b$ and $f(a) < f(b)$, then $a < x_1 < b \Rightarrow f(a) < f(x_1) < f(b)$. I call this result $(*)$.

To show that f is increasing, let $x_1 < x_2$ with $x_1, x_2 \in [a, b]$, we need to show that $f(x_1) < f(x_2)$.

We have $a \leq x_1 < x_2 \leq b$

We know f is continuous and injective on $[x_1, b]$ and $f(x_1) < f(b)$. Also $x_2 \in (x_1, b]$. Hence (by $(*)$) we have $f(x_1) < f(x_2) \leq f(b)$. Importantly, $f(x_1) < f(x_2)$. Hence f is increasing in this case.

The proof that $f(a) > f(b) \Rightarrow f$ is strictly decreasing, is a symmetric argument to what we have done here. In either case, we have shown that f is strictly monotone.

I call what we have proved so far $(**)$

We did assume that I was a closed interval (which it might not be). If it is not, we do the following:

$$\text{Let } a < b \text{ (for some } a, b \in I) \text{ these are fixed}$$

$$\text{In the case that } f(a) < f(b) \text{ we want to show that } f \text{ is increasing.}$$

$$\text{Then consider } [a, b]. \text{Since } f(a) < f(b), \text{by } (**) \ f \text{ is increasing on } [a, b]$$

$$\text{We still need to show that } f \text{ is increasing everywhere else in } I$$

$$\text{Suppose } u \leq a < b \leq v \text{ (for some } u, v \in I). \big[\text{This is so that } [u, v] \text{ contains } [a, b]\big]$$

$$\text{Then by } (**) \ [u, v] \text{ is strictly monotone.}$$

Since $a < b$ and $f(a) < f(b)$, and $a, b \in [u, v]$, f cannot be strictly decreasing on $[u, v]$

so it must be strictly increasing on $[u, v]$.

(This is true for any u, v with $u \leq a < b \leq v$)

Finally, let $x_1, x_2 \in I$ with $x_1 < x_2$.

If we can show that $f(x_1) < f(x_2)$ then we have shown that f is increasing on I.

Let $u = \min\{x_1, a\}, \qquad v = \max\{x_2, b\}$

This choice of u and v ensures that that $u \leq a < b \leq v$ (so that f is strictly increasing on $[u, v]$) and ensures that $x_1, x_2 \in [u, v]$.

$x_1, x_2 \in [u, v]$, f is increasing on $[u, v]$, and $x_1 < x_2$, so $f(x_1) < f(x_2)$.

Hence $x_1 < x_2 \Rightarrow f(x_1) < f(x_2)$ meaning f is strictly increasing on I.

The case for $f(b) < f(a)$ is almost identical, but we end up with $x_1 < x_2 \Rightarrow f(x_2) < f(x_1)$ and we have f is strictly decreasing on I in this case.

A very small Lemma next.

Lemma 4.8

If f is strictly increasing Then f^{-1} is strictly increasing

pf:

Suppose $f(a) < f(b)$.

$a = b \Rightarrow f(a) = f(b), \qquad a > b \Rightarrow f(a) > f(b)$ (both contradictions)

$a < b \Rightarrow f(a) < f(b), \qquad$ (not a contradiction)

Meaning, $f(a) < f(b) \Rightarrow a < b$

Let $u = f(a), \qquad v = f(b)$

Then $u < v \Rightarrow f^{-1}(u) < f^{-1}(v)$

Hence f^{-1} is strictly increasing

Similarly if f is strictly decreasing, then f^{-1} is strictly decreasing.

We will now use this Lemma to prove the main theorem we were trying to prove.

Theorem 4.9:

If f, defined on some interval I is bijective and is continuous at c

Then f^{-1} is continuous at $f(c)$

pf:

Let $\varepsilon > 0$

Choose $\delta \cdots$

We come back later to choose a $\delta > 0$ which will work.

Suppose $|y - f(c)| < \delta$

We know that $y \in (f(c) - \delta, f(c) + \delta)$.

We wish to show that $|f^{-1}(y) - c| < \varepsilon$ and we will have proved that f^{-1} is continuous at $f(c)$.

Since f is bijective, it is injective. Also f is continuous. This means, by Lemma 4.7, that f is strictly monotone.

Case 1: f is strictly increasing

Then f^{-1} is also strictly increasing (Lemma 4.8). Assume that ε is small enough that $(c - \varepsilon, c + \varepsilon) \subset I$ (recall $\subset$ means "is a subset of"). If this assumption is false, the δ we find under this assumption would still work for any larger ε. We make this assumption to make sure that this entire interval is within the domain of f.

Since f is monotone increasing, the image of this interval under f is

$$A = \big(f(c - \varepsilon), f(c + \varepsilon)\big)$$

Also (since f is monotone increasing) $f(c - \varepsilon) < f(c) < f(c + \varepsilon)$ which means $f(c) \in A$.

We now choose our $\delta > 0$ such that $(f(c) - \delta, f(c) + \delta) \subset A$.

[$\delta = \min\{f(c + \varepsilon) - f(c),\ f(c) - f(c - \varepsilon)\}$ works for this.]

Since $y \in (f(c) - \delta, f(c) + \delta)$, we now also know that $y \in A$.

Using the fact that f^{-1} is increasing, the pre-image of this interval is $(c - \varepsilon, c + \varepsilon)$. By definition of the pre-image, $f^{-1}(y) \in (c - \varepsilon, c + \varepsilon)$

meaning $|f^{-1}(y) - c| < \varepsilon$

Hence $f^{-1}(y)$ is continuous at $f(c)$ as required.

Some immediate applications of this theorem are that functions like $\sqrt{x}$ (inverse of $x^2, x \geq 0$) and $\log x$ (inverse of e^x) are continuous.

We also then get for free that x^α, (at least for $x > 0$) is continuous for any $\alpha \in \mathbb{R}$, since

$$x^\alpha = e^{\alpha \log x}$$

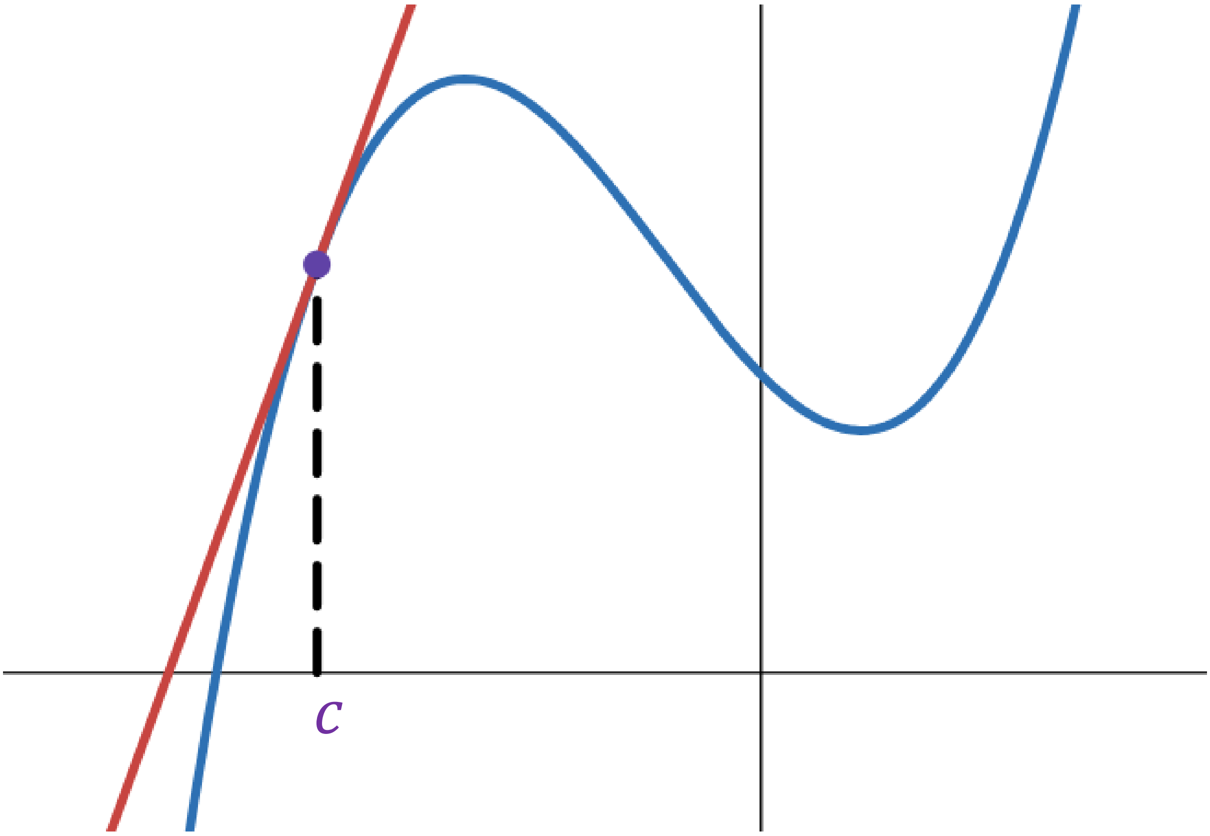
And we now know that e^x and $\log x$ are continuous.

Differentiation and Differentiability

Definition of Differentiation

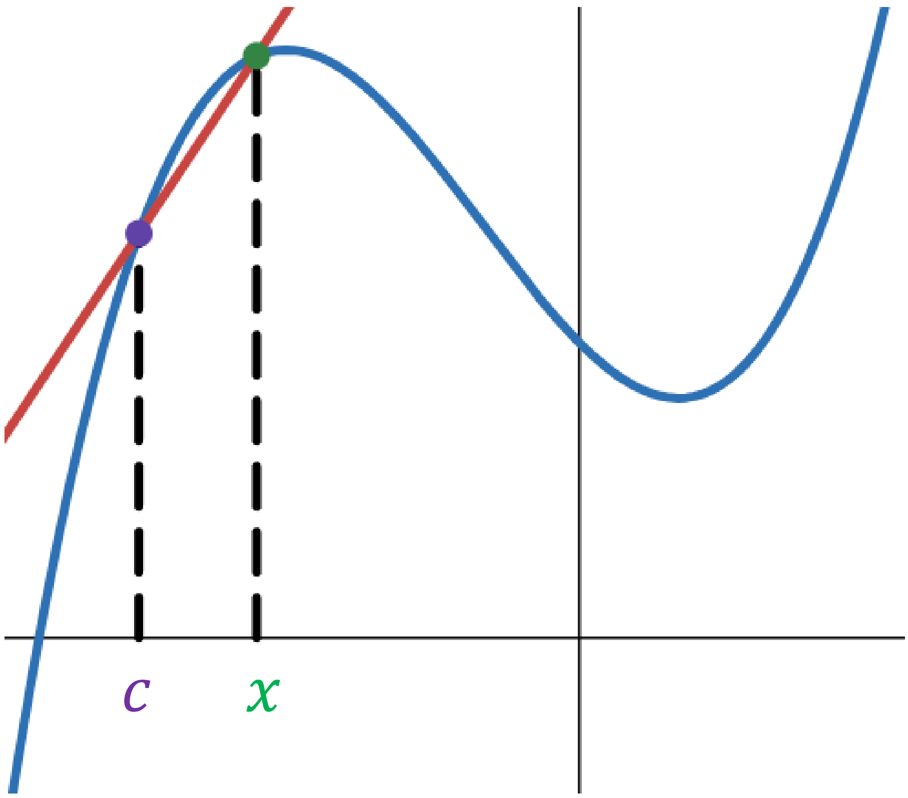
Differentiation is a key concept from calculus. We have now developed the tools necessary to make it rigorous.

Roughly speaking, the derivative of a function f at some point x, tells you how quickly the value of the function f will change as you increase the input value, when the input value is close to x. One way of thinking about the derivative is as the gradient (slope) of the tangent line to the function at that point.



In this example, the derivative of the function at the point c is the gradient of the tangent line.

We can approximate the tangent line by drawing a line through $(c, f(c))$ and through a point near c, let's say $(x, f(x))$ where x is some number close to c:



Which might look something like this. Of course, the closer the points are together, the better this approximation will be.

Since we know that this line passes through $(c, f(c))$ and $(x, f(x))$, it's gradient (change in f divided by change in x) will be

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

Since this is only an approximation (which gets better when x gets closer to c) it makes sense to define the derivative of f at c (which we write as $f'(c)$) by:

$$f'(c) \coloneqq \lim_{x \to c} \frac{f(x) - f(c)}{x - c}$$

We say that a function is differentiable at c, if this limit exists. If it is differentiable at each point in it's domain, we simply call it differentiable.

An alternative definition which we will sometimes use can be obtained by using the substitution, $h = x - c$. (You might want to check back to the "limits by substiution" section to verify that this substitution is allowed. $g(x) = x - c$ is continuous at $x = c$. Also $g(x)$ is injective, meaning $g(x) \neq g(c)$ whenever $x \neq c$. The substitution is allowed.)

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

We may think of the derivative as itself being a function. To remind ourselves of this we often write:

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

We sometimes write $\frac{\mathrm{d}}{\mathrm{d}x}$ to denote "the derivate of [] with respect to x", i.e., $\frac{\mathrm{d}}{\mathrm{d}x} f(x) = f'(x)$.

$$\text{E.g.}, f(x) = k \text{ where } k \text{ is some constant}$$

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

$$\text{E.g.}, f(x) = x \Rightarrow f'(x) = \lim_{h \to 0} \frac{x+h-x}{h} = \lim_{h \to 0} \frac{h}{h} = \lim_{h \to 0} 1 = 1$$

$$\text{E.g.}, f(x) = x^2 \Rightarrow f'(x) = \lim_{h \to 0} \frac{(x+h)^2 - x^2}{h} = \lim_{h \to 0} \frac{x^2 + 2xh + h^2 - x^2}{h} = \lim_{h \to 0} \frac{2xh + h^2}{h}$$

$$= \lim_{h \to 0} (2x + h) = 2x$$

One of the first results you learn in a calculus course is that $f(x) = x^n \Rightarrow f'(x) = nx^{n-1}$ (for $n \in \mathbb{W}$). This can be easily proved by the binomial theorem and the definition of the derivative (problem sheet).

It is easy to conceive of a function which is continuous but not differentiable, e.g., $f(x) = |x|$ is continuous everywhere, but it is not differentiable at $x = 0$ since,

$$f'(0) = \lim_{h \to 0} \frac{|h| - |0|}{h} = \lim_{h \to 0} \frac{|h|}{h}$$

The right limit is $\lim_{h \to 0^+} \frac{|h|}{h} = \lim_{h \to 0^+} \frac{h}{h} = 1$

The left limit is $\lim_{h \to 0^-} \frac{|h|}{h} = \lim_{h \to 0^-} \frac{-h}{h} = -1 \neq \lim_{h \to 0^+} \frac{|h|}{h}$

Hence the limit does not exist at zero. Continuity $\not\Rightarrow$ differentiability.

We are however fortunate enough that differentiability $\Rightarrow$ continuity. This can be quickly proved by using a clever, but not at all complicated, trick.

$$\text{Theorem 5.1:}$$

$$\text{If } f \text{ is differentiable at } c \text{ then } f \text{ is continuous at } c$$

pf:

$$\lim_{x\to c}[f(x)-f(c)] = \lim_{x\to c}\left[\frac{f(x)-f(c)}{x-c}(x-c)\right] = \lim_{x\to c}\frac{f(x)-f(c)}{x-c}\cdot\lim_{x\to c}(x-c)$$

$$= f'(c)\cdot 0 = 0$$

$$\text{But also } \lim_{x\to c}[f(x)-f(c)] = \lim_{x\to c} f(x) - \lim_{x\to c} f(c) = \lim_{x\to c} f(x) - f(c)$$

$$\text{So } \lim_{x\to c} f(x) - f(c) = 0$$

$$\Rightarrow \lim_{x\to c} f(x) = f(c)$$

Hence we have continuity.

I will show you an alternative way of defining differentiability, which gives us an explicit form for f, which makes many proofs much easier.

The intuitive idea first. When x is close to c,

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

$\approx$ means "approximately equal to". We will call the error in this approximation the remainder, denoted by $r(x)$. Then

$$\frac{f(x)-f(c)}{x-c} = f'(c) + r(x)$$

$$\Rightarrow f(x) = f(c) + (x-c)f'(c) + (x-c)r(x)$$

So $f(x)$ is differentiable if it can be written in this form.

Also, as $x \to c$, we expect the remainder to $\to 0$, meaning $\lim_{x\to c} r(x) = 0$.

We want to show that this new definition is equivalent to the old one.

Theorem 5.2:

$f(x)$ is differentiable at c iff it can be written in the form:

$$f(x) = f(c) + f'(c)(x-c) + r(x)(x-c)$$

where $r(x)$ is some function with $\lim_{x\to c} r(x) = 0$

This is an iff (if an only if) statement. We need to prove both directions.

pf:

" $\Rightarrow$ "

$$\text{Suppose } f \text{ is differentiable at } c$$

$$\text{This means } \lim_{x\to c} \frac{f(x)-f(c)}{x-c} = f'(c)$$

$$\text{Then we define } r(x) \coloneqq \frac{f(x)-f(c)}{x-c} - f'(c), \qquad x \neq c$$

It doesn't matter what $r(c)$ is because we only care about the limit as $x \to c$. $r(c)$ can be anything as long as it is defined.

We need to show that $f(x) = f(c) + f'(c)(x-c) + r(x)(x-c)$ (for all x).

$$\text{For } x \neq c \text{ we rearrange:}$$

$$r(x) = \frac{f(x)-f(c)}{x-c} - f'(c) \Rightarrow f(x) = f(c) + f'(c)(x-c) + r(x)(x-c)$$

$$\text{For } x = c:$$

$$\text{LHS} = f(x) = f(c)$$

$$\text{RHS} = f(c) + f'(c)(x-c) + r(x)(x-c) = f(c) + f'(c)\cdot 0 + r(c)\cdot 0 = f(c) = \text{LHS}$$

So we always have $f(x) = f(c) + f'(c)(x-c) + r(x)(x-c)$.

We finally need to show that $\lim_{x\to c} r(x) = 0$.

$$\lim_{x\to c} r(x) = \lim_{x\to c}\left(\frac{f(x)-f(c)}{x-c} - f'(c)\right) = \lim_{x\to c}\frac{f(x)-f(c)}{x-c} - f'(c) = f'(c) - f'(c) = 0$$

" $\Leftarrow$ "

Suppose $f(x) = f(c) + M(x-c) + r(x)(x-c)$ where $\lim_{x\to c} r(x) = 0$

We need to show that f is differentiable at c with $f'(c) = M$. Rearranging:

$$\frac{f(x)-f(c)}{x-c} = M + r(x), \qquad x \neq c$$

$$\lim_{x\to c}\frac{f(x)-f(c)}{x-c} = \lim_{x\to c}\big(M + r(x)\big)$$

$$\lim_{x\to c}\frac{f(x)-f(c)}{x-c} = M$$

Hence f is differentiable at c with $f'(c) = M$.

We can modify this defintion slightly by factoring out the $(x-c)$

$$f(x) = f(c) + \big(f'(c) + r(x)\big)(x-c)$$

And then by letting $f_1(x) = f'(c) + r(x)$.

$$f(x) = f(c) + f_1(x)(x-c), \qquad \text{where} \lim_{x \to c} f_1(x) = f'(c)$$

You can show that this goes both ways by letting $r(x) = f_1(x) - f'(c)$.

I will refer to this new definition, as the remainder definition. This definition is useful because it gives us the form for $f(x)$ so we can talk about f in a somewhat more explicit way than we could before.

Rules for Differentiation

We have so far seen how to differentiate functions of the form $f(x) = x^n, n \in \mathbb{W}$. I will hold off for now on differentiating other familiar functions ($x^\alpha, e^x, \ln x$) because we still need some more knowledge before we can tackle these.

We have seen with limits that $\lim(f+g) = \lim f + \lim g$, $\lim(fg) = \lim f \cdot \lim g$, $\lim \frac{f}{g} = \frac{\lim f}{\lim g}$ (for $g, \lim g \neq 0$), and (if f is continuous) $\lim f\big(g(x)\big) = f(\lim g(x))$.

We might hope for similar results to hold for derivatives.

Indeed it is not difficult to show that $\frac{\mathrm{d}}{\mathrm{d}x}\big(f(x) + g(x)\big) = f'(x) + g'(x)$, but when it comes to multiplication, division and composition, the corresponding derivative results are a bit more complicated than these.

$$\text{Theorem(s) 5.3:}$$

$$\text{Let } f \text{ and } g \text{ be differentiable, then:}$$

$$(f+g)' = f' + g'$$

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

$$\left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2} \text{ (as long as } g(x) \neq 0)$$

$$(f \circ g)' = (f' \circ g) \times g'$$

$$(f^{-1})' = \frac{1}{f' \circ f^{-1}} \ \big(\text{as long as } f'\big(f^{-1}(x)\big) \neq 0\big)$$

These are usually proved by using the limit definition of the derivative, starting with the LHS, manipulating until you have the RHS, but this often requries clever manipulations which are difficult to come up with, certainly if you don't already suspect what the RHS is going to be.

I will proof these results (except the first which is easy using limits) by using the remainder definition, which will allow us to derive them (so the motivation for each step should be clear).

Addition is certainly the easiest. Simply by considering limits.

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

$$= \lim_{h \to 0} \left(\frac{f(x+h) - f(x)}{h} + \frac{g(x+h) - g(x)}{h} \right)$$

$$= \lim_{h \to 0} \left(\frac{f(x+h) - f(x)}{h} \right) + \lim_{h \to 0} \left(\frac{g(x+h) - g(x)}{h} \right)$$

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

Next, the product rule. This is often done using the limit definition of the derivative, but I will use our remainder definition instead.

Since f and g are both differentiable at c:

$$f(x) = f(c) + f_1(x)(x-c), \qquad g(x) = g(c) + g_1(x)(x-c)$$

$$\text{Where } \lim_{x \to c} f_1(x) = f'(c) \text{ and } \lim_{x \to c} g_1(x) = g'(c)$$

We then let $h(x) = f(x)g(x)$. Then we then try to write $h(x)$ in the form $h(c) + h_1(x)(x-c)$. Then to determine $h'(c)$, we find $\lim_{x \to c} h_1(x)$.

$$h(x) = \big(f(c) + f_1(x)(x-c)\big)\big(g(c) + g_1(x)(x-c)\big)$$

$$= f(c)g(c) + f_1(x)(x-c)g(c) + g_1(x)(x-c)f(c) + f_1(x)g_1(x)(x-c)^2$$

$$= h(c) + [f_1(x)g(c) + f(c)g_1(x) + f_1(x)g_1(x)(x-c)](x-c)$$

Then we let $h_1(x) = f_1(x)g(c) + f(c)g_1(x) + f_1(x)g_1(x)(x-c)$

$$\text{so that } h(x) = h(c) + h_1(x)(x-c)$$

Finally, $\lim_{x \to c} h_1(x) = f'(c)g(c) + f(c)g'(c) + f'(c)g'(c) \cdot 0 = f'(c)g(c) + f(c)g'(c)$

Hence the derivative of $f(x)g(x)$ at c is $f'(c)g(c) + f(c)g'(c)$. This is the product rule.

(Note that the product rule along with $\frac{\mathrm{d}}{\mathrm{d}x}(k) = 0$ gives us $\frac{\mathrm{d}}{\mathrm{d}x}[kf(x)] = k\frac{\mathrm{d}}{\mathrm{d}x}[f(x)]$, where k is any constant).

Next, the quotient rule. We can simplify the task by considering only $\frac{1}{g(x)}$, and then use the product rule.

$$\text{Let } h(x) = \frac{1}{g(x)} = \frac{1}{g(c) + g_1(x)(x-c)}$$

Remember we want this in the form $h(c) + h_1(x)(x-c)$. Assuming $g(c) \neq 0$

$$h(x) = \frac{1}{g(c)} - \frac{1}{g(c)} + \frac{1}{g(c) + g_1(x)(x-c)}$$

$$= h(c) - \frac{g(c) + g_1(x)(x-c)}{g(c)[g(c) + g_1(x)(x-c)]} + \frac{g(c)}{g(c)[g(c) + g_1(x)(x-c)]}$$

$$= h(c) - \frac{g_1(x)}{g(c)[g(c) + g_1(x)(x-c)]}(x-c)$$

Then let $h_1(x) = -\frac{g_1(x)}{g(c)[g(c)+g_1(x)(x-c)]}$,

$$\text{then } h(x) = h(c) + h_1(x)(x-c)$$

$$\text{Then } h'(c) = \lim_{x\to c} h_1(x) = -\frac{g'(c)}{g(c)[g(c) + g'(c)(0)]} = \frac{-g'(c)}{g(c)^2}$$

Hence $\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{g(x)}\right) = -\frac{g'(x)}{g(x)^2}$.

Then use the product rule to get the full quotient rule:

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{f(x)}{g(x)}\right) = f'(x) \times \frac{1}{g(x)} + f(x) \times -\frac{g'(x)}{g(x)^2}$$

$$= \frac{f'(x)}{g(x)} - \frac{f(x)g'(x)}{g(x)^2} = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}$$

Next, the chain rule. Again the remainder definition makes this easier than it would otherwise have been. Instead of f and g each being differentiable at c, we suppose that g is differentiable at c and f is differentiable at $g(c)$.

Then we can write:

$$f(u) = f\big(g(c)\big) + f_1(u)\big(u - g(c)\big), \qquad \text{where } \lim_{u\to g(c)} f_1(u) = f'\big(g(c)\big)$$

$$g(x) = g(c) + g_1(x)(x-c), \qquad \text{where } \lim_{x\to c} g_1(x) = g'(c)$$

$$\text{Let } h(x) = f\big(g(x)\big)$$

$$= f\big(g(c)\big) + f_1\big(g(x)\big)\big(g(x) - g(c)\big)$$

$$= h(c) + f_1\big(g(c) + g_1(x)(x-c)\big)\big(g(c) + g_1(x)(x-c) - g(c)\big)$$

$$= h(c) + f_1\big(g(c) + g_1(x)(x-c)\big)g_1(x)(x-c)$$

Then we let $h_1(x) = f_1\big(g(c) + g_1(x)(x-c)\big)g_1(x)$.

So that $h(x) = h(c) + h_1(x)(x-c)$.

Finally, $h'(c) = \lim_{x\to c} h_1(x)$

$$\lim_{x\to c} h_1(x) = \lim_{x\to c} f_1\big(g(c) + g_1(x)(x-c)\big)g_1(x)$$

$$= \lim_{x\to c} f_1\big(g(c) + g_1(x)(x-c)\big) \times \lim_{x\to c} g_1(x)$$

$$= \lim_{x\to c} f_1\big(g(c) + g_1(x)(x-c)\big) \times g'(c)$$

We know that $\lim_{x\to c}\big(g(c) + g_1(x)(x-c)\big) = g(c)$.

We could bring the limit inside of f_1 if we knew that f_1 were continuous at $g(c)$. We know that $\lim_{u\to g(c)} f_1(u) = f'\big(g(c)\big)$. For continuity then, we need $f_1\big(g(c)\big) = f'\big(g(c)\big)$. Note that we have not yet defined $f_1\big(g(c)\big)$, so I now define $f_1\big(g(c)\big) = f'\big(g(c)\big)$. Now f_1 is continuous at $g(c)$ and we can take the limit inside.

$$h'(c) = f_1\left(\lim_{x\to c}\big(g(c) + g_1(x)(x-c)\big)\right) \times g'(c)$$

$$= f_1(g(c) + g'(c)\cdot 0) \times g'(c)$$

$$= f_1\big(g(c)\big) \times g'(c)$$

$$= f'\big(g(c)\big) \times g'(c)$$

$$\text{Hence } \frac{\mathrm{d}}{\mathrm{d}x}(f\big(g(x)\big) = f'\big(g(x)\big)g'(x)$$

Finally we have the inverse function rule. For this we can simply use the chain rule:

$$f\big(f^{-1}(x)\big) = x$$

Differentiate both sides.

$$f'\big(f^{-1}(x)\big) \times \frac{\mathrm{d}}{\mathrm{d}x} f^{-1}(x) = 1$$

$$\Rightarrow \frac{\mathrm{d}}{\mathrm{d}x} f^{-1}(x) = \frac{1}{f'\big(f^{-1}(x)\big)}$$

$$\text{(assuming } f \text{ is differentiable at } f^{-1}(x) \text{ and that } f'\big(f^{-1}(x)\big) \neq 0$$

We have now successfully proved the standard results for differentiating functions which are composed of other standard functions, but for these to be practically useful, we need to be able to differentiate those standard functions themselves. So far we know how to differentiate x^n where $n \in \mathbb{W}$ (recall $\mathbb{W} := \{0,1,2,3,\cdots\}$ is the whole numbers).

We might also wish to differentiate x^α where $\alpha \in \mathbb{R}$ can be any real number. The definition of x^α is

$$x^\alpha = e^{\alpha \log x}$$

So we need to be able to first differentiate e^x and $\log x$. Since these are inverses of each other, if I can differentiate one, I can differentiate the other by using the inverse function rule.

If we were to differentiate the infinite series of $e^x := \sum_{n=0}^{\infty} \frac{x^n}{n!}$ term-by-term, we would find the same series we started with, which suggests that $\frac{\mathrm{d}}{\mathrm{d}x} e^x = e^x$. Note that this is not a proof because we haven't proved that infinite series can be differentiated termwise (term-by-term).

We can take a more direct approach by using the definition of the derivative. The limit definition is easy enough to work with in this case, though one small detail makes the proof harder than it first seems.

$$\text{Theorem 5.4:}$$

$$\frac{\mathrm{d}}{\mathrm{d}x} e^x = e^x$$

pf:

$$\frac{\mathrm{d}}{\mathrm{d}x} e^x = \lim_{h \to 0} \frac{e^{x+h} - e^x}{h} = e^x \lim_{h \to 0} \frac{e^h - 1}{h} = e^x \lim_{h \to 0} \sum_{n=1}^{\infty} \frac{h^{n-1}}{n!}$$

$$= e^x \left(1 + \lim_{h \to 0} \sum_{n=2}^{\infty} \frac{h^{n-1}}{n!} \right)$$

It might seem obvious that $\lim_{h \to 0} \sum_{n=2}^{\infty} \frac{h^{n-1}}{n!} = 0$ but you can't just take the limit inside the series because an infinite series is a limit (of partial sums), and you can't always swap the order of any two limits. We can prove that this limit is indeed 0 by using the comparison test, and then the squeeze theorem.

We have seen previously that $n! \geq 2^{n-1}$ for $n \in \mathbb{N}$ (problem sheet Q13).

$$\Rightarrow \frac{|h^{n-1}|}{n!} \leq \frac{|h^{n-1}|}{2^{n-1}}$$

$$\Rightarrow -\frac{|h^{n-1}|}{2^{n-1}} \leq \frac{h^{n-1}}{n!} \leq \frac{|h^{n-1}|}{2^{n-1}}$$

Then use the comparison test,

$$\sum_{n=2}^{\infty} \frac{|h^{n-1}|}{2^{n-1}} \text{ converges (for } |h| < 2) \text{ to} \frac{|h|}{2-|h|} \text{ (geometric series)}$$

$$\text{Likewise, } \sum_{n=2}^{\infty} -\frac{|h^{n-1}|}{2^{n-1}} = \frac{|h|}{|h|-2}$$

$$\text{So } \frac{|h|}{|h|-2} \leq \sum_{n=2}^{\infty} \frac{h^{n-1}}{n!} \leq \frac{|h|}{2-|h|}$$

Then use the squeeze theorem,

$$\lim_{h\to 0} \frac{|h|}{|h|-2} = \frac{0}{0-2} = 0 \text{ and } \lim_{h\to 0} \frac{|h|}{2-|h|} = \frac{0}{2-0} = 0$$

$$\text{So } \lim_{h\to 0} \sum_{n=2}^{\infty} \frac{h^{n-1}}{n!} = 0$$

Which was what we wanted. Returning to the original thing we were trying to prove:

$$\frac{\mathrm{d}}{\mathrm{d}x} e^x = e^x \left(1 + \lim_{h\to 0} \sum_{n=2}^{\infty} \frac{h^{n-1}}{n!}\right) = e^x(1+0) = e^x$$

$$\frac{\mathrm{d}}{\mathrm{d}x} e^x = e^x$$

We now can use the inverse function rule, with $f(x) = e^x \Rightarrow f^{-1}(x) = \log x$

$$\frac{\mathrm{d}}{\mathrm{d}x} \log x = \frac{1}{e^{\log x}} = \frac{1}{x}$$

We can also now differentiate $x^{\alpha}, x > 0$.

$$\frac{\mathrm{d}}{\mathrm{dx}} x^{\alpha} = \frac{\mathrm{d}}{\mathrm{d}x} e^{\alpha \log x} = e^{\alpha \log x} \times \frac{\alpha}{x} = x^{\alpha} \times \frac{\alpha}{x} = \alpha x^{\alpha-1}$$

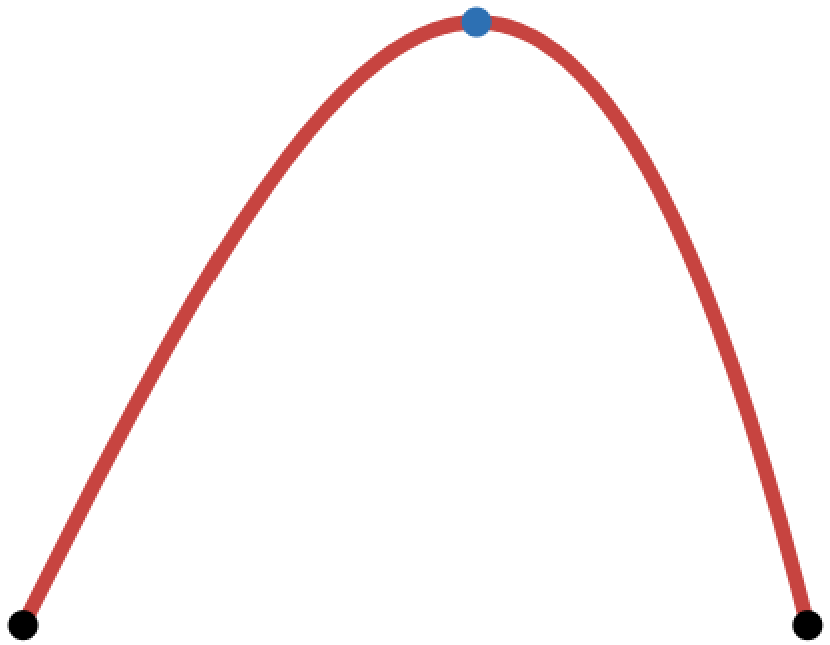
Which is the same as we have for $\alpha \in \mathbb{W}$, which is good news.

The Mean Value Theorem

Previously we have seen the intermediate and extreme value theorems. There are a few more theorems similar to these, which we will look at here.

The first is **Rolle's Theorem** which states, if f is differentiable on $[a, b]$ and $f(a) = f(b)$ then there exists some $c \in (a, b)$ such that $f(c) = 0$.

E.g.,



The black points are the end points. This function is differentiable everywhere between these points and so there exists a point (the blue point) where the gradient is 0. If you go up you must come back down, if you go down you must come back up, so you must have a turning point where the derivative is zero. That is at least the intuition, of course we also need proof.

First we prove a related lemma (it's relevance is clear from the intuition).

$$\text{Lemma 5.5:}$$

$$\text{If } f \text{ is differentiable at } c \text{ and } f(c) \text{ is a local maximum (or minmum)}$$

$$\text{and } c \text{ is an interior point in the domain of } f$$

$$\text{Then } f'(c) = 0$$

"Interior point" just means "not an endpoint". We say that $f(c)$ is a "local maximum" if $f(c)$ is the maximum in some local nieghbourhood. More precisely, if there exists $\delta > 0$

such that $|x-c|<\delta \Rightarrow f(c) \geq f(x)$. Similarly $f(c)$ is a local minimum if $\exists_{\delta>0}\ \text{s.t.}, |x-c|<\delta \Rightarrow f(c) \leq f(x)$.

This proof is short and sweet. It uses left and right limits. Recall that a limit exists if and only if the left and right limits both exist and they agree.

pf:

We suppose that $f(c)$ is a maximum (the minimum case is essentially the same).

Since we are given that f is differentiable at c, $f'(c) = \lim_{x \to c} \frac{f(x)-f(c)}{x-c}$ exists.

This means the left and right limits both exist and are each equal to $f'(c)$

We know that there is some $\delta > 0$ so that $|x-c|<\delta \Rightarrow f(c) \geq f(x)$. This is because $f(c)$ is a local maximum and is an interior point.

[x being an interior point is important as it guarantees that we can choose δ small enough that $(c-\delta, c+\delta)$ lies entirely within $[a,b]$.]

Let $x \in (c-\delta, c+\delta)$, then $f(c) \geq f(x)$.

When $x > c$, $f(x)-f(c) \leq 0$ and $x-c>0$ meaning that $\frac{f(x)-f(c)}{x-c} \leq 0$.

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

When $x < c$, $f(x)-f(c) \leq 0$ and $x-c<0$ meaning that $\frac{f(x)-f(c)}{x-c} \geq 0$.

Hence $\lim_{x \to c^-} \frac{f(x)-f(c)}{x-c} = f'(c) \geq 0$.

$$f'(c) \leq 0 \text{ and } f'(c) \geq 0$$

$$\Rightarrow f'(c) = 0$$

Rolle's theorem follows almost immediately.

Corollary 5.6: **Rolle's Theorem**

If f is differentiable on $[a,b]$ and $f(a) = f(b)$

Then there exists some $c \in (a,b)$ such that $f'(c) = 0$

pf:

Recalling our intuition before, we expect that a maximum or minimum point will achieve this.

By the extreme value theorem, there exists a $m, M \in [a, b]$ such that $f(m)$ is a global minimum and $f(M)$ is a global maximum. $f(m) \le f(x) \le f(M)$ for every $x \in [a, b]$.

There are 3 cases.

Case 1: $f(m) = f(M) = f(a)$. If this is not true then either $f(m) < f(a)$ or $f(M) > f(a)$ (or both).

$$\text{If } f(m) = f(M) \text{ then } f \text{ is constant.}$$

$$\text{The derivative of a constant is zero everywhere, so any point } c \in (a, b) \text{ will do.}$$

$$\text{If } f(m) < f(a) = f(b)$$

$$\text{Then } m \ne a \text{ and } m \ne b \text{ so } m \text{ is an interior point.}$$

$$\text{Also since } f \text{ is differentiable everywhere, } f \text{ is differentiable at } m.$$

$$\text{By the previous lemma, } f'(m) = 0$$

$$\text{If } f(M) > f(a) = f(b)$$

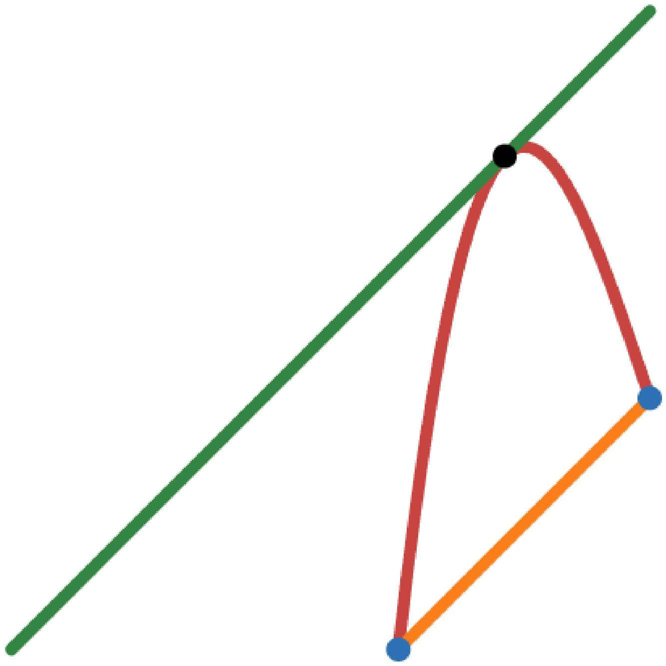
$$\text{Then } M \ne a \text{ and } M \ne b \text{ so } M \text{ is an interior point. Also } f \text{ is differentiable at } M.$$

$$\text{By the previous lemma, } f'(M) = 0$$

We are done.

The next theorem is a generalization of Rolle's theorem. It is called the **Mean Value Theorem** (MVT). It could be formulated in terms of integrals, but we have not defined integrals yet. (In fact we will need to use MVT to prove one of the most important fact about integrals, the fundamental theorem of calculus.)

We define the mean value theorem for now in terms of derivatives. If I have a function f differentiable on $[a, b]$ then there is some point $c \in (a, b)$ where the derivative of f at c is equal to the average (mean) slope of the curve. By "mean slope" of the curve, I really mean the gradient of the line which connects the endpoints.



Here, there is **a black point** with an x coordinate between the **endpoints** so that the **tangent line** at the **black point** has the same gradient as the **secant line** between the **endpoints**.

A "secant line" is just a line connecting 2 points. Rolle's theorem is a special case of MVT in which $f(a) = f(b)$.

The gradient of the tangent line is $f'(c)$ and the gradient of the secant line is $\frac{f(b)-f(a)}{b-a}$.

Since the core idea of this theorem is similar to Rolle's theorem, we might be able to cut out the bulk of the work by invoking Rolle's theorem in some clever way.

If we consider the difference between the **secant line** and the **curve** we notice that the difference is zero at the **endpoints**. If we consider then the function which tells us this difference, we can apply Rolle's theorem to that function. Then hopefully this will give us the result we want.

Mean Value Theorem (5.7):

If f is differentiable on $[a, b]$

$$\text{Then } \exists_{c\in(a,b)} \text{ s.t.}, f'(c) = \frac{f(b) - f(a)}{b - a}$$

pf:

First we determine the equation of the secant line. The line has equation $y = mx + c$. It has gradient $\frac{f(b)-f(a)}{b-a}$ so this becomes $y = \frac{f(b)-f(a)}{b-a}x + c$. It passes through $(a, f(a))$ so $c = y - \frac{f(b)-f(a)}{b-a}x = f(a) - \frac{f(b)-f(a)}{b-a}a$. We now have the equation of the secant line:

$$y = \frac{f(b) - f(a)}{b - a}x + f(a) - \frac{f(b) - f(a)}{b - a}a$$

We are now prepared to make our educated guess of a function which might be helpful. The difference between the curve $f(x)$ and this line.

$$\text{Let } g(x) \coloneqq f(x) - \frac{f(b) - f(a)}{b - a}x - f(a) + \frac{f(b) - f(a)}{b - a}a$$

f is continuous on $[a, b]$ so g is too. We need to then check that $g(a) = g(b)$.

$$g(a) = f(a) - \frac{f(b) - f(a)}{b - a}a - f(a) + \frac{f(b) - f(a)}{b - a}a = 0$$

$$g(b) = f(b) - \frac{f(b) - f(a)}{b - a}b - f(a) + \frac{f(b) - f(a)}{b - a}a$$

$$= [f(b) - f(a)] + \frac{f(b) - f(a)}{b - a}(a - b)$$

$$= [f(b) - f(a)] - [f(b) - f(a)] = 0 = f(a)$$

Then we apply Rolle's theorem.

$$\text{There exists a } c \in (a, b) \text{ such that } g'(c) = 0$$

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

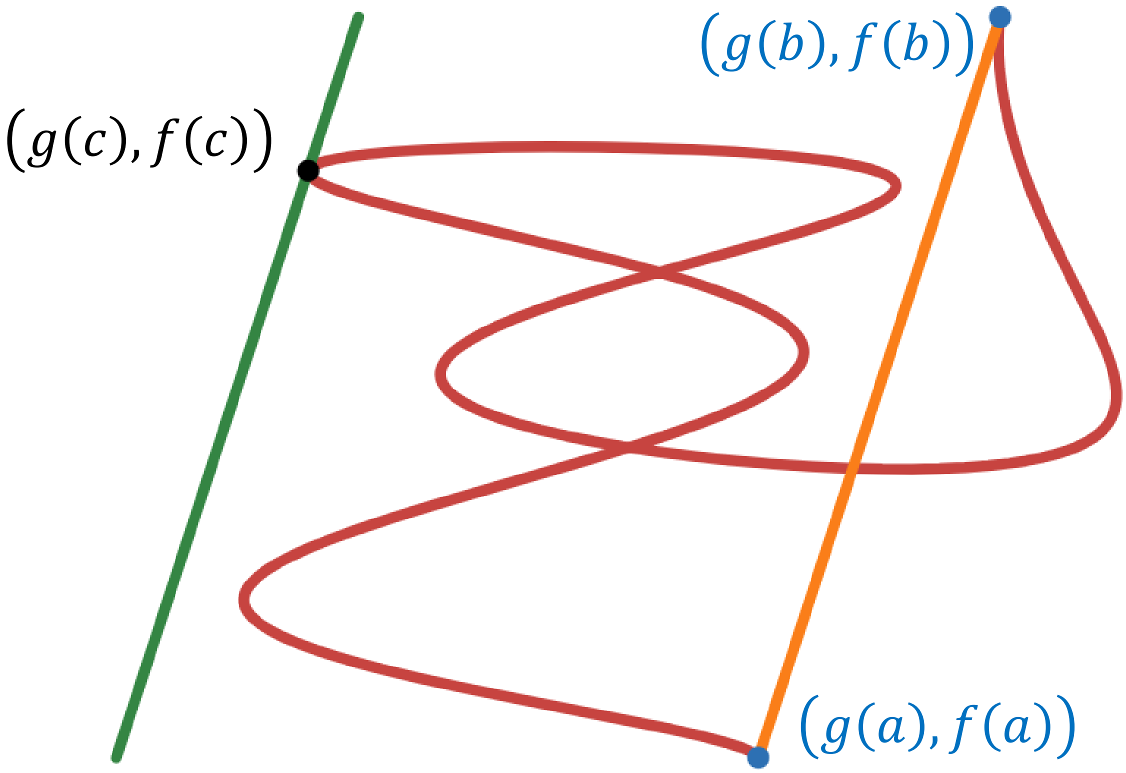
$$\Rightarrow f'(c) - \frac{f(b) - f(a)}{b - a} = 0$$

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

Which is precisely the Mean Value Theorem. We are done.

The mean value theorem will be incredibly important when we talk about integrals later on. We can generalise MVT even further with
Cuachy's Generalised Mean Value Theorem (GMVT).

For the intuition of this one, consider parametric equations. Bear in mind the following is just intuition meaning I am allowed to be as unrigorous as I want. If you are unfamiliar with parametric equations, the idea is that instead of thinking about $y = f(x)$ (you give me some x and I give you y) we think of $y = f(t), x = g(t)$ (you give me some t and I give you x and y). Parametric equations allow us to describe a wider variety of curves (e.g., there may be vertical lines which cross the curve more than once, and the curve may even cross over itself). The graph of $y = f(x)$ on some interval, I is the set of all points $\{(x, f(x)): x \in I\}$. The graph of some parametric equations where t ranges over I is the set of points $\{(g(t), f(t)): t \in I\}$.



Here is a diagram of the graph of some arbitrary parametric equations $x = g(t), y = f(t)$, for $t \in [a, b]$ with f and g both differentiable on that interval.

Cauchy's Generalised Mean Value Theorem (GMVT) essentially says, if we call the gradient of the **secant line** m, then there exists some $c \in (a, b)$ where the gradient of the curve at $t = c$ is equal to m. For the regular MVT, the gradient of the curve was given by $f'(c)$. What would it be in this case? Roughly speaking (we are still in the intuition stage) $x = g(t) \Rightarrow \frac{\mathrm{d}x}{\mathrm{d}t} = g'(t)$ and $y = f(t) \Rightarrow \frac{\mathrm{d}y}{\mathrm{d}t} = f'(t)$. Therefore the gradient, $\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}y}{\mathrm{d}t} \div \frac{\mathrm{d}x}{\mathrm{d}t} = \frac{f'(t)}{g'(t)}$. The gradient of the **tangent line** at $t = c$ is $\frac{f'(c)}{g'(c)}$. The gradient of the **secant line** is $\frac{f(b)-f(a)}{g(b)-g(a)}$. Note that MVT is just a special case of GMVT in which $g(t) = t$. We are now equipped to precisely state the GMVT.

To avoid any accidental division by zero, instead of writing $\frac{f(b)-f(a)}{g(b)-g(a)} = \frac{f'(c)}{g'(c)}$, we instead write $(f(b) - f(a))g'(c) = (g(b) - g(a))f'(c)$. Writing it like this also allows us to include the case where the secant line is vertical $(g(b) - g(a) = g'(c) = 0)$.

Cauchy's Generalised Mean Value Theorem (5.8):

If f and g are each differentiable on $[a, b]$

Then $\exists_{c \in (a,b)}$ s.t., $\quad (f(b) - f(a))g'(c) = (g(b) - g(a))f'(c)$

pf:

We can actually prove this directly from Rolle's theorem (meaning we could have skipped the proof of MVT if we wanted to).

We use a similar idea, guess some function $h(t)$ and apply Rolle's theorem to it.

$$\text{Let } h(t) = f(t) - \frac{f(b) - f(a)}{g(b) - g(a)} g(t) - f(a) + \frac{f(b) - f(a)}{g(b) - g(a)} g(a)$$

$$\big(\text{Assuming for now that } g(a) \neq g(b)\big)$$

(This represents the vertical difference between the **curve** and the **secant line**. I encourage you to derive this guess for yourself.).

$$\text{Then } h(a) = f(a) - \frac{f(b) - f(a)}{g(b) - g(a)} g(a) - f(a) + \frac{f(b) - f(a)}{g(b) - g(a)} g(a) = 0$$

$$h(b) = [f(b) - f(a)] - \frac{f(b) - f(a)}{g(b) - g(a)} [g(b) - g(a)]$$

$$= [f(b) - f(a)] - [f(b) - f(a)] = 0 = h(a)$$

Hence we can apply **Rolle's Theorem**.

There exists a $c \in (a, b)$ with $h'(c) = 0$. Differentiate h:

$$h'(t) = f'(t) - \frac{f(b) - f(a)}{g(b) - g(a)} g'(t)$$

$$\Rightarrow f'(c) - \frac{f(b) - f(a)}{g(b) - g(a)} g'(c) = 0$$

$$\Rightarrow \big(g(b) - g(a)\big)f'(c) = \big(f(b) - f(a)\big)g'(c)$$

In the edge case where $g(a) = g(b)$, the **secant line** is vertical. We wish to find a $c \in (a, b)$ such that.

$$\big(f(b) - f(a)\big)g'(c) = 0$$

We can easly achieve this using Rolle's Theorem. $g(a) = g(b)$ and g is differentiable on $[a, b]$ meaning there is some $c \in (a, b)$ with $g'(c) = 0$.

$$\Rightarrow \big(f(b) - f(a)\big)g'(c) = \big(f(b) - f(a)\big) \cdot 0 = 0$$

And the edge case is dealt with. We are done.

If we further assume that $g(a) \neq g(b)$ and $g'(x) \neq 0$ for any x, then we can write the original $\frac{f(a)-f(b)}{g(a)-g(b)} = \frac{f'(c)}{g'(c)}$.

Interpretation of Differentiation

We have seen that we get turning points when the derivative is zero. What about when the derivative is positive (or negative)? Clearly we should have a function which is increasing (or decreasing respectively). Also, we expect the function to be constant if its derivative is zero everywhere.

These are our intuitive ways to interpret differentiation and so we may be tempted to assume them to be true, but in the spirit of analysis, we shall prove them.

$$\text{Theorem 5.9:}$$

$$\text{If } f \text{ is differentiable on some interval } I, \text{and } f'(x) \geq 0 \text{ for all } x \in I$$

$$(\text{or } f'(x) \leq 0 \text{ for all } x \in I)$$

$$\text{Then } f \text{ is monotone increasing (or decreasing respectively) on } I$$

$$[\text{Similarly}, f'(x) > 0 (< 0) \Rightarrow f \text{ is } \textit{strictly} \text{ increasing (decreasing)}]$$

$$\text{and } f'(x) = 0 \Rightarrow f \text{ is constant}$$

This proof is relatively easy by using the mean value theorem (MVT).

pf:

$$\text{Suppose } x < y \text{ (with } x, y \in I).$$

$$\text{Then (by MVT) } \exists_{c \in I} \text{ such that } \frac{f(y) - f(x)}{y - x} = f'(c)$$

$$y - x > 0 \text{ meaning:}$$

$$f'(x) \geq 0\ \forall_{x \in I} \Longrightarrow f'(c) \geq 0 \Rightarrow f(y) - f(x) \geq 0 \Rightarrow f(x) \leq f(y) \text{ (monotone increasing)}$$

$$f'(c) \leq 0 \Rightarrow f(x) \geq f(y) \text{ (monotone decreasing)}$$

$$f'(c) > 0 \Rightarrow f(x) < f(y) \text{ (strictly monotone increasing)}$$

$$f'(c) < 0 \Rightarrow f(x) > f(y) \text{ (strictly monotone decreasing)}$$

$$f'(c) = 0 \Rightarrow f(x) = f(y) \text{ (constant function)}$$

L'Hospital's Rule

(The "H" and "s" are silent. You can either write L'Hospital or L'Hôpital.)

L'Hospital's rule is evey calculus student's favourite theorem as it makes many limits much easier to calculate.

When dealing with limits we often break down the expression into smaller parts and deal with each part separetely by using the theorems for combining limits:

$$\lim(f+g) = \lim f + \lim g\,, \qquad \lim fg = \lim f \lim g\,, \qquad \lim \frac{f}{g} = \frac{\lim f}{\lim g}$$

If you can caluculate $\lim f$ and $\lim g$ then you can always find the original limit, except in the case of $\lim \frac{f}{g} = \frac{\lim f}{\lim g}$ when $\lim g$ happens to be 0. In particular, if $\lim f = \lim g = 0$ then $\lim \frac{f}{g}$ may converge, but we don't know what it converges to (it may also diverge). L'Hospitals rule is a tool to solve this problem.

First, the intuitive idea:

Suppose $\lim_{x\to c} f(x) = \lim_{x\to c} g(x) = 0$ and we want to find $\lim_{x\to c} \frac{f(x)}{g(x)}$.

Since we dealing with limits, we only care about what happens whenever x is close to c. If we "zoom in" near c (assuming f and g are differentiable near c), f and g should be approximately straight lines (I am still being intentionally imprecise).

The gradients of these lines will be $f'(c)$ and $g'(c)$ respectively (the slope of the curve at c). The equation of the line approximating f near c will be $y = xf'(c) + k$. To find k, the lines passes through $(c, 0)$ so $0 = cf'(c) + k$ meaning the approximation is:

$$f(x) \approx xf'(c) - cf'(c) = f'(c)(x-c)$$

Similarly, the approximation for g near c is $g(x) \approx xg'(c) - cg'(c) = g'(c)(x-c)$.

These "linear" approximations of f and g will be discussed in more detail later when we see Taylor Series. Finally use these to determine the limit.

$$\lim_{x\to c} \frac{f(x)}{g(x)} = \lim_{x\to c} \frac{f'(c)(x-c)}{g'(c)(x-c)} = \lim_{x\to c} \frac{f'(c)}{g'(c)}$$

Which makes sense when you think about it. f is getting close to zero at a rate $\frac{f'(c)}{g'(c)}$ "faster" than g is, meaning $\frac{f}{g} \to \frac{f'(c)}{g'(c)}$. (It is crucial that both f and g each approach 0. The above argument fails if they approach some other value. If $f \to 3$ and $g \to 3$ then $\frac{f}{g} \to 1$).

Before going into full on rigour mode, I will mention some more details to make this even more general. First, we don't actually need f or g to be differentiable at c, we only care about what happens close to c, so we can use $\lim_{x \to c} \frac{f'(x)}{g'(x)}$ instead. Secondly, this should also work for endpoints and in cases where the left and right limits are not necessarily the same.

Since we have $f \to 0$, $g \to 0$ and are considering $\frac{f}{g}$, we call this the "$\frac{0}{0}$" case of L'Hospital's rule. There is another case I will discuss after.

Theorem 5.10: L'Hôpital's rule (a):

If f and g are differentiable on (a, b), $\lim_{x \to a^+} f(x) = \lim_{x \to a^+} g(x) = 0$, and $\lim_{x \to a^+} \frac{f'(x)}{g'(x)}$ exists.

Also assume that $g(x) \neq 0$ and $g'(x) \neq 0$ for any $x \in (a, b)$.

Then $\lim_{x \to a^+} \frac{f(x)}{g(x)} = \lim_{x \to a^+} \frac{f'(x)}{g'(x)}$

One might notice this is not the first time we have seen something like $\frac{f'(x)}{g'(x)}$. In the previous section, **Cauchy's Generalised Mean Value Theorem** mentions something of a similar form ($\frac{f(b)-f(a)}{g(b)-g(a)} = \frac{f'(c)}{g'(c)}$ for some c) and so attempting to apply GMVT might be something to look out for.

pf:

First, since the limits of f and g exist $(= 0)$, we can actually make them continuous by simply defining $f(a) = g(a) = 0$.

For any sequence $x_n \in (a, b)$ with $\lim x_n = a$, we have $\lim f(x_n) = \lim g(x_n) = 0$

$$\left[\text{since } \lim_{x \to a^+} f(x) = \lim_{x \to a^+} g(x) = 0\right]$$

Then we use GMVT: For each x_n in this sequence, there exists some $y_n \in (a, x_n)$ such that $\frac{f(x_n)-f(a)}{g(x_n)-g(a)} = \frac{f'(y_n)}{g'(y_n)}$.

Then (since $f(a) = g(a) = 0$), $\frac{f(x_n)}{g(x_n)} = \frac{f'(y_n)}{g'(y_n)}$

Then by using squeezing:

$$a < y_n < x_n \text{ and } \lim x_n = a \Longrightarrow \lim y_n = a$$

$$\lim_{x\to a^+}\frac{f(x)}{g(x)} = \lim\frac{f(x_n)}{g(x_n)} = \lim\frac{f'(y_n)}{g'(y_n)} = \lim_{x\to a^+}\frac{f'(x)}{g'(x)}$$

We can immediately extend this to two-sided limits. If $L = \lim_{x\to a}\frac{f'(x)}{g'(x)}$ exists then $L = \lim_{x\to a^+}\frac{f'(x)}{g'(x)} = \lim_{x\to a^-}\frac{f'(x)}{g'(x)}$. Using the above: $\lim_{x\to a^+}\frac{f(x)}{g(x)} = \lim_{x\to a^+}\frac{f'(x)}{g'(x)} = L$ and $\lim_{x\to a^-}\frac{f(x)}{g(x)} = \lim_{x\to a^-}\frac{f'(x)}{g'(x)} = L$. Hence $\lim_{x\to a}\frac{f(x)}{g(x)} = L = \lim_{x\to a}\frac{f'(x)}{g'(x)}$.

And we are done!

We have seen one and two sided limits, but we can also apply L'Hospitals rule with limits where $x\to\infty$ or $x\to-\infty$. Let's consider $+\infty$.

$$\text{Theorem 5.10: L'Hôpital's rule (b):}$$

$$\text{If } f \text{ and } g \text{ are differentiable on } (A,\infty) \text{ (for some } A>0\text{)}$$

$$\text{and } \lim_{x\to\infty} f(x) = \lim_{x\to\infty} g(x) = 0, \text{and } \lim_{x\to\infty}\frac{f'(x)}{g'(x)} \text{ exists.}$$

$$\text{Also assume that } g(x)\neq 0 \text{ and } g'(x)\neq 0 \text{ for any } x\in(A,\infty).$$

$$\text{Then } \lim_{x\to\infty}\frac{f(x)}{g(x)} = \lim_{x\to\infty}\frac{f'(x)}{g'(x)}$$

pf:

Let $\phi(x) = f\left(\frac{1}{x}\right)$ and $\psi(x) = g\left(\frac{1}{x}\right)$. Then $\lim_{x\to 0^+}\phi(x) = \lim_{x\to\infty} f(x) = 0$ and $\lim_{x\to 0^+}\psi(x) = 0$.

(ϕ is greek letter phi, ψ is greek letter psi).

Also define $\phi(0) = \psi(0) = 0$, so ϕ and ψ are right-continuous (continuous from the right hand side) at 0. f is differentiable and so is $\frac{1}{x}$ (for $x>0$) which means that $\phi(x)$ is differentiable also. Similarly $\psi(x)$ is differentiable. We also have $\lim_{x\to 0^+}\phi(x) = \lim_{x\to 0^+}\psi(x) = 0$ so apply part (a).

$$\lim_{x\to 0^+}\frac{f\left(\frac{1}{x}\right)}{g\left(\frac{1}{x}\right)} = \lim_{x\to 0^+}\frac{\phi(x)}{\psi(x)} = \lim_{x\to 0^+}\frac{\phi'(x)}{\psi'(x)} = \lim_{x\to 0^+}\frac{-\frac{1}{x^2}f'\left(\frac{1}{x}\right)}{-\frac{1}{x^2}g'\left(\frac{1}{x}\right)} = \lim_{x\to 0^+}\frac{f'\left(\frac{1}{x}\right)}{g'\left(\frac{1}{x}\right)}$$

Then use the substitution $t = \frac{1}{x}$, (as $x\to 0^+$, $t\to+\infty$). The right hand side becomes $\lim_{t\to\infty}\frac{f'(t)}{g'(t)}$. We are given that this limit exists and so this is okay. The left hand side becomes $\lim_{t\to\infty}\frac{f(t)}{g(t)}$

$$\lim_{t\to\infty}\frac{f(t)}{g(t)} = \lim_{t\to\infty}\frac{f'(t)}{g'(t)}$$

And we are done!

There is yet another case where L'Hospital's rule still applies. The intuition is similar to the "$\frac{0}{0}$" case. It is referred to as the "$\frac{\infty}{\infty}$" case.

As the name would suggest, you can apply it whenever $\lim_{x\to a} f(x) = \infty$ and $\lim_{x\to a} g(x) = \infty$.

$$\text{Theorem 5.10: L'Hôpital's rule (c):}$$

$$\text{If } f \text{ and } g \text{ are differentiable in } (a, b) \text{ and } \lim_{x\to a^+} f(x) = \lim_{x\to a^+} g(x) = \infty,$$

$$\text{and } \lim_{x\to a^+}\frac{f'(x)}{g'(x)} \text{ exists, and } \lim_{x\to a^+}\frac{f(x)}{g(x)} \text{ exists}$$

$$\text{and } g'(x), f'(x) \neq 0 \text{ for all } x \text{ sufficiently close to } a.$$

$$\text{Then } \lim_{x\to a^+}\frac{f(x)}{g(x)} = \lim_{x\to a^+}\frac{f'(x)}{g'(x)}$$

I have written $x \to a^+$, but this should work for one-sided, two-sided, and infinite ($x \to \pm\infty$) limits too.

pf:

Since $\frac{1}{f(x)} \to 0$ and $\frac{1}{g(x)} \to 0$, we apply the "$\frac{0}{0}$" case. Since $f(x), g(x) \to \infty$, we know that $f(x) \neq 0$ and $g(x) \neq 0$ for all x sufficiently close to a.

$$\text{Let } L = \lim_{x\to a^+}\frac{f(x)}{g(x)} = \lim_{x\to a^+}\frac{1/g(x)}{1/f(x)}$$

Differentiating the top and bottom:

$$L = \lim_{x\to a^+}\frac{-g'(x)/g(x)^2}{-f'(x)/f(x)^2} \text{ (Assuming that this limit exists)}$$

$$L = \left(\lim_{x\to a^+}\frac{f(x)}{g(x)}\right)^2 \lim_{x\to a^+}\frac{g'(x)}{f'(x)} = L^2 \lim_{x\to a^+}\frac{g'(x)}{f'(x)}$$

$$(\text{assuming } L \neq 0) \Rightarrow L = \lim_{x\to a^+}\frac{f'(x)}{g'(x)}$$

I will refer to the result we have so far ($L \neq 0$ case) as (†). If $L = 0$ then we can be clever. Instead of considering $\frac{f(x)}{g(x)}$, consider $\frac{f(x)+g(x)}{g(x)}$ (which is fine since $f(x) + g(x) \to \infty$).

$$\text{Let } L_2 = \lim_{x \to a^+} \frac{f(x) + g(x)}{g(x)} = \lim_{x \to a^+} \frac{f(x)}{g(x)} + 1 = L + 1 = 1$$

We can therefore apply (†) to $\frac{f(x)+g(x)}{g(x)}$ (since the limit is not zero).

$$1 = \lim_{x \to a^+} \frac{f(x) + g(x)}{g(x)} = \lim_{x \to a^+} \frac{f'(x) + g'(x)}{g'(x)} = \lim_{x \to a^+} \frac{f'(x)}{g'(x)} + 1$$

$$\text{So } \lim_{x \to a^+} \frac{f'(x)}{g'(x)} + 1 = 1$$

$$\Rightarrow \lim_{x \to a^+} \frac{f'(x)}{g'(x)} = 0 = L$$

We Are Done.

If $f \to \infty$ and $g \to -\infty$ or $f \to -\infty$ and $g \to \infty$ or $f \to -\infty$ and $g \to -\infty$, we can simply factor out the minuses from the limit, and the result still works.

We have proved all cases of L'Hospital's rule.

To summarise the results from this section. If f and g are differentiable near a and $L = \lim_{x \to a} \frac{f'(x)}{g'(x)}$ exists, and $\lim_{x \to a} f(x) = \lim_{x \to a} g(x) = 0$.

$$\text{Then } \lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{x \to a} \frac{f'(x)}{g'(x)}$$

If instead we have $\lim_{x \to a} f(x) = \pm\infty$ and $\lim_{x \to a} g(x) = \pm\infty$, and we also know that both of the following limits exist, then:

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{x \to a} \frac{f'(x)}{g'(x)}$$

These also work for one sided limits and limits with $x \to \infty$ or $x \to -\infty$.

Part (a) is the only version of this theorem which we will be using, but the rest of the cases are useful for practical purposes of evaluating limits.

Functions as Infinite Series

Taylor's Theorem

The intuition of Maclaurin series is that if I have a function f which I wish to approximate, I assume that it can be written as some kind of "infinite polynomial".

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots$$

Then to find the coefficients, a_0, a_1 etc., we differentiate and evaluate at 0.

$$f(0) = a_0$$

$$f'(x) = a_1 + 2a_2 x + 3a_3 x^2 + \cdots \Rightarrow f'(0) = a_1$$

$$f''(x) = 2a_2 + 3!\, a_3 x + \cdots \Rightarrow f''(0) = 2a_2 \Rightarrow a_2 = \frac{f''(0)}{2}$$

$$f'''(x) = 3!\, a_3 + \cdots \Rightarrow f'''(0) = 3!\, a_3 \Rightarrow a_3 = \frac{f'''(0)}{3!}$$

$$\vdots$$

$$a_n = \frac{f^{(n)}(0)}{n!}$$

Where $f^{(n)}$ is the n^{th} derivative of f.

$$\text{Then } f(x) = \frac{f(0)}{0!} + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \cdots$$

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

This is ***not*** a proof, because we made some unjustified assumptions. Namely that we can write f as one of these "infinite polynomials" and that we are allowed to differentiate this infinite series term-wise (term-by-term).

This idea provides some more motivation to the series definition of the exponential function. We want a function which is its own derivative and such that $f(0) = 1$. Then $f(0) = f'(0) = f''(0) = \cdots = 1 \Rightarrow f(x) = \sum_{n=0}^{\infty} \frac{1}{n!} x^n$.

One function which we do not have an infinite series representation for is $\log x$, which we have defined as the inverse of the exponential function. Unfortunately, we cannot create a Maclaurin series for $\log x$, because $\log 0$ is undefined. In fact, each of the derivatives, $\frac{1}{x}, -\frac{1}{x^2}, \frac{2}{x^3}, \cdots$ are each undefined at $x = 0$. We can however create a Maclaurin series for $f(x) = \log(1 + x)$. $f(0) = \log 1 = 0$. $f'(x) = \frac{1}{1+x} \Rightarrow f'(0) = 1$. $f''(x) = -\frac{1}{(1+x)^2} \Rightarrow f''(0) = -1, f'''(x) = \frac{2}{(1+x)^3} \Rightarrow f'''(0) = 2 \cdots$. In general we can show (by induction) that $f^{(n)}(x) = \frac{(-1)^{n-1}(n-1)!}{(1+x)^n} \Rightarrow f^{(n)}(0) = (-1)^{n-1}(n-1)!$ [for $n \geq 1$].

$$\text{Then } f(x) = \log(1+x) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}(n-1)!}{n!} x^n = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} x^n$$

$$= x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots$$

We can then replace x by $x - 1$ to get:

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

This then motivates the idea of Taylor series:

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

For situations where it might not be possible to evaluate the function or its derivatives at 0. We call this a Taylor series centred at c. A Maclaurin series is just a Taylor series centred at 0.

The intuition I have provided so far is likely familiar to those of you who have taken calculus before, but of course the goal of this course is to make these ideas more formal. We do this via **Taylor's Theorem** (among other things) which will tell use just how good of an approximation this kind of series will be.

We define the n^{th} order Taylor polynomial of f as:

$$\sum_{k=0}^{n} \frac{f^{(k)}(c)}{k!} (x-c)^k$$

We wish to determine how well this polynomial approximates f. One thing to note is that if f is differentiable at c then $f(x) = f(c) + f'(c)(x-c) + r(x)(x-c)$ with $\lim_{x \to c} r(x) = 0$ (remainder definition of differentiability). Note that $f(c) + f'(c)(x-c)$ is the 1^{st} order Taylor polynomial of f centred at c. Perhaps we will have a similar result if f is n times differentiable at c (meaning $f^{(n)}(c)$ is defined).

Taylor's Theorem (Peano) 6.1:

If f is n times differentiable at c, where $n \in \mathbb{N}$, and $f^{(n)}$ is continuous at c

$$\text{Then } f(x) = \sum_{k=0}^{n} \frac{f^{(n)}(c)}{n!}(x-c)^n + r_n(x)(x-c)^n \text{ where } \lim_{x \to c} r_n(x) = 0$$

We call $r_n(x)(x-c)^n$ the **Peano Form of the Remainder**.

pf:

The proof is relatively simple.

$$\text{Define } r_n(x) = \frac{f(x) - \sum_{k=0}^{n} \frac{f^{(n)}(c)}{n!}(x-c)^n}{(x-c)^n}, x \neq c$$

To make this easier to write down,

$$\text{Let } R_n(x) = f(x) - \sum_{k=0}^{n} \frac{f^{(k)}(c)}{k!}(x-c)^k$$

$$\left[= f(x) - \left(f(c) + f'(c)(x-c) + \frac{f''(c)}{2!}(x-c)^2 + \frac{f'''(c)}{3!}(x-c)^3 + \cdots\right)\right]$$

$$\text{Then } r_n(x) = \frac{R_n(x)}{(x-c)^n}$$

We wish to show that $\lim_{x \to c} r_n(x) = 0$ and we will be done.

Notice that $\lim_{x \to c} R_n(x) = f(c) - f(c) = 0$.

(f is differentiable at c meaning it is continuous at c meaning $\lim_{x \to c} f(x) = f(c)$).

We can therefore apply L'Hospital's rule:

$$\lim_{x \to c} r_n(x) = \lim_{x \to c} \frac{f'(x) - \left(f'(c) + f''(c)(x-c) + \frac{f'''(c)}{2!}(x-c)^2 + \cdots\right)}{n(x-c)^{n-1}}$$

$$= \lim_{x \to c} \frac{f'(x) - \sum_{k=1}^{n} \frac{f^{(k)}(c)}{(k-1)!}(x-c)^{k-1}}{n(x-c)^{n-1}}$$

We repeat this (applying L'Hospital's rule a total of n times).

You may wish to check that we can cnotinue to apply L'Hospital's rule (that the limit of the numerator and denominator are always both 0).

After n applications we have:

$$\lim_{x\to c}\frac{f^{(n)}(x)-\dfrac{f^{(n)}(c)}{n!}n!\,(x-c)^0}{n!\,(x-c)^0}=\lim_{x\to c}\frac{f^{(n)}(x)-f^{(n)}(c)}{n!}=\frac{f^{(n)}(c)-f^{(n)}(c)}{n!}=0$$

Using continuity of $f^{(n)}(x)$ at c in the last step.

There is a much more useful version of Taylor's theorem where the remainder carries more information.

$$\text{Taylor's Theorem (Lagrange) 6.1:}$$

$$\text{If } f: I \to \mathbb{R} \text{ is } (n+1) \text{ times differentiable in } I \text{ and } c \in I$$

$$\text{Then } f(x)=\sum_{k=0}^{n}\frac{f^{(n)}(c)}{n!}(x-c)^n+\frac{f^{(n+1)}(\xi)}{(n+1)!}(x-c)^{n+1}$$

$$\text{where } \xi \text{ is strictly between } x \text{ and } c.$$

(ξ is the Greek letter "xi")

We call $\frac{f^{(n+1)}(\xi)}{(n+1)!}(x-c)^{n+1}$ the **Lagrange Form of the Remainder.**

The intuition behind this idea requires some very basic integration and the mean value theorem:

Define the remainder:

$$R_n(x) \coloneqq f(x)-\sum_{k=0}^{n}\frac{f^{(k)}(c)}{k!}(x-c)^k$$

$$=f(x)-\left(f(c)-f'(c)(x-c)-\frac{f''(c)}{2}(x-c)^2-\cdots-\frac{f^{(n)}(c)}{n!}(x-c)^n\right)$$

Then differentiate both sides n times (As this causes most of the terms to go to zero) to get: $R_n^{(n)}(x)=f^{(n)}(x)-f^{(n)}(c)$. We can simplify RHS using MVT.

Using MVT, there is a ξ strictly between x and c such that

$$f^{(n+1)}(\xi)(x-c)=f^{(n)}(x)-f^{(n)}(c)=R_n^{(n)}(x)$$

Then anti-differentiate both sides n times, (easy calculus).

$$R_n(x)=\frac{f^{n+1}(\xi)}{(n+1)!}(x-c)^{n+1}, \text{for some } \xi \text{ strictly between } x \text{ and } c$$

This almost amounts to a proof but not quite. The only real problem is that there are many different anti-derivatives of a particular function (more on this later when we discuss integration) and we haven't justified choosing this particular one. (We also haven't yet discussed anti-derivatives in any capacity).

For a more rigorous approach we will do something slightly different and use Cauchy's GMVT.

pf:

We define:

$$r_n = \frac{f(x) - \sum_{k=0}^{n} \frac{f^{(k)}(c)}{k!}(x-c)^k}{(x-c)^{n+1}}$$

This is slightly different to how we defined r_n in the Peano case, because we are using $(x-c)^{n+1}$ rather than $(x-c)^n$. This is because the form of the Lagrange remainder contains an $(x-c)^{n+1}$.

Then the remainder $R_n = r_n(x-c)^{n+1}$. We wish to show that $r_n = \frac{f^{(n+1)}(\xi)}{(n+1)!}$ for some ξ between x and c (so that the remainder $R_n = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-c)^{n+1}$, the lagrange form).

Since r_n is defined by a function divided by another function, we will use the GMVT. We know that we will find some ξ between x and c, so we want a differentiable function $F(t)$ such that

$$F(c) - F(x) = f(x) - \sum_{k=0}^{n} \frac{f^{(k)}(c)}{k!}(x-c)^k \,(= R_n)$$

And a differentiable function $G(t)$ such that

$$G(c) - G(x) = (x-c)^{n+1}$$

It isn't too difficult to come up with functions which work.

$$F(t) \coloneqq f(x) - \sum_{k=0}^{n} \frac{f^{(k)}(t)}{k!}(x-t)^k, \qquad G(t) \coloneqq (x-t)^{n+1}$$

$$F(c) = f(x) - \sum_{k=0}^{n} \frac{f^{(k)}(c)}{k!}(x-c)^k, \qquad F(x) = f(x) - f(x) - \sum_{k=1}^{n} \frac{f^{(k)}(c)}{k!}(x-x)^k$$

So $F(c) - F(x) = R_n$ as we wanted.

$$G(c) = (x-c)^{n+1}, \qquad G(x) = 0$$

So $G(c) - G(x) = (x-c)^{n+1}$ as we wanted.

This is great news because it means that $r_n = \frac{F(c)-F(x)}{G(c)-G(x)}$ where F and G are both differentiable between c and x. So we can apply the GMVT:

$r_n = \frac{F'(\xi)}{G'(\xi)}$ for some ξ strictly between x and c.

F' and G' are the derivatives of F and G respectively with respect to t (not x). Evaluating these: (F' takes some effort but terms cancel out nicely and you end up with this)

$$F(t) = f(x) - \sum_{k=0}^{n} \frac{f^{(k)}(t)}{k!}(x-t)^k$$

$$\Rightarrow F'(t) = \frac{-f^{(n+1)}(t)}{n!}(x-t)^n$$

(prove that this is $F'(t)$ on the problem sheet)

G' is much easier.

$$G(t) = (x-t)^{n+1} \Rightarrow G'(t) = -(n+1)(x-t)^n$$

$$\Rightarrow r_n = \frac{F'(\xi)}{G'(\xi)} = \frac{-f^{(n+1)}(\xi)}{n!}(x-\xi)^n \div -(n+1)(x-\xi)^n$$

$$= \frac{f^{(n+1)}(\xi)}{(n+1)!}$$

$$\text{Then } R_n = r_n(x-c)^{n+1} = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-c)^{n+1}$$

$$R_n = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-c)^{n+1} \text{ for some } \xi \text{ strictly between } x \text{ and } c$$

This is called the **Lagrange Form of the Remainder**.

We are done!

This remainder tells us the error in a Taylor polynomial approximation, but it does not tell us that the Taylor Series converges to f (unless we can show that $R_n \to 0$ as $n \to \infty$).

If we wish to say that $f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(c)}{k1}(x-c)^k$, we need to show that $\lim_{n\to\infty} R_n = 0$ which we have not yet done. If the remainder does $\to 0$ (on some interval around x_0) we say that f is analytic at x_0.

Radius of Convergence

A power series is any infinite series of the form:

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

A Taylor series is a type of power series where $a_n = \frac{f^{(n)}(c)}{n!}$.

An important question regarding power series is, for which values of x do they converge? We can determine the answer to this question (similarly to what we did for the exponential function) by using the ratio test.

The series will converge when:

$$\lim\left|\frac{a_{n+1}(x-c)^{n+1}}{a_n(x-c)^n}\right| = \lim\left|\frac{a_{n+1}}{a_n}\right||x-c| < 1$$

$$\Leftrightarrow |x-c| < \lim\left|\frac{a_n}{a_{n+1}}\right|$$

We call $R = \lim\left|\frac{a_n}{a_{n+1}}\right|$ the $\mathbf{Radius\ of\ Convergence}$ of the power series. The term "radius" might seem odd as there are no circles or spheres here. The reason for this becomes more apparent once you study complex analysis. Instead of a number line, complex numbers lie on a plane. A power series with complex numbers converges not on some interval of the number line, but on some circle of the number plane, with a centre c, and radius R.

The series converges when $|x-c| < R$ and diverges when $|x-c| > R$. It may either converge or diverge at $x = c \pm R$ (the endpoints).

$R = \infty$ is also okay ($\lim\left|\frac{a_{n+1}}{a_n}\right| = 0$). In this special case the series converges for all $x \in \mathbb{R}$ (this is the case for the e^x series).

A power series will always converge at $x = c$ (to the value a_0). If $\lim\left|\frac{a_{n+1}}{a_n}\right| = \infty$ then $R = 0$ and the power series converges only at $x = c$.

We have been assuming that $\lim\left|\frac{a_{n+1}}{a_n}\right|$ exists (or is infinity). This may not be the case, but it would be nice if we still had a radius of convergence. This motivates the following theorem.

$$\text{Cauchy-Hadamard Theorem (6.2):}$$

$$\text{Every power series has a radius of convergence } R \in [0,\infty]$$

$$\text{Such that the series absolutely converges for } |x-c| < R \text{ and diverges for } |x-c| > R$$

Note that the interval $[0, \infty]$ "includes ∞". ∞ is of course not a real number, but we use it to represent the case where the series converges everywhere.

It may seem that we already have proved this theorem by using the ratio test, but we assumed that $\lim \left|\frac{a_{n+1}}{a_n}\right|$ exists (or is ∞). This may not always be the case.

We can use the stronger version of the ratio to include the cases where the limit does not exist, but this test is only good so long as we don't have $\liminf \left|\frac{a_{n+1}}{a_n}\right| \leq 1 \leq \limsup \left|\frac{a_{n+1}}{a_n}\right|$. If we want something which will work *all* of the time, we ought to use the strong root test.

pf:

We always have convergence at $x = c$, so suppose that $x \neq c$.

If $|a_k|^{\frac{1}{k}} \cdot |x - c|$ (as a sequence of k) is unbounded above, then by the **Strong Root Test**, the series diverges for all x. This case can only occur if $x \neq c$ since if $x = c$, the sequence would be bounded. In this case we have $R = 0$ and convergence only at c.

Otherwise, $\limsup |a_k|^{\frac{1}{k}} \cdot |x - c|$ exists, meaning that $\limsup |a_k|^{\frac{1}{k}}$ exists.

$$\text{Let } a = \limsup |a_k|^{\frac{1}{k}} \geq 0$$

The applying the **Strong Root Test**: (assume for now that $a > 0$)

$$\limsup |a_k(x - c)^k|^{\frac{1}{k}} = a|x - c|$$

$$\text{If } a|x - c| < 1 \left(\Leftrightarrow |x - c| < \frac{1}{a}\right)$$

$$\text{Then } \sum_{k=0}^{\infty} a_k(x - c)^k \text{ converges absolutely}$$

$$\text{If } a|x - c| > 1 \left(\Leftrightarrow |x - c| > \frac{1}{a}\right)$$

$$\text{Then } \sum_{k=0}^{\infty} a_k(x - c)^k \text{ diverges}$$

Edge case, $a = 0$. Then $a|x - c| = 0 < 1$ and we have convergence for all x ($R = \infty$).

We now have the following 3 cases:

$$|a_k|^{\frac{1}{k}} \text{ is unbounded above} \Rightarrow R = 0$$

Otherwise $a = \limsup |a_k|^{\frac{1}{k}}$ exists (and is ≥ 0).

$$a > 0 \Rightarrow R = \frac{1}{a}$$

$$a = 0 \Rightarrow R = \infty$$

We have now proved Cauchy-Hadamard. Power series only converge on an interval centred at c.

Uniform Convergence

The next thing I wanted to talk about was if power series are continuous, and whether we can differentiate them term-wise.

[SPOILER] the answer turns out to be yes to both. In fact there are more general classes of series which are continuous and which can be differentiated term-wise. The proof that power series can be differentiated term-wise actually requires the more general case for continuity (at least the way I will present it), so we will jump straight to the more general case for when an infinite series of functions is continuous.

When we define a function by an infinite series, e.g., $f(x) = \sum_{k=0}^{\infty} a_k x^k$, we are *really* defining it as a limit of partial sums.

$$\text{Let } f_n(x) = \sum_{k=0}^{n} a_k x^k \text{ then } f(x) = \lim f_n(x)$$

So if we want to learn more about series of functions, it might be easier to first consider sequences of functions.

By definition of the limit of a sequence:

If $f_n(x)$ is a sequence of functions with domain X and $f(x)$ is a function with domain X,

$$\lim f_n(x) = f(x) \text{ means}$$

$$\forall_{x\in X}, \forall_{\varepsilon>0}, \qquad \exists_{N\in\mathbb{N}} \text{ s.t.}, \qquad \forall_{n\geq N}, \qquad |f_n(x) - f(x)| < \varepsilon$$

(i.e., $\lim f_n(x) = f(x)$ for every x).

Note that the natural number N may depend on ε and may also depend on x. We often refer to convergence of a sequence of functions as **pointwise convergence**.

Similar to the idea of uniform continuity, we say that the sequence $f_n(x)$ **converges uniformly** to $f(x)$ if N does not need to depend on x (if there is a value of N which works for every x). Formally:

$$f_n(x) \text{ converges uniformly to } f(x) \text{ means}$$

$$\forall_{\varepsilon>0}, \qquad \exists_{N\in\mathbb{N}} \text{ s.t.}, \qquad \forall_{n\geq N}, \forall_{x\in X} \ |f_n(x) - f(x)| < \varepsilon$$

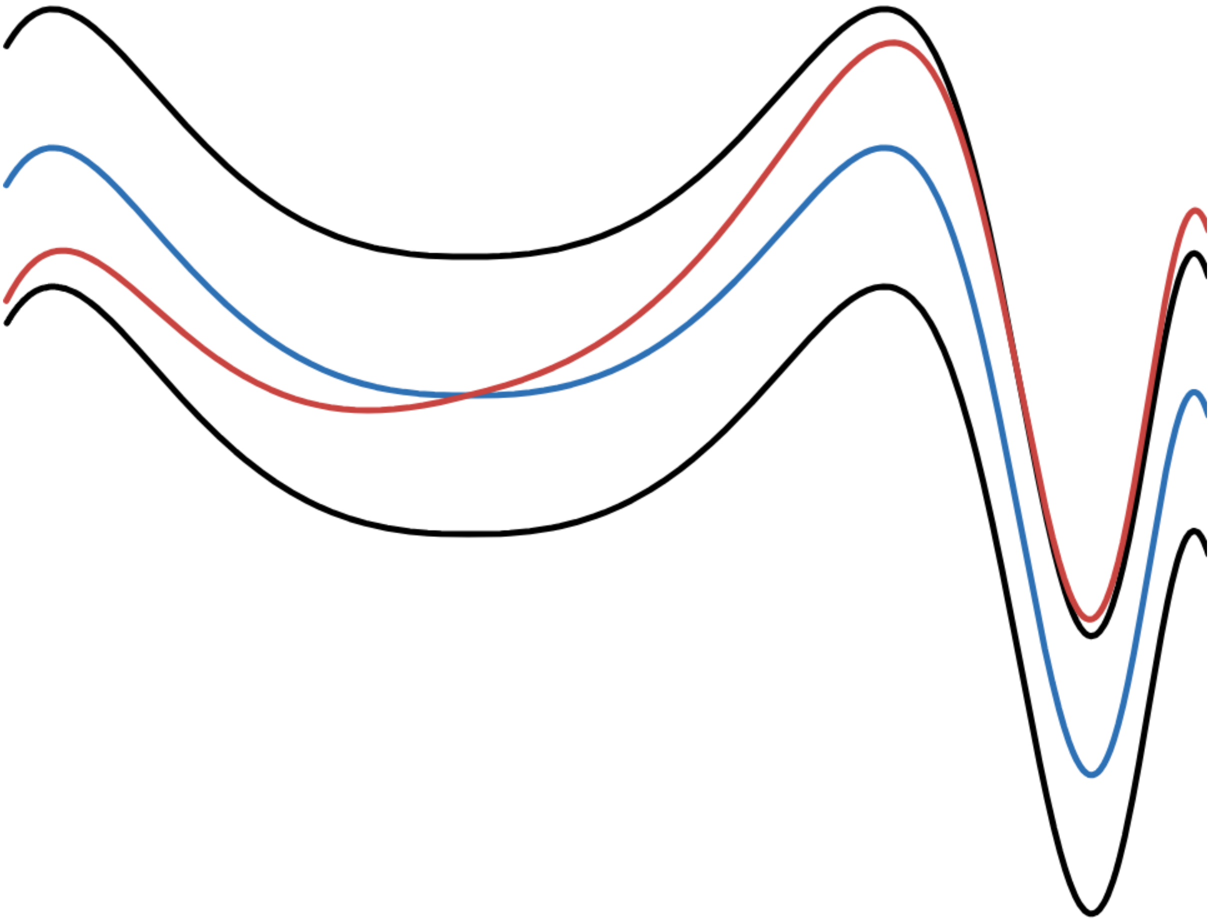
The only difference is that we are now choosing x after we choose N. This value of N works for every x. This definition might seem a bit out of nowhere but its use and motivation will soon become apparent.

Clearly, if a sequence converges **uniformly**, it converges **pointwise**.

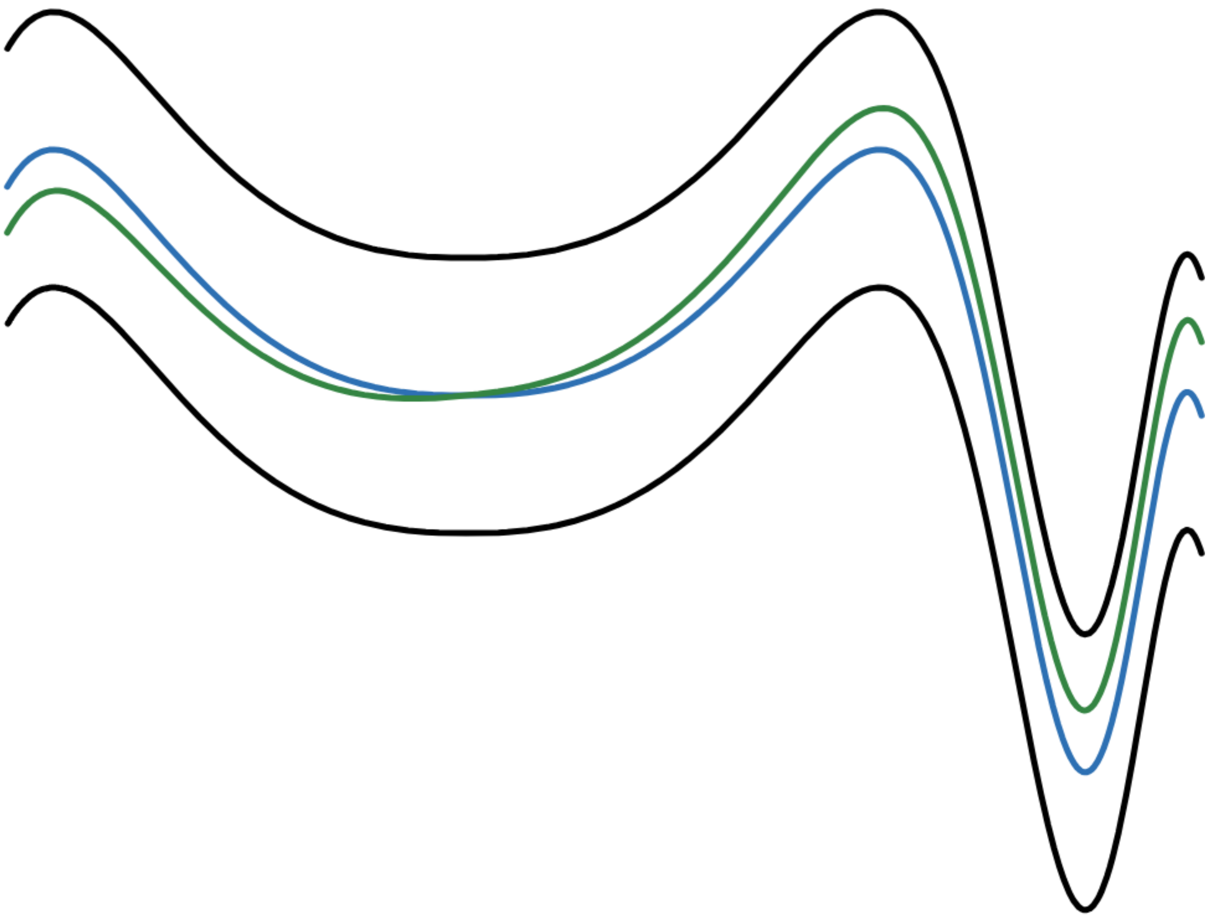
For a more intuitive idea of uniform convergence:

A sequence of functions f_n converges uniformly to f, if one of the functions in our sequence, f_N (and every function after it) is close to f everywhere at once (this N works for every x).

This is a diagram of some arbitrary function f and $\varepsilon = 0.3$:



Here the **blue curve** represents $f(x)$. The **black curves** are $f(x) \pm \varepsilon$. The **red** curve represents the 3$^{\text{rd}}$ term of our sequence $f_3(x)$. Even though the red curve is between the black curves in some places, it isn't between them everywhere. If I choose $N = 10$ however,



The green curve represents $f_{10}(x)$ which lies entirely between the black curves. In this example, the sequence only gets closer to f meaning $N = 10$ will work here (but $N = 3$ would not work).

For a smaller value of ε you may need a larger value of N. The big picture is that for any $\varepsilon > 0$ there is some N such that for each $n \geq N$, $f_n(x)$ lies between the black curves ($f(x) \pm \varepsilon$).

To help this click, I have created an interactive desmos graph:

https://www.desmos.com/calculator/rbni9dupog

Where you can change the values of ε and n. For any $\varepsilon > 0$ you choose, you can make n large enough that the curve of f_n lies entirely between the black curves.

If I have a sequence of functions $f_n(x)$ with $f(x) = \lim f_n(x)$ (a pointwise limit), where each $f_n(x)$ is continuous, is it necessarily the case that $f(x)$ is continuous? It seems like a reasonable thing to suppose but it isn't too difficult to think of a counter example. There are many examples you could come up with, but I think this is the simplest one.

Let $f_n(x) = x^n$ for $x \in [0,1], n \in \mathbb{N}$. Then $f_n(x)$ is continuous.

$$\text{Let } f(x) = \lim f_n(x) = \lim x^n = \begin{cases} 0 & x \in [0,1) \\ 1 & x = 1 \end{cases}$$

Which is clearly discontinous at 1. We need an extra condition in order to ensure that our sequence of continuous functions converges to a continuous limit. I will not yet spoil what this condition is (though you may be able to guess based on the topic of this section), we will figure it out together.

Theorem 6.3 (**Incomplete**):

Let $f_n(x)$ be a convergent sequence of continuous functions with $f(x) \coloneqq \lim f_n(x)$

Then $f(x)$ is continuous.

I write **Incomplete** because as it stands, this is false. We need to figure out what the extra condition is.

pf:

$$\text{Let } \varepsilon > 0, c \in X$$

Definition of limit:

$$\lim f_n(x) = f(x)$$

$$\Rightarrow \exists_{N \in \mathbb{N}} \text{ such that } n \geq N \Rightarrow |f_n(x) - f(x)| < \frac{\varepsilon}{3}$$

Take this N is given.

(The reason for using $\frac{\varepsilon}{3}$ instead of just ε will become apparent).

Continuity:

$$f_n(x) \text{ is continuous at } c$$

$$\Rightarrow \exists_{\delta_n > 0} \text{ such that } |x - c| < \delta_n \Rightarrow |f_n(x) - f_n(c)| < \frac{\varepsilon}{3}$$

(I write δ_n as this value may be different depending on which $f_n(x)$ we are considering).

(This is the definition of continuity assuming c is an interior point.)

Next let $n = N$. Let $\delta = \delta_N$. Suppose $|x - c| < \delta$. We want to show that $|f(x) - f(c)| < \varepsilon$. Using the things which we already know can be made small:

$$|f(x) - f(c)| = |f(x) - f_N(x) + f_N(x) - f_N(c) + f_N(c) - f(c)|$$

$$\leq |f(x) - f_N(x)| + |f_N(x) - f_N(c)| + |f_N(c) - f(c)|$$

(We used $\frac{\varepsilon}{3}$ because here we have 3 things added together, we hope to have each to be $< \frac{\varepsilon}{3}$ so that the sum is $< \varepsilon$).

We kow that $|f(x) - f_N(x)|, |f_N(x) - f_N(c)| < \frac{\varepsilon}{3}$, but what about $|f_N(c) - f(c)|$?

We know that $f_N(x)$ is close to $f(x)$, but remember the value of N depends on x meaning we cannot gaurantee that $f_N(c)$ is close to $f(c)$. We could solve this problem if we knew that there was some N (independent of x) such that $|f_N(x) - f(x)| < \varepsilon$ for every $x \in X$, (including for $x = c$).

This requirement, that N does not depend on x is precisely the idea of uniform convergence. If we have this additional requirement, then we have $|f_N(x) - f(x)| < \frac{\varepsilon}{3}$ and $|f_N(c) - f(c)| < \frac{\varepsilon}{3}$. We still have (by continuity of f_N) $|f_N(x) - f_N(c)| < \frac{\varepsilon}{3}$.

$$\text{Hence } |f(x) - f(c)| < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon$$

Since the definition of continuity at endpoints is (ever so slightly) different, we would need to make a few minor changes to show that this works at the endpoints (I leave this to you).

And we are done!

To state the complete theorem:

Theorem 6.3 (**Complete**):

Let $f_n(x)$ be a **uniformly** convergent sequence of continuous functions where

$$f(x) := \lim f_n(x)$$

Then $f(x)$ is continuous.

i.e., uniform convergence preserves continuity.

We might want to use this to show that power series are continuous.

$$\text{Let } f(x) = \sum_{k=0}^{\infty} a_k x^k \text{ be an arbitrary power series (centred at 0).}$$

$$\text{Then } f(x) \text{ is defined as } \lim f_n(x)$$

$$\text{where } f_n(x) = \sum_{k=0}^{n} a_k x^k \text{ (this is a polynomial hence continuous)}$$

If we can show that $f_n(x)$ converges *uniformly* (this turns out to be be quite true, but things will still work out) then we can use Theorem 6.3 to conclude that $f(x)$ is continuous. Determining whether an infinite series converges uniformly can be difficult, the next section provides a useful test to determine if it does.

Weierstrass M-Test

The Weierstrass M-Test will give us a way to determine if a series converges uniformly.

The simplest case of a series converging uniformly (meaning N does not depend on x) is a sequence of constant functions, $f_n(x) = M_n$. Each M_n depends on n but not on x, it is a sequence of numbers.

The idea of the M-test is to compare our series of functions $\sum_{k=0}^{\infty} f_k(x)$ against some convergent series of constants, $\sum_{k=0}^{\infty} M_k$ (using the comparison test). It seems reasonable that the choice of N which works for $\sum M_k$ (which is independent of x) will also work for $\sum f_k(x)$, meaning we have a choice of N which is independent of x which works for $\sum f_k(x)$, so we have uniform convergence of the series. I will now state this idea more precisely and then will prove it:

Theorem 6.4 (**Weierstrass M-Test**):

If $f_k(x)$ is a sequence of functions with domain X

and there is a sequence M_k such that $|f_k(x)| \leq M_k \ \forall_{x \in X}$ and $\sum_{k=0}^{\infty} M_k$ converges

$$\text{Then } \sum_{k=0}^{\infty} f_k(x) \text{ converges uniformly (and absolutely)}$$

We use $|f_k(x)| \leq M_k$ because we need both sequences to be ≥ 0 in order to use the comparison test.

pf:

$$\text{By the } \textbf{Comparison Test}, (0 \leq |f_k(x)| \leq M_k), \qquad \sum_{k=0}^{\infty} f_k(x) \text{ converges (absolutely)}$$

$$\text{We can therefore let } f(x) = \sum_{k=0}^{\infty} f_k(x)$$

$$\text{Let } \varepsilon > 0$$

[Our goal is to find an N independent of x such that $\forall_{n \geq N}$, we have $|f_n(x) - f(x)| < \varepsilon$]

$$\text{Let } L = \sum_{k=0}^{\infty} M_k$$

We know that there exists an $N \in \mathbb{N}$ such that $n \geq N \Rightarrow |\sum_{k=0}^{n} M_k - L| < \varepsilon$

$$\Rightarrow \left|\sum_{k=0}^{n} M_k - L\right| = \left|L - \sum_{k=0}^{n} M_k\right| = \left|\sum_{k=0}^{\infty} M_k - \sum_{k=0}^{n} M_k\right| = \left|\sum_{k=n+1}^{\infty} M_k\right| = \sum_{k=n+1}^{\infty} M_k < \varepsilon$$

Each $M_k \geq 0$ so we can get rid of the modulus signs at the end.

This N is of course independent of x, we suspect that this N will also work for the sequence of partial sums, $\sum_{k=0}^{n} f_k(x)$.

$$\text{Let } n \geq N$$

We finally wish to show that $\left|\sum_{k=0}^{n} f_k(x) - f(x)\right| < \varepsilon$ and then we are done.

Proving that this is indeed the case is not too bad:

$$\left|\sum_{k=0}^{n} f_k(x) - f(x)\right| = \left|f(x) - \sum_{k=0}^{n} f_k(x)\right| = \left|\sum_{k=0}^{\infty} f_k(x) - \sum_{k=0}^{n} f_k(x)\right|$$

$$= \left|\sum_{k=n+1}^{\infty} f_k(x)\right| \leq \sum_{k=n+1}^{\infty} |f_k(x)| \text{ (triangle inequality)}$$

$$\leq \sum_{k=n+1}^{\infty} M_k \text{ (comparison test)}$$

$$< \varepsilon$$

And we are done!

The motivation for this test was to show that power series converge uniformly (and are hence continuous).

$$\text{Let } f(x) \coloneqq \sum_{k=0}^{\infty} a_k (x-c)^k$$

Be a power series. We will assume (for now) that this power series converges on a closed interval, $(c-R, c+R)$ with $R \neq \infty$.

$$\text{Let } f_k(x) \coloneqq a_k(x-c)^k, \qquad \text{for } x \in [c-R, c+R]$$

Then $|x-c| \leq R$. Then we apply the Weierstrass M-Test.

$$|f_k(x)| = |a_k||x-c|^k \leq |a_k|R^k$$

$$\text{So let } M_k = |a_k|R^k$$

Finally, we need to show that $\sum_{k=0}^{\infty} M_k = \sum_{k=0}^{\infty} |a_k|R^k$ converges.

This is just $\sum_{k=0}^{\infty}|a_k(x-c)^k|$ but with $x-c=R$. We know that power series converge absolutely within their radius of convergence, and $c+R$ is in the radius of convergence, $[c-R, c+R]$, meaning $\sum_{k=0}^{\infty} a_k(x-c)^k$converges absolutely at $x=c+R$, i.e., $\sum_{k=0}^{\infty} a_k R^k$ converges absolutely and hence $\sum_{k=0}^{\infty}|a_k|R^k$ converges.

We have shown that $\sum_{k=0}^{\infty} M_k$ converges, hence the original power series converges uniformly.

There are two small problems with this proof. Namely the assumptions that $R \neq \infty$ (that the power series does not converge everywhere) and that the interval of convergence is closed (the endpoints converge). If these are not the case, we cannot conclude that the power series converges uniformly, but we can cheat slightly (don't worry it's still rigorous).

Call the interval where the series converges I (may be an infinite interval). Let r be a real number with $0 \leq r < R$ (if $R=\infty$ then r can be any non-negative real number).

Then we definitely have absolute convergence at the endpoints, $c \pm r$.

Restrict the domain of $f_k(x)$ to $[c-r, c+r]$ (we know we have absolute convergence everywhere on this interval, since this interval is entirely contained within I).

$$\text{Finally, Let } M_k = |a_k|r^k$$

$$\Rightarrow\ |f_k(x)| = |a_k||x-c|^k \leq |a_k|r^k = M_k$$

$$\text{This time we can guarantee that } \sum_{k=0}^{\infty} M_k = \sum_{k=0}^{\infty}|a_k|r^k \text{ converges}$$

$$\text{since } c+r \in [c-r, c+r]$$

So then by the Weierstrass M-Test, $\sum_{k=0}^{\infty} f_k(x)$ converges uniformly on $[c-r, c+r]$.

We now have:

Theorem 6.5:

A power series converges uniformly on any closed interval

which lies within its interval of convergence

[IMPORTANT: The restriction of the domain to $[c-r, c+r]$ is absolutely necessary! Power series do NOT generally uniformly converge on their interval of convergence, but they do converge uniformly on any closed interval which lies within the interval of convergence. If the original interval is already closed at both ends (meaning it must also be finite) then this extra step is not necessary.]

We are now ready to prove that power series are continuous.

Theorem 6.6:

Power series are continuous.

Let $\sum_{k=0}^{\infty} a_k(x-c)^k$ be a power series with interval of convergence I.

To show that f is continuous we show that it is continuous at every point in I.

Let $x_0 \in I$. We now show continuity at x_0. Let $r = |x_0 - c|$ (then $[c-r, c+r] \subseteq I$).

By theorem 6.5, the power series converges uniformly on $[c-r, c+r]$.

Each $\sum_{k=0}^{n} a_k(x-c)^k$ is a polynomial and is hence continuous.

Then (since uniform convergence preserves continuity) $\sum_{k=0}^{\infty} a_k(x-c)^k$ is continuous on $[c-r, c+r]$. In particular it is contnuous at $x_0 = c + r$ as required.

Term-wise Differentiation of Power Series!

I put an exclamation mark in the title of this section because I am excited for this one. This is one of those facts which is often taken for granted. We have now put a lot more effort than may have been anticipated in getting to the point where we are finally ready (more or less) to prove that power series can be term-wise differentiated (term by term).

That is:

$$\frac{\mathrm{d}}{\mathrm{d}x}\sum_{k=0}^{\infty} a_k x^k = \sum_{k=1}^{\infty} k a_k x^{k-1}$$

(the $k = 0$ term vanishes when we differentiate)

Before I proceed there is one small useful fact we will need. The motivation for this fact will become clear once we do the proof for term-wise differentiation.

At one point of the proof we will be required to factorise an $(x - y)$ out of expressions of the form $x^n - y^n$,

For $n = 1$, this is trivial: $x - y = (x - y)$

$n = 2$ is easy by using the difference of squares: $x^2 - y^2 = (x + y)(x - y)$

$n = 3$ is a bit harder. We want to write $x^3 - y^3 = \text{something} \times (x - y)$. We probably want this "something" to have an x^2 and y^2 term (so that we get an $x^3 - y^3$ when we expand. Trying $x^2 + y^2$:

$$(x^2 + y^2)(x - y) = x^3 - y^3 + xy^2 - x^2y$$

Which is close but we have these extras $xy^2 - x^2y$ terms at the end. If we factorise these terms:

$$(x^2 + y^2)(x - y) = x^3 - y^3 - xy(x - y)$$

$$\Rightarrow (x^2 + xy + y^2)(x - y) = x^3 - y^3$$

So the correct answer is $x^2 + xy + y^2$ (we we could have also just guessed with some more attempts).

You might start to notice a pattern here and guess that

$$x^4 - y^4 = (x^3 + x^2y + xy^2 + y^3)(x - y)$$

Indeed, this is correct. We might then make a general guess that:

Lemma 6.7:

$$x^n - y^n = (x - y)(x^{n-1} + x^{n-2}y + x^{n-3}y^2 + \cdots + xy^{n-2} + y^{n-1})$$

$$= (x - y)\sum_{k=0}^{n-1} x^{n-1-k}y^k \text{, for } n \in \mathbb{N}$$

pf:

We can use induction, the method of differences or geometric series. This is the method of differences approach:

$$(x-y)\sum_{k=0}^{n-1} x^{n-1-k}y^k = \sum_{k=0}^{n-1} x^{n-1-k}y^k(x-y)$$

$$= \sum_{k=0}^{n-1}(x^{n-k}y^k - x^{n-1-k}y^{k+1})$$

$$= (x^n - x^{n-1}y) + (x^{n-1}y - x^{n-2}y^2) + \cdots + (xy^{n-1} - y^n)$$

$$= x^n + (-x^{n-1}y + x^{n-1}y) + (-x^{n-2}y^2 + x^{n-2}y^2) + \cdots + (-xy^{n-1} + xy^{n-1}) - y^n$$

$$= x^n - y^n$$

Try proving this yourself by induction, or geometric series.

Now for the main event! We want to show that:

Theorem 6.8 (Differentiation of Power Series)

If $f(x) = \sum_{k=0}^{\infty} a_k(x-x_0)^k$ has radius of convergence R

Then $f(x)$ is differentiable, with $f'(x) = \sum_{k=1}^{\infty} ka_k(x-x_0)^{k-1}$

also with radius of convergence R.

We call $\sum_{k=1}^{\infty} ka_k(x-x_0)^{k-1}$ the "formal derivative" of f. Our goal is to show that the formal derivative is the actual derivative.

I will denote the formal derivative as $h(x)$. We aim to show that $f'(x) = h(x)$.

pf:

For now, we assume that $x_0 = 0$ (this makes our job easier, and the case for general x_0 will follow quickly).

$$f(x) = \sum_{k=0}^{\infty} a_k x^k$$

It is easy to show using the strong root test that $\sum_{k=1}^{\infty} ka_k x^{k-1}$ has the same radius of convergence as $f(x)$ (problem sheet). Preservation of convergence at endpoints is not guaranteed.

$$\text{Let } I \text{ be the interval of convergence of } f(x), \text{then let } c \in I \text{ be an interior point}$$

$$(c \text{ is not an endpoint})$$

We need to show that $\lim_{x \to c} \frac{f(x)-f(c)}{x-c}$ exists and is equal to $\sum_{k=0}^{\infty} ka_k c^{k-1}$.

$$\text{Let } g(x) = \frac{f(x) - f(c)}{x - c} \text{ (for } x \neq c)$$

$$g(x) = \frac{1}{x - c} \sum_{k=0}^{\infty} a_k (x^k - c^k)$$

$$= \frac{1}{x - c} \sum_{k=1}^{\infty} a_k (x^k - c^k)$$

(The $k = 0$ term is $a_0(x^0 - c^0) = a_0(1 - 1) = 0$)

This is where we apply Lemma 6.7:

$$x^k - c^k = (x - c) \sum_{i=0}^{k-1} x^{k-1-i} c^i, \qquad \text{for } k \geq 1$$

So that the $(x - c)'$s cancel out.

$$g(x) = \sum_{k=1}^{\infty} a_k \sum_{i=0}^{k-1} x^{k-1-i} c^i$$

But g is so far only defined for $x \neq c$. I then set

$$g(c) = \sum_{k=1}^{\infty} a_k \sum_{i=0}^{k-1} c^{k-1-i} c^i$$

So that $g(x) = \sum_{k=1}^{\infty} a_k \sum_{i=0}^{k-1} x^{k-1-i} c^i$ for every x.

$$\text{Then } g(c) = \sum_{k=1}^{\infty} a_k \sum_{i=0}^{k-1} c^{k-1-i} c^i = \sum_{k=1}^{\infty} a_k \sum_{i=0}^{k-1} c^{k-1} = \sum_{k=1}^{\infty} a_k c^{k-1} \sum_{i=0}^{k-1} 1 = \sum_{k=1}^{\infty} ka_k c^{k-1}$$

$$[= h(c)] \text{ (remember, } h \text{ represents the "formal derivative")}$$

Which is precisely what we hoped for $f'(c)$ to be!

But we are not quite done yet. $f'(c) = \lim_{x \to c} \frac{f(x)-f(c)}{x-c} = \lim_{x \to c} g(x)$.

We need to show that $\lim_{x \to c} g(x) = g(c)$, that is we need to show that g is continuous at c. Then we will be done.

$$g(x) = \sum_{k=1}^{\infty} a_k \sum_{i=0}^{k-1} x^{k-1-i} c^i$$

This is a series of functions. Each partial sum is continuous.

We can guarantee that the limit g is continuous if the series converges uniformly. We will use the Weierstrass M-Test for this. When we saw that power series are continuous, we had to restrict the domain to some closed interval, so here I do the same thing.

As an aside, we know that $g(x)\left(= \sum_{k=1}^{\infty} a_k \sum_{i=0}^{k-1} x^{k-1-i} c^i\right)$ converges on $(-R, R)$ because, (for $x \neq c$) f converges on $(-R, R)$ and $g(x) = \frac{f(x)-f(c)}{x-c}$, hence g converges. For $x = c$, $g(c) = h(c)$. We know that h (the formal derivate) converges for $x \in (-R, R)$ which includes $x = c$ (since c is an interior point) so the series converges at c.

Next, we restrict g to the interval $[-c, c]$ (which is fine since $[-c, c] \subseteq (-R, R)$).

Finally, we are ready to apply the M-Test.

$$g_k(x) = a_k \sum_{i=0}^{k-1} x^{k-1-i} c^i \text{, } x \in [-c, c] \text{ (note } |x| \leq |c|\text{)}$$

$$\text{Then } |g_k(x)| = |a_k| \left| \sum_{i=0}^{k-1} x^{k-1-i} c^i \right|$$

$$\leq |a_k| \sum_{i=0}^{k-1} \left| x^{k-1-i} c^i \right| = |a_k| \sum_{i=0}^{k-1} |x|^{k-1-i} |c|^i$$

$$\leq |a_k| \sum_{i=0}^{k-1} |c|^{k-1-i} |c|^i = |a_k| \sum_{i=0}^{k-1} |c|^{k-1}$$

$$= k|a_k||c|^{k-1}$$

$$\text{So let } M_k = k|a_k||c|^{k-1}$$

$$\text{Then } \sum_{k=1}^{\infty} M_k = \sum_{k=1}^{\infty} k|a_k||c|^{k-1} = \sum_{k=1}^{\infty} k|a_k||c|^{k-1}$$

$$\sum_{k=1}^{\infty} ka_k c^{k-1} \text{ is the formal derivative evaluated at } c$$

c is in the interval of convergence of the formal derivative meaning this series converges absolutely, i.e.,

$$\sum_{k=1}^{\infty} M_k = \sum_{k=1}^{\infty} k|a_k||c|^{k-1} \text{ converges}$$

By the Weierstrass M-Test, we have shown that

$$\sum_{k=1}^{\infty} a_k \sum_{i=0}^{k-1} x^{k-1-i} c^i \text{ converges uniformly}$$

We know that each term of this series is continuous at c. Hence g is continuous at c, which means

$$f'(c) = \lim_{x \to c} \frac{f(x) - f(c)}{x - c} = \lim_{x \to c} g(x) = g(c) = \sum_{k=0}^{\infty} ka_k c^{k-1}$$

i.e.,

$$f'(x) = \sum_{k=0}^{\infty} ka_k x^{k-1}$$

For the case where $x_0 \neq 0$:

$$f(x) = \sum_{k=0}^{\infty} a_k (x - x_0)^k \Rightarrow f(x + x_0) = \sum_{k=0}^{\infty} a_k x^k \Rightarrow f'(x + x_0) = \sum_{k=0}^{\infty} ka_k x^{k-1}$$

$$\Rightarrow f'(x) = \sum_{k=0}^{\infty} ka_k (x - x_0)^{k-1}$$

And We Are Done!

That took way more effort than we would have liked but we have now shown that power series are term-wise differentiable!

The (arguably) hardest part of this chapter is now over. The rest should be easy in comparison.

Power Series are Unique

Clearly if I have two sequences $a_k = b_k$ (two names for the same sequence) then

$$\sum_{k=0}^{\infty} a_k (x-c)^k = \sum_{k=0}^{\infty} b_k (x-c)^k$$

The converse is not so obvious. If $\sum_{k=0}^{\infty} a_k (x-c)^k = \sum_{k=0}^{\infty} b_k (x-c)^k$, does this imply that $a_k = b_k$?

Different sums often add to the same thing, for example $2+2=4$ and $1+3=4$. Two different calculations give the same result. Is it possible that two different power series result in the same function? The answer is no!

$$\text{Let } f(x) = \sum_{k=0}^{\infty} a_k (x-c)^k, \qquad g(x) = \sum_{k=0}^{\infty} b_k (x-c)^k$$

$$\text{With } f(x) = g(x)$$

Then $f(c) = a_0$

Then since we can differentiate term-wise,

$$f'(x) = \sum_{k=1}^{\infty} k a_k (x-c)^{k-1}, \qquad \text{and so } f'(c) = a_1$$

Differentiating again, $f''(c) = 2a_2,\ f'''(c) = 3!\, a_3, \cdots f^{(k)}(c) = k!\, a_k$ (you can prove this by induction, see problem sheet Q39).

So, we have $a_k = \frac{f^{(k)}(c)}{k!}$. Similarly, $b_k = \frac{g^{(k)}(c)}{k!}$.

$$f(x) = g(x) \Rightarrow \frac{f^{(k)}(c)}{k!} = \frac{g^{(k)}(c)}{k!} \Rightarrow a_k = b_k$$

And we are done.

$$\text{Notice also that } f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(c)}{k!}(x-c)^k$$

$$\text{Is the taylor series of } f \text{ centred at } c$$

This means that if a function f can be represented by a power series (which is not always the case) then the power series is the Taylor series (which we can find by differentiating).

Trigonometric Functions and π

We are finally ready to define the sine and cosine functions (and the famous constant π). We could have done this before, but without being able to term-wise differentiate it would have been much more difficult to do anything with these functions.

Similarly to how we defined the exponential function, we define the trig function by power series (intuition comes from their Maclaurin series).

$$\sin(x) := \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!} \left(= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots \right)$$

$$\cos(x) := \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!} \left(= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots \right)$$

$$[\text{We are of course using radians here.} \, \pi \text{ radians} = 180°]$$

We can use the ratio test to show that these converge for all x:

$$\sin x \colon \lim \left| \frac{(-1)^{n+1} x^{2n+3}}{(2n+3)!} \div \frac{(-1)^n x^{2n+1}}{(2n+1)!} \right| = \lim \left| \frac{x^2}{(2n+3)(2n+2)} \right|$$

$$= x^2 \lim \frac{1}{(2n+3)(2n+2)} = 0 < 1$$

Cosine is similar.

We can easily evaluate $\sin 0 = 0$ and $\cos 0 = 1$.

It is clear from the definition that sine is odd $(\sin(-x) = -\sin x)$ and cosine is even $(\cos(-x) = \cos x)$.

We can also differentiate term-wise:

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

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

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

We would like to prove some of the famous trig identities. Importantly, $\sin^2 x + \cos^2 x = 1$, $\sin(x+y) = \sin x \cos y + \sin y \cos x$, and $\cos(x+y) = \cos x \cos y - \sin x \sin y$. The rest are easy enough to derive from these alone so we will focus on these.

These seem intimidating at first, but we can use a clever trick. First, we wish to show that $\sin^2 x + \cos^2 x = 1$ for every x. That is, we want to show that $\sin^2 x + \cos^2 x$ is constant with value 1.

$$\text{Let } f(x) = \sin^2 x + \cos^2 x$$

$$f'(x) = 2 \sin x \cos x + 2 \cos x\,(-\sin x)$$

$$= 2 \sin x \cos x - 2 \sin x \cos x = 0$$

Since $f'(x) = 0$ everywhere, $f(x)$ must be constant. Also $f(0) = \sin^2 0 + \cos^2 0 = 1$.

Hence $f(x) = 1$ for every x.

We can use a similar tactic to prove the other identities. For $\sin(x + y) = \sin x \cos y + \sin y \cos x$, let's try:

$$\text{Fix } y \text{ and let } f(x) = \sin(x + y) - \sin x \cos y - \sin y \cos x$$

$$f'(x) = \cos(x + y) - \cos x \cos y + \sin y \sin x$$

Since we don't yet know that $\cos(x + y) = \cos x \cos y - \sin x \sin y$, we cannot conclude that $f'(x) = 0$. We need to try something else.

We want to define a function f which is equal to $\sin(x + y)$ for some input, is equal to $\sin x \cos y + \sin y \cos x$ for another input, and with a derivative of zero everywhere (meaning it is constant). After some tinkering, we might come up with:

$$\text{fix } x, y$$

$$\text{Let } f(t) \coloneqq \sin(x + t) \cos(y - t) + \sin(y - t) \cos(x + t)$$

$$\text{Then } f(0) = \sin x \cos y + \sin y \cos x\,, \qquad f(y) = \sin(x + y)$$

Where x and y are fixed. Hopefully $f'(t)$ will be 0.

$$f'(t)$$

$$= \cos(x + t) \cos(y - t) + \sin(x + t) \sin(y - t)$$

$$- \cos(y - t) \cos(x + t) - \sin(y - t) \sin(x + t)$$

$$= 0$$

Therefore $f(t)$ is constant so $f(y) = f(0) \Rightarrow \sin(x + y) = \sin x \cos y + \sin y \cos x$.

We could try something similar for $\cos(x + y)$, but we can make things easier by using the result for $\sin(x + y)$.

$$\text{fix } y, \qquad \sin(x + y) = \sin x \cos y + \cos x \sin y$$

$$\text{Differentiate with respect to } x\text{:}$$

$$\cos(x+y) = \cos x \cos y - \sin x \sin y$$

Another important fact about the sine and cosine is that they are bounded. $|\sin x| \leq 1$ and $|\cos x| \leq 1$.

$$\text{We know that} \sin^2 x + \cos^2 x = 1$$

$$\Rightarrow \sin^2 x = 1 - \cos^2 x \leq 1$$

$$\Rightarrow |\sin x| \leq 1$$

Similarly, $\cos^2 x = 1 - \sin^2 x \leq 1 \Rightarrow |\cos x| \leq 1$.

A very important property of the trig functions is that they are periodic, meaning they repeat. In particular $\sin x = \sin(x + 2\pi)$ and $\cos x = \cos(x + 2\pi)$ for every x. Currently we have no chance of proving this because we haven't even defined π yet!

We might define π (Greek letter pi) as the smallest positive real number where $\sin \pi = 0$. This can be done but it will be slightly easier to define π as the smallest positive real number such that $\cos \frac{\pi}{2} = 0$.

$$\text{Theorem 6.9 (existence of } \pi)$$

$$\text{There exists smallest a real number } \pi > 0 \text{ such that} \cos \frac{\pi}{2} = 0$$

We expect that $\pi \approx 3.14, \frac{\pi}{2} \approx 1.57 < 2$. The idea is to show that $\cos 0 > 0$ and $\cos 2 < 0$ and then use the intermediate value theorem.

pf:

$$\text{Cosine is defined by a power series, hence is a continuous function}$$

$$[\text{since power series are continuous}]$$

$$\cos 0 = 1 > 0$$

$$\cos 2 = \sum_{n=0}^{\infty} \frac{(-1)^n 2^{2n}}{(2n)!} = \sum_{n=0}^{\infty} \frac{(-1)^n 4^n}{(2n)!}$$

We originally proved convergence using the ratio test. We could also use the alternating series test, which here will be more useful as it tells us something about the sign of the answer. Clearly this is an alternating series.

Also $\frac{4^n}{(2n)!}$ is decreasing (for $n \geq 1$). A quick proof of this fact:

$$\frac{4^n}{(2n)!} = \frac{(2n+1)(2n+2)}{4} \times \frac{4^{n+1}}{(2n+2)!} \geq \frac{4^{n+1}}{(2n+2)!}$$

(since $(2n+1)(2n+2) \geq 4$ for $n \geq 1$)

[Each term is $\geq$ the next term hence decreasing]

By the alternating series test, this series converges (which we already knew). Recalling how the proof of the alternating series test went, we split the series into increasing and decreasing subsequences of partial sums. The subsequence of the partial sums which end in a positive term was a decreasing subsequence, and moreover each element of this subsequence was an upper bound for the alternating series.

Considering the partial sum consisting of the first 3 terms:

$$\sum_{n=0}^{2} \frac{(-1)^n 4^n}{(2n)!} = 1 - \frac{4}{2!} + \frac{4^2}{4!} = 1 - 2 + \frac{2}{3} = -\frac{1}{3}$$

This partial sum ends in a positive term $\left(\frac{4^2}{4!}\right)$ hence is $-\frac{1}{3}$ is an upper bound for the sum.

$$\cos 2 \leq -\frac{1}{3} < 0$$

We now apply the intermediate value theorem, there exists an $x \in (0,2)$ such that $\cos x = 0$.

We still need to show that there is a *smallest* positive x with this property. To find the smallest number with a certain property, the *infimum* of some set might be a good thing to consider.

$$\text{Let } S = \{x > 0 : \cos x = 0\}$$

[The set of positive zeroes of the cosine function.]

We already showed that there is at least one $x > 0$ with $\cos x = 0$ meaning S is non-empty.

S is also bounded below (by definition, 0 is a lower bound for S).

Therefore $\alpha = \inf S$ exists. Since 0 is a lower bound for S, $\alpha \geq 0$. There exists a sequence $x_n \in S$ with $\lim x_n = \alpha$.

By definition $\cos x_n = 0$ (since $x_n \in S \Rightarrow \cos x_n = 0$), meaning $\cos x_n$ is a constant sequence with $\lim \cos x_n = 0$.

Since cosine is continuous, $0 = \lim \cos x_n = \cos \lim x_n = \cos \alpha$.

We know $\alpha \geq 0$. If $\alpha = 0$ then $\cos \alpha = \cos 0 = 0$, but we know that $\cos 0 = 1$, meaning $\alpha > 0$.

Since α is the infimum, $0 < x < \alpha \Rightarrow \cos x \neq 0$ meaning α is indeed the smallest positive real number with $\cos \alpha = 0$. We can then define $\pi = 2\alpha$. We are done!

We can then use our trig identities to evaluate sine and cosine at various multiple of π.

$$\sin\frac{\pi}{2} = 1 - \cos^2\frac{\pi}{2} = 1 - 0^2 = 1$$

$$\sin\pi = \sin\left(\frac{\pi}{2} + \frac{\pi}{2}\right) = \sin\frac{\pi}{2}\cos\frac{\pi}{2} + \sin\frac{\pi}{2}\cos\frac{\pi}{2} = 0$$

$$\cos\pi = \cos\frac{\pi}{2}\cos\frac{\pi}{2} - \sin\frac{\pi}{2}\sin\frac{\pi}{2} = -1$$

And so on, $\sin\frac{3\pi}{2} = -1, \sin 2\pi = 0, \cos\frac{3\pi}{2} = 0, \cos 2\pi = 1$

sine and cosine each repeat every 2π:

$$\sin(x + 2\pi) = \sin x\cos 2\pi + \sin 2\pi\cos x = \sin x$$

$$\cos(x + 2\pi) = \cos x\cos 2\pi - \sin x\sin 2\pi = \cos x$$

We can also find out where sine and cosine are positive/negative and where they are increasing/decreasing so that we might get a better idea of what their graphs look like.

Also note that since $\cos 0 = 1$ and $\cos x$ is continuous and $\frac{\pi}{2}$ is the smallest positive real number with $\cos\frac{\pi}{2} = 0$, it must be the case that $\cos x > 0$ for $x \in \left[0, \frac{\pi}{2}\right)$ since otherwise we could use the intermedaite value theorem to find an $x \in \left(0, \frac{\pi}{2}\right)$ with $\cos x = 0$ (which is not possible since $\frac{\pi}{2}$ is the smallest which works).

Then since cosine is even, $\cos x > 0$ for $x \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ and $\cos\left(-\frac{\pi}{2}\right) = \cos\frac{\pi}{2} = 0$.

This also means that $\sin x$ is strictly increasing on $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ (its derivative is $\cos x > 0$).

Then since $\sin 0 = 0$ and it is strictly incresing on $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, we must have $\sin x > 0$ for $x \in \left(0, \frac{\pi}{2}\right)$ and $\sin x < 0$ for $x \in \left(-\frac{\pi}{2}, 0\right)$.

Also sine is just a shifted version of cosine:

$$\sin\left(x + \frac{\pi}{2}\right) = \sin x\cos\frac{\pi}{2} + \sin\frac{\pi}{2}\cos x = \cos x$$

Which means

$$\cos x > 0 \text{ for } x \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \Longrightarrow \sin x > 0 \text{ for } x \in (0, \pi)$$

Then since sine is an odd function we must have $\sin x < 0$ for $x \in (-\pi, 0)$.

Then again using that cosine is a shifted version of sine: $\cos x < 0$ for $x \in \left(-\frac{3\pi}{2}, -\frac{\pi}{2}\right)$, in particular, for $x \in \left(-\pi, -\frac{\pi}{2}\right)$.

So far we have:

$\sin x < 0$ for $x \in (-\pi, 0)$ meaning $\cos x$ is increasing for $x \in (-\pi, 0)$.

$\sin x > 0$ for $x \in (0, \pi)$ meaning $\cos x$ is decreasing for $x \in (0, \pi)$.

$\cos x > 0$ for $x \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ meaning $\sin x$ is increasing for $x \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$.

$\cos x < 0$ for $x \in \left(-\pi, -\frac{\pi}{2}\right) \cup \left(\frac{\pi}{2}, \pi\right)$ meaning $\sin x$ is decreasing on $\left(-\pi, -\frac{\pi}{2}\right) \cup \left(\frac{\pi}{2}, \pi\right)$.

$\sin x = 0$ at $x = 0, \pm\pi$ so $\cos x$ has stationary points for these values.

$\cos x = 0$ at $x = \pm\frac{\pi}{2}$ so $\sin x$ as stationary points for these values.

There are other multiples of π where the trig functions have nice values.

$$\text{Namely } \frac{\pi}{3}, \frac{\pi}{4}, \frac{\pi}{6}$$

Using the $\cos(x+y)$ identity setting $x = y$ we get $\cos 2x = \cos^2 x - \sin^2 x$. Then using $\sin^2 x = 1 - \cos^2 x$, we get $\cos 2x = 2\cos^2 x - 1$. Now set $x = \frac{\pi}{4}$,

$$\cos\frac{\pi}{2} = 0 = 2\cos^2\frac{\pi}{4} - 1$$

$$\Rightarrow \cos^2\frac{\pi}{4} = \frac{1}{2}$$

$$\Rightarrow \cos\frac{\pi}{4} = \frac{1}{\sqrt{2}}$$

We take the positive root since $\frac{\pi}{4} \in \left(0, \frac{\pi}{2}\right)$ and cosine is positive on this interval.

$\sin\frac{\pi}{4} = \sqrt{1 - \cos^2\frac{\pi}{4}} = \frac{1}{\sqrt{2}}$ (sine is also positive on $\left(0, \frac{\pi}{2}\right)$).

We can also find triple angle formulae:

$$\sin 3x = \sin(2x + x) = \sin 2x \cos x + \sin x \cos 2x = 2\sin x \cos^2 x + \sin x \cos^2 x - \sin^3 x$$

$$= 3\sin x\cos^2 x - \sin^3 x$$

Then Let $x = \frac{\pi}{3}$:

$$\sin\pi = 0 = 3\sin\frac{\pi}{3}\cos^2\frac{\pi}{3} - \sin^3\frac{\pi}{3}$$

Dividing by $\sin\frac{\pi}{3} (\neq 0)$:

$$0 = 3\cos^2\frac{\pi}{3} - \sin^2\frac{\pi}{3}$$

Using $\sin^2 x = 1 - \cos^2 x$:

$$\cos^2\frac{\pi}{3} = \frac{1}{4}$$

Since $\frac{\pi}{3} \in \left(0, \frac{\pi}{2}\right)$, $\cos\frac{\pi}{3} > 0$:

$$\cos\frac{\pi}{3} = \frac{1}{2}$$

$\sin\frac{\pi}{3} > 0$ also:

$$\sin\frac{\pi}{3} = \sqrt{1 - \left(\frac{1}{2}\right)^2} = \frac{\sqrt{3}}{2}$$

We can then also find sine and cosine of $\frac{\pi}{6}$:

$$\sin\frac{\pi}{6} = \sin\left(\frac{\pi}{2} - \frac{\pi}{3}\right) = \sin\frac{\pi}{2}\cos\frac{\pi}{3} - \sin\frac{\pi}{3}\cos\frac{\pi}{2} = \cos\frac{\pi}{3} = \frac{1}{2}$$

$$\cos\frac{\pi}{6} = \cos\left(\frac{\pi}{2} - \frac{\pi}{3}\right) = \cos\frac{\pi}{2}\cos\frac{\pi}{3} + \sin\frac{\pi}{2}\sin\frac{\pi}{3} = \sin\frac{\pi}{3} = \frac{\sqrt{3}}{2}$$

We now know all of the basic facts about sine and cosine. We can also define:

$$\tan x = \frac{\sin x}{\cos x}, \qquad \sec x = \frac{1}{\cos x}, \qquad \csc x = \frac{1}{\sin x}, \qquad \cot x = \frac{\cos x}{\sin x}$$

Whenever these denominators are non-zero.

A Power Series for the Logarithm

The exponential function, sine and cosine, all have known power series (as they were defined by their power series), but the logarithm does not.

The best we have is its Taylor series centred at 1:

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

Which only converges for x between 0 and 2 by using the ratio test (we need to check endpoints separately). The series has convergence for $x = 2$ (alternating sign test) but divergence at $x = 0$ (harmoic series).

The series only converges for $x \in (0,2]$. (This won't end up being too much of a problem thankfully).

Even where it does converge, we don't yet know that it converges to $\log x$.

Showing that the Taylor series converges to $\log x$ is equivalent to showing that the remainder → 0 as $n \to \infty$. We could attempt to do this directly by using the Lagrange Remainder but this turns out to be very difficult (I will explain why later).

Thankfully there exists a much easier way to prove that this power series converges to $\log x$. First I will prove a *very small* Lemma.

Lemma 6.10:

If $f'(x) = g'(x)$ then $f(x) = g(x) + c$ where c is some arbitrary constant.

pf:

Let $h(x) = f(x) - g(x)$. Then $h'(x) = f'(x) - g'(x) = 0$

Therefore h is a constant function, $h(x) = c$

$$\Rightarrow f(x) - g(x) = c \Rightarrow f(x) = g(x) + c$$

Theorem 6.11 (Taylor series of the logarithm):

$$\log x = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n}(x-1)^n \text{ for } x \in (0,2]$$

pf:

$$\text{Let } f(x) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n}(x-1)^n, \qquad x \in (0,2]$$

$$\Rightarrow f'(x) = \sum_{n=1}^{\infty}(-1)^{n-1}(x-1)^{n-1} = \sum_{n=0}^{\infty}(-1)^n(x-1)^n = \sum_{n=0}^{\infty}(1-x)^n$$

This is a geometric series. It converges for $|1-x| < 1 \Leftrightarrow x \in (0,2)$, so assume for now that $x \neq 2$.

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

We know that $\log' x = \frac{1}{x}$ so by Lemma 6.10,

$$f(x) = \log x + c$$

$$f(1) = 0 \text{ and } \log(1) = 0 \Rightarrow c = 0$$

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

For $x = 2$ we can simply use the fact that log is continuous and power series are continuous, then take the limit $x \to 2$ on both sides:

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

$$\log 2 = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots$$

Another way to handle $x = 2$ is to use the **Lagrange Remainder** (which is harder, but I think is more fun). I said before that using the Lagrange Remainder approach is very difficult (which it is for general $x \in (0,2]$), but if we only care about $x = 2$, it isn't so bad.

Recall:

$$R_n = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-1)^{n+1} \text{ for some } \xi \text{ strictly between } x \text{ and } 1$$

The $(n+1)^{\text{th}}$ derivative of $\log x$ is $f^{(n+1)}(x) = \frac{(-1)^n n!}{x^{n+1}}$.

$$R_n = \frac{(-1)^n n!}{(n+1)!\,\xi^{n+1}}(x-1)^{n+1} = \frac{(-1)^n}{(n+1)}\left(\frac{x-1}{\xi}\right)^{n+1}$$

Also $x = 2$:

$$R_n = \frac{(-1)^n}{(n+1)\xi^{n+1}}, \text{for some } \xi \in (1,2)$$

$$|R_n| = \frac{1}{(n+1)\xi^{n+1}} \to 0 \text{ Hence } R_n \to 0$$

And we are done!

The fact that this works at $x = 2$ means $\log 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \cdots$, though this converges rather slowly. The first 10 terms are correct to only 1 decimal place.

Dealing with the lagrange remainder for $x > 1$ is relatively easy (you can show that $\left|\frac{x-1}{\xi}\right| < 1$ giving a geometric sequence $\to 0$. It is a bit harder, but not too hard, to show that $\left|\frac{x-1}{\xi}\right| < 1$ in the case that $x \geq \frac{1}{2}$. For $0 < x < \frac{1}{2}$ it becomes much more difficult to prove (mainly because it isn't true for general $\xi \in (x, 1)$, though it it true for the specific value chosen by the Lagrange Remainder Theorem). This difficulty is why we avoided the Lagrange remainder for the whole proof, though you may wish to show that $R_n \to 0$ for $x \geq \frac{1}{2}$ using the Lagrange remainder directly as this is not too bad (problem sheet).

What should we do in the case that we wish to calculate $\log x$ for some $x > 2$? In this case our taylor series diverges. We could create a new taylor series centred somewhere else, $x = c$ but then we would need to know what $\log c$ is.

A cleaner approach is to notice that $x > 2 \Rightarrow \frac{1}{x} \in (0,2]$ and then we can use our log laws:

$$\log x = -\log x^{-1} = -\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n}\left(\frac{1}{x} - 1\right)^n$$

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

$$\text{E.g.}, \log 4 = \sum_{n=1}^{\infty} \frac{1}{n}\left(\frac{3}{4}\right)^n$$

$$= \frac{3}{4} + \frac{9}{32} + \frac{9}{64} + \cdots$$

For $x \in [0.5,2]$ we also have $\frac{1}{x} \in (0,2]$ meaning you may use either version. In such cases, the first version $\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n}(x-1)^n \left(= \sum_{n=1}^{\infty} \frac{-1}{n}(1-x)^n\right)$ typically converges faster for $x < 1$, and the second version $\sum_{n=1}^{\infty} \frac{1}{n}\left(1 - \frac{1}{x}\right)^n$ typically converges faster for $x > 1$.

This means we can also use the second version to write:

$$\log 2 = \sum_{n=1}^{\infty} \frac{1}{n \cdot 2^n}$$

The first 10 terms are correct to 3 decimal places

The rest of this section is just an interesting aside.

An even faster converging series can be obtained by considering,

$$\log 2 = \log\left(\frac{4}{3} \div \frac{2}{3}\right) = \log\left(\frac{4}{3}\right) - \log\left(\frac{2}{3}\right)$$

$$= \sum_{n=1}^{\infty} \frac{1}{n \cdot 4^n} + \sum_{n=1}^{\infty} \frac{1}{n \cdot 3^n}$$

$$\left(\text{using the second version for} \log \frac{4}{3} \text{ and the first version for} \log \frac{2}{3}\right)$$

$$\log 2 = \sum_{n=1}^{\infty} \frac{1}{n}\left[\frac{1}{4^n} + \frac{1}{3^n}\right]$$

Which converges *even* more quickly to $\log 2$. The first 10 terms give an answer correct to 5 decimal places. (Though perhaps this is cheating as each term of the series requires adding 2 things together).

We can also determine arbitrarily precise lower and upper bounds on $\log 2$ (or any other logarithm we like). We could use the alternating series to obtain bounds, but it is easier to use the second version: First compute as many terms as you like, I will do 10.

$$\log 2 = \frac{44711}{64512} + \sum_{n=11}^{\infty} \frac{1}{n \cdot 2^n}$$

$$\frac{1}{n \cdot 2^n} \le \frac{1}{11} \frac{1}{2^n} \text{ for } n \ge 11 \text{ (with equality only when } n = 11)$$

$$\Rightarrow \sum_{n=11}^{\infty} \frac{1}{n \cdot 2^n} < \frac{1}{11} \sum_{n=11}^{\infty} \frac{1}{2^n} = \frac{1}{11 \cdot 2^{10}} = \frac{1}{11264}$$

$$\Rightarrow \frac{44711}{64512} < \frac{44711}{64512} + \sum_{n=11}^{\infty} \frac{1}{n \cdot 2^n} < \frac{44711}{64512} + \frac{1}{11264} = \frac{122971}{177408}$$

$$\text{So } \frac{44711}{64512} < \log 2 < \frac{122971}{177408}$$

rounding $\frac{44711}{64512}$ down and rounding $\frac{122971}{177408}$ up:

$$0.69306 < \log 2 < 0.69316$$

so $\log 2 \approx 0.693$ is correct to 3 decimal places

In general, choose some large N and $\sum_{n=1}^{N} \frac{1}{n2^n} < \log 2 < \sum_{n=1}^{N} \frac{1}{n2^n} + \frac{1}{(N+1)2^N}$

Integration

Step Functions

The idea of an integral can be thought of in many ways, but perhaps the easiest to consider is that of area under a curve. We have not defined "area" in any sense and so we must be more precise. Before we can be precise we will be intuitive.

Defining the area under an arbitrary curve seems like a difficult task. We can make the task easier by first considering curves consisting only of rectangles, we (roughly speaking) call functions of this kind "step functions". Then we can easily define the integral of step functions (by adding the rectangle areas together).

If we want to define the integral of an arbitrary function f, we find a sequence of step functions converging to f, and define the integral of f as the limit of the integrals of the step functions.

There are a few details which we will need to iron out here, but this idea is called "regulated integration".

Integrals are often defined by using the approach of "Riemann Integration", defining the integral as a limit of "Riemann Sums". There are some functions which are Riemann Integrable but are not regulated integrable, but these are not worth worrying about. There are more general definitions such as "Lebesgue Integration" which we will also not be touching here. I chose Regulated integration because it is probably the easiest to teach (and it relates to sequences of functions which is an idea we have already built up) and so it is a good choice for a first course in real analysis.

For now we will only consider functions on closed intervals. Extending the definition to infinite intervals will be fairly easy (see the section on improper integrals).

First, we define step functions.

$$\text{Def: Step Function}$$

$$f\colon [a,b] \to \mathbb{R} \text{ is a step function if there exists a partition}$$

$$a = x_0 < \cdots < x_n = b$$

$$\text{such that } f \text{ is constant on each interval } (x_k, x_{k+1}) \text{ for each integer } 0 \le k \le n-1\text{:}$$

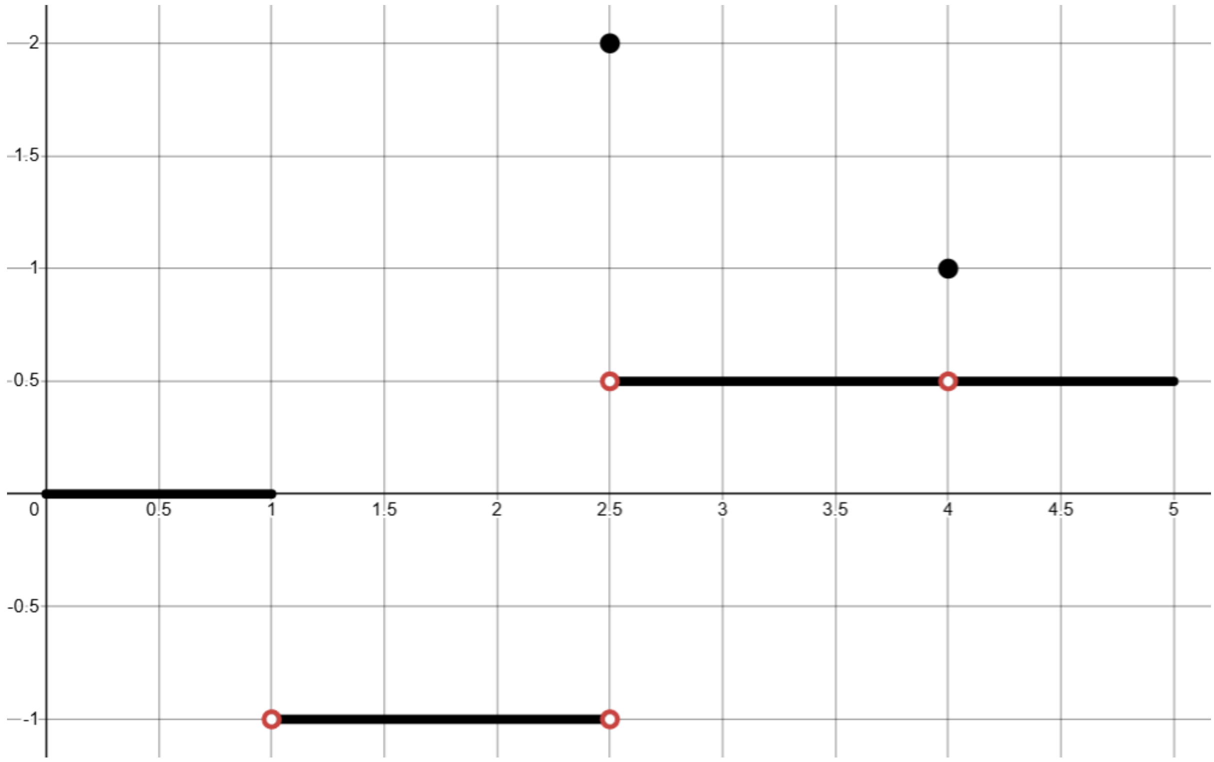
We might also consider a function defined only at a single point to be a step function.

"Partition" just means an increasing (finite) sequence of $n+1$ numbers, begging with $x_0 = a$, ending with $x_n = b$. The term "partition" may be used to refer to the endpoints x_k, or the intervals (x_k, x_{k+1}) between them.

A common refinement of two partitions P and Q is a partition which includes every endpoint from P and every endpoint from Q (you can always use the union of P and Q).

The partition splits the interval $[a, b]$ into n subintervals $(x_0, x_1), (x_1, x_2), \cdots (x_{n-1}, x_n)$. Notice that these are all open intervals, meaning the endpoints of each subinterval (which includes a and b) can be whatever you like. This makes sense in the context of defining integrals because we would expect that altering or removing individual points would not change the area.

E.g., (graph of a step function):



[The red circles represent gaps.]

For this example our partition could be $a = 0 < 1 < 2.5 < 4 < 5 = b$ because the function is constant on $(0,1), (1,2.5), (2.5,4), (4,5)$. Note that partitions are not unique, I could also use $a = 0 < 0.5 < 1 < 1.5 < 2 < 2.5 < 3 < 3.5 < 4 < 4.5 < 5 = b$. This is a valid partition, because the function is constant on each of those intervals.

Let $f: [a, b] \to \mathbb{R}$ be a step function with partition $a = x_0 < \cdots < x_n = b$. For each integer $0 \leq k \leq n - 1$, choose some $x_k^* \in (x_k, x_{k+1})$. We define the integral of f as:

$$\text{Def: } \int_a^b f(x)\mathrm{d}x := \sum_{k=0}^{n-1} (x_{k+1} - x_k) f(x_k^*)$$

$(x_{k+1} - x_k)$ represents the width of the rectangle, $f(x_k^*)$ is the height. Multiplying these and adding them together gives us the total area (which is the idea). If a rectangle is below the graph $f(x_k^*) < 0$ meaning it contributes a negative area, for this reason we call it a "signed area".

If our step function is of the kind defined only at a single point a, then this definition does not work. In this case we define $\int_a^a f(x)\mathrm{d}x \coloneqq 0$.

Before proceeding we need to make sure that this definition is coherent.

We choose an arbitrary $x_k^* \in (x_k, x_{k+1})$. Since f is constant on this interval (by definition of step functions), our choice does not affect the value of $f(x_k^*)$.

I also mentioned that the choice of partition for a step function is not unique. It is easy to see, from the intuitive notion of adding rectangle areas, that the choice of partition does not matter.

A more rigorous proof is possible (and conceptually very simple):

Let P and Q be any two partitions for f, we want to show that they give the same integral.

Let R be a common refinement of P and Q (e.g., the union). Then for each interval, (x_k, x_{k+1}) of P, there are endpoints in R, $(y_i, y_j \text{ with } i < j)$ such that $y_i = x_k, y_j = x_{k+1}$, meaning $x_{k+1} - x_k = y_j - y_i = (y_j - y_{j-1}) + (y_{j-1} - y_{j-2}) + \cdots + (y_{i+1} - y_i)$.

f is constant on (x_k, x_{k+1}), so it is constant on each of $(y_i, y_{i+1}), \cdots (y_{j-1}, y_j)$.

$$(x_{k+1} - x_k)f(x_k^*) = [(y_j - y_{j-1}) + \cdots + (y_{i+1} - y_i)]f(x_k^*)$$

$$= (y_j - y_{j-1})f(x_k^*) + \cdots + (y_{i+1} - y_i)f(x_k^*)$$

$$= (y_j - y_{j-1})f(y_{j-1}^*) + \cdots + (y_{i+1} - y_i)f(y_i^*)$$

Where $y_i^* \in (y_i, y_{i+1}), \cdots, y_{j-1}^* \in (y_{j-1}, y_j)$ are arbitrary points in their respective intervals (f is constant on all of these intervals so the choices of $y_i^*, \cdots y_{j-1}^*$ do not matter).

Suppose P has n endpoints, and R has m endpoints.

$$\text{Then } \sum_{k=0}^{n-1} (x_{k+1} - x_k)f(x_k^*) = \sum_{i=0}^{m-1} (y_{i+1} - y_i)f(y_i^*)$$

So, the integral given by P equals the integral given by R. By the same argument, the integral given by Q equal the integral given by R. Hence the integral given by P equals the integral given by Q. This means the choice of partition does not matter.

Regulated Functions

I will not define what a regulated function is just yet. We will proceed with the plan, and then the motivation for the definition should become clear.

We want to define $\int_a^b f(x)\mathrm{d}x$ (to save space I will use the notation $I(f) = \int_a^b f(x)\mathrm{d}x$. Suppose there is a sequence of step functions f_n which converges to f. Then we wish to define $I(f) = \lim I(f_n)$, (that is $I(\lim f_n) = \lim I(f_n)$).

For this definition to be any good, we need to check that $\lim I(f_n)$ always converges. We also said that f_n is some sequence of step functions with $\lim f_n = f$. There may well be multiple sequences of step functions which converge to f, we need to be sure that $\lim I(f_n)$ converges to *the same thing* regardless of which sequence we choose.

It turns out that these two conditions are not always met, meaning the definition is no good. Luckily there is a small extra condition we can impose which makes it work. We will proceed blindly for now, and the extra condition will become apparent during the proof.

First, we wish to show that $\lim I(f_n)$ always converges. We will use an ε-N proof.

pf:

Since $\lim f_n = f$:

$$\text{Let } \varepsilon > 0$$

$$\text{There exists } N \in \mathbb{N} \text{ such that } n \geq N \Rightarrow |f_n(x) - f(x)| < \varepsilon$$

(The N chosen may depend on x as we have only pointwise convergence).

We wish to show that $I(f_n)$ converges, but we have no candidate limit. Luckily, we do have a way to show convergence without a candidate limit, that is by showing the sequence to be Cauchy.

$$\text{Let } n, m \geq N$$

(So that $|f_n(x) - f(x)|$ and $|f_m(x) - f(x)| < \varepsilon$)

We wish to show that $|I(f_n) - I(f_m)| < \varepsilon$ as this is what it means for a sequence to be Cauhcy. f_n and f_m are step functions each with some number of partitions. Let $a = x_0 < \cdots < x_M = b$ be a common refinement of these partitions. (Meaning it is a partition of both f_n and of f_m).

$$|I(f_n) - I(f_m)| = \left| \sum_{k=0}^{M-1} (x_{k+1} - x_k) f_n(x_k^*) - \sum_{k=0}^{M-1} (x_{k+1} - x_k) f_m(x_k^*) \right|$$

$$= \left| \sum_{k=0}^{M-1} (x_{k+1} - x_k)\big(f_n(x_k^*) - f_m(x_k^*)\big) \right|$$

$$\leq \sum_{k=0}^{M-1} |x_{k+1} - x_k||f_n(x_k^*) - f_m(x_k^*)|$$

We know that $|f_n(x) - f_m(x)| = |f_n(x) - f(x) + f(x) - f_m(x)| \leq |f_n(x) - f(x)| + |f(x) - f_m(x)| = |f_n(x) - f(x)| + |f_m(x) - f(x)| < 2\varepsilon$, but this only works for this particular value of x (because N depends on x). If we had uniform convergence then there is a value of N which works for every x, (including x_k^*) meaning we have $|f_n(x_k^*) - f_m(x_k^*)| < 2\varepsilon$.

Also $x_k < x_{k+1} \Rightarrow |x_{k+1} - x_k| = x_{k+1} - x_k$.

$$\text{Hence } |I(f_n) - I(f_m)| \leq \sum_{k=0}^{M-1} |x_{k+1} - x_k||f_n(x_k^*) - f_m(x_k^*)|$$

$$< 2\varepsilon \sum_{k=0}^{M-1} (x_{k+1} - x_k) = 2\varepsilon(b - a)$$

(We could have chosen to make $|f_n(x) - f(x)| < \frac{\varepsilon}{2(b-a)}$ so that this final inequality would read $|I(f_n) - I(f_m)| < \varepsilon$).

And we are done!

To write the same proof more concisely (now that we know what the extra condition is):

$$\text{Let } \varepsilon > 0, \qquad \text{then } \frac{\varepsilon}{2(b-a)} > 0 \text{ so}$$

$$\exists_{N \in \mathbb{N}} \text{ s.t.,} \qquad \forall_{x \in [a,b], n \geq N}, \qquad |f_n(x) - f(x)| < \frac{\varepsilon}{2(b-a)}$$

$$\text{Let } n, m \geq N \Rightarrow |f_n(x_k^*) - f_m(x_k^*)| \leq |f_n(x_k^*) - f(x_k^*)| + |f_m(x_k^*) - f(x_k^*)|$$

$$< \frac{\varepsilon}{2(b-a)} + \frac{\varepsilon}{2(b-a)} = \frac{\varepsilon}{b-a}$$

$$\text{and } |I(f_n) - I(f_m)| = \left| \sum_{k=0}^{M-1} (x_{k+1} - x_k)\big(f_n(x_k^*) - f_m(x_k^*)\big) \right|$$

$$\leq \sum_{k=0}^{M-1} |x_{k+1} - x_k||f_n(x_k^*) - f_m(x_k^*)| < \frac{\varepsilon}{b-a} \sum_{k=0}^{M-1} (x_{k+1} - x_k) = \frac{\varepsilon}{b-a}(b-a) = \varepsilon$$

Hence $I(f_n)$ is cauchy and is therefore convergent and we are done!

We are only certain that the limits exists if the convergence of the step functions is uniform. From now on we will assume uniform convergence. We also need to prove the second condition for the definition to make sense,

$$\text{If } f_n \text{ and } g_n \text{ are both sequences of step functions which converge uniformly to } f$$

$$\text{Then} \lim I(f_n) = \lim I(g_n)$$

We already know that both limits exist, we just need to show that they are equal. One way to do this is to show that $\lim[I(f_n) - I(g_n)] = 0$

pf:

This proof is almost identical to the last one.

$$\text{Let } \varepsilon > 0$$

Using uniform convergence to f:

$$\text{Let } N \text{ be such that } \forall_{x\in[a,b],n\geq N}, \qquad |f_n(x) - f(x)| < \varepsilon \text{ and } |g_n(x) - f(x)| < \frac{\varepsilon}{2(b-a)}$$

(We are using $\frac{\varepsilon}{2(b-a)}$, as we did before, because it works out the same.)

We wish to show that $|I(f_n) - I(g_n)| < \varepsilon$.

As before, let $a = x_0 < \cdots < x_M = b$ be a common refinement of the partitions of f_n and g_n (so it is both a partition of f_n and of g_n).

$$|I(f_n) - I(g_n)| = \left|\sum_{k=0}^{M-1}(x_{k+1} - x_k)f_n(x_k^*) - \sum_{k=0}^{M-1}(x_{k+1} - x_k)g_n(x_k^*)\right|$$

$$= \left|\sum_{k=0}^{M-1}(x_{k+1} - x_k)\big(f_n(x_k^*) - g_n(x_k^*)\big)\right|$$

$$\leq \sum_{k=0}^{M-1}(x_{k+1} - x_k)|f_n(x_k^*) - g_n(x_k^*)|$$

We know that $|f_n(x_k^*) - g_n(x_k^*)| \leq |f_n(x_k^*) - f(x)| + |g_n(x_k^*) - f(x)| < \frac{\varepsilon}{b-a}$

$$|I(f_n) - I(g_n)| < \frac{\varepsilon}{b-a}\sum_{k=0}^{M-1}(x_{k+1} - x_k) = \frac{\varepsilon}{b-a}(b-a) = \varepsilon$$

$$\text{Hence} \lim\big(I(f_n) - I(g_n)\big) = 0 \ (\text{and} \lim I(f_n), \lim I(g_n) \ \text{exist})$$

$$\Rightarrow \lim\big(I(f_n)\big) = \lim\big(I(g_n)\big)$$

And we are done!

So, our definition is good:

If there exists a sequence of step functions f_n which converges uniformly to f on $[a, b]$ then,

$$\int_a^b f(x)\mathrm{d}x = \lim \int_a^b f_n(x)\mathrm{d}x$$

We now have our motivation to define a regulated function:

$$\text{Def: we say that } f \text{ is regulated iff}$$

$$\text{there exists a sequence of step functions which converges uniformly to } f$$

So we can integrate a function if it is regulated.

Some important observations can then be made:

If f_n, g_n are step functions and $\alpha, \beta \in \mathbb{R}$ then $\alpha f_n + \beta g_n$ is a step function. $f_n g_n$ is a step function. Also $|f_n|$ is a step function. These are relatively obvious statements.

Then if f and g are regulated functions, (meaning there are sequences of step functions f_n, g_n which converge uniformly to f and g respectively) then $\alpha f + \beta g = \lim(\alpha f_n + \beta g_n)$ meaning $\alpha f + \beta g$ is regulated also (it easy to show that $\lim(\alpha f_n + \beta\, g_n)$ converges uniformly). Similarly, $f(x)g(x)$ is regulated and $|f(x)|$ is regulated (problem sheet).

Typically, we talk about whether a function is “Riemann Integrable” or “Lebesgue integrable” etc.,

In our formulation, our definition for integration applies iff our function is regulated, so I may interchange the terms “integrable” and “regulated” (even though in other contexts, e.g., Riemann integration, they do not mean the same thing).

The idea of a regulated function is that it can be approximated arbitrarily well by step functions, so we can reformulate regulated functions in this way:

$$\text{Theorem 7.1:}$$

$$f(x): [a, b] \to \mathbb{R} \text{ is regulated iff}$$

$$\forall_{\varepsilon>0} \text{ there exists a step function } g: [a, b] \to \mathbb{R} \text{ such that } |f(x) - g(x)| < \varepsilon\ \forall_{x\in[a,b]}$$

i.e., iff you can find step functions which approximate f arbitrarily well.

pf:

" $\Rightarrow$ "

Let f be regulated, then there exists a sequence of step functions $f_n \to f$ uniformly. Let $\varepsilon > 0$, there is an N such that $n \geq N \Rightarrow |f(x) - f_n(x)| < \varepsilon \, \forall_{x \in [a,b]}$. $g(x) = f_N(x)$ works.

" $\Leftarrow$ "

Let $\varepsilon_n = \frac{1}{n}$. Then there is a step function $g_n(x)$ such that $|f(x) - g_n(x)| < \frac{1}{n}$ for all $x \in [a, b]$.

Let $\varepsilon > 0$, there exists an N such that $N > \frac{1}{\varepsilon}$ (archimedes).

Let $n \geq N \Rightarrow |f(x) - g_n(x)| < \frac{1}{n} \leq \frac{1}{N} < \varepsilon$. Hence $g_n \to f$ uniformly and so f is regulated.

We are done!

We might not expect every function to be integrable (by which I mean regulated) e.g., the famous Dirichlet function:

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

But most functions we would be interested in in practical situations we would expect to be regulated. E.g., all continuous functions.

Theroem 7.2:

Continuous Functions are Regulated

pf:

Let f be continuous on $[a, b]$

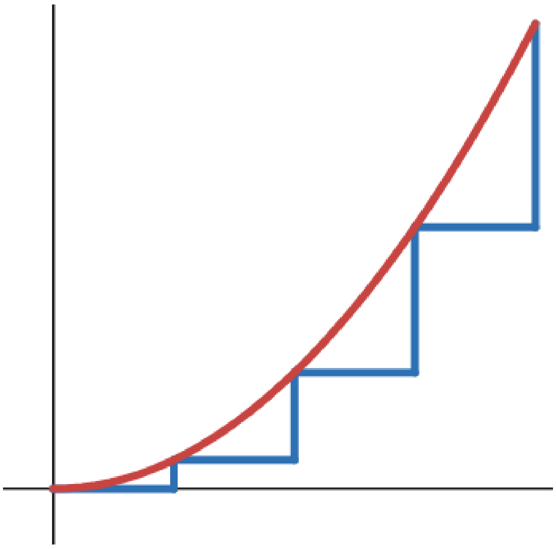
Since f is continuous on a closed interval, it is uniformly continuous on that interval.

(Theorem 4.2)

Let $\varepsilon > 0$

There exists $\delta > 0$ such that $\forall_{x,y \in [a,b]}, \quad |x - y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon$

The idea is to create a step function where each interval is small ($< \delta$) and equal in length, for this we need $n > \frac{b-a}{\delta}$. We also want the endpoints to touch f:



Hopefully, if we make the interval small enough, we can show that our step function is close to f everywhere and then use Theorem 7.1.

Let $n > \frac{b-a}{\delta}$ be a natural number. Define the partition, $x_k = a + (b-a)\frac{k}{n}$ for $0 \leq k \leq n$ (this is a partition where each interval is equal in length). Each partition has length $\frac{b-a}{n} < \delta$. Then define $g(x) = \begin{cases} f(x_k) & x \in [x_k, x_{k+1}) \\ f(b) & x = x_n \end{cases}$.

We finally wish to show that $|f(x) - g(x)| < \varepsilon$ for every x.

If $x = x_n = b$, then $|f(x) - g(x)| = |f(b) - f(b)| = 0 < \varepsilon$. Otherwise $x \in [x_k, x_{k+1})$ for some k. Relabel x as x_k^* accordingly.

$$|f(x) - g(x)| = |f(x_k^*) - g(x_k^*)| = |f(x_k^*) - f(x_k)|$$

Since we know that $x_k^* \in [x_k, x_{k+1})$, we know that $|x_k^* - x_k| < |x_{k+1} - x_k| = \frac{b-a}{n} < \delta$, hence (by uniform continuity of f), $|f(x_k^*) - f(x_k)| < \varepsilon$.

$$\Rightarrow |f(x) - g(x)| < \varepsilon$$

So by Theorem 7.1, f is regulated and we are done!

This is excellent news as it means we can integrate any continuous function.

Of course you can have functions which are not continuous but are regulated (e.g., any non-constant step function).

The Mean Value Theorem (for integrals)

We have seen the mean value theorem, if f is differentiable on $[a, b]$ there exists some $c \in (a, b)$ with $f'(c) = \frac{f(b)-f(a)}{b-a}$. $f'(c)$ is essentially the mean value of the derivative. There is another version of this theorem which says:

(MVT):

If f is continuous on $[a, b]$ then there exists a $c \in [a, b]$ with

$$f(c) = \frac{1}{b-a} \int_a^b f(x)\mathrm{d}x$$

$f(c)$ is essentially the mean value of function f over the interval $[a, b]$.

We already know that $\int_a^b f(x)\mathrm{d}x$ exists (continuous functions are regulated).

Before we can prove this theorem, we have some other facts of integration which we need to uncover.

Theorem 7.3: Linearity of Integration:

If $\alpha, \beta \in \mathbb{R}$ and f and g are regulated on $[a, b]$

$$\text{Then} \int_a^b \big(\alpha f(x) + \beta g(x)\big)\mathrm{d}x = \alpha \int_a^b f(x)\mathrm{d}x + \beta \int_a^b f(x)\mathrm{d}x$$

pf:

We know that $\alpha f + \beta g, f, g$ are regulated so all integrals here exist automatically.

$$f(x) = \lim f_n(x), \qquad g(x) = \lim g_n(x) \text{ where these are uniform limits.}$$

$$\alpha I(f_n) + \beta I(g_n) = \alpha \sum_{k=0}^{M-1} (x_{k+1} - x_k) f_n(x_k^*) + \beta \sum_{k=0}^{M-1} (x_{k+1} - x_k) g_n(x_k^*)$$

$$= \sum_{k=0}^{M-1} (x_{k+1} - x_k)\big(\alpha f_n(x_k^*) + \beta g_n(x_k^*)\big)$$

$$= I(\alpha f_n + \beta g_n)$$

And $\alpha f + \beta g = \lim(\alpha f_n + \beta g_n)$ where this is a uniform limit.

Hence $I(\alpha f + \beta g) = \lim I(\alpha f_n + \beta g_n) = \lim\big(\alpha I(f_n) + \beta I(g_n)\big) = \alpha \lim I(f_n) + \beta \lim I(g_n) = \alpha I(f) + \beta I(g)$

$$\text{i.e.,} \qquad \alpha \int_a^b f(x) + \beta \int_a^b g(x)\mathrm{d}x = \int_a^b \big(\alpha f(x) + \beta g(x)\big)\mathrm{d}x \text{ as required}$$

We also have monotonicity:

Theorem: 7.4:

$$f(x) \geq 0 \ \forall_{x \in [a,b]} \Rightarrow \int_a^b f(x)\mathrm{d}x \geq 0$$

pf:

Let $f(x) \geq 0 \ \forall_{x \in [a,b]}$

Then there exists a sequence of step functions g_n which converges uniformly to f. Then $|g_n|$ converges uniformly to $|f| = f$ ($|f| = f$ because $f(x) \geq 0$). Let $f_n(x) = |g_n(x)| \geq 0$. Then f_n is a sequence of step functions which converges uniformly to f and $f_n(x) \geq 0$ for every x and every n.

$$\int_a^b f(x)\mathrm{d}x = \lim I(f_n) = \lim_{n \to \infty} \sum_{k=0}^{M-1} (x_{k+1} - x_k) f_n(x_k^*)$$

but $f_n(x_k^*) \geq 0$ and $x_{k+1} - x_k > 0$ for every k

$$\Rightarrow \sum_{k=0}^{M-1} (x_{k+1} - x_k) f_n(x_k^*) \geq 0$$

$$\Rightarrow \int_a^b f(x)\mathrm{d}x \geq 0$$

Corollary 7.5:

If $f(x) \geq g(x)$ for every $x \in [a,b]$

Then $\int_a^b f(x)\mathrm{d}x \geq \int_a^b g(x)\mathrm{d}x$

pf:

$f(x) - g(x) \geq 0$ for every $x \in [a,b]$

$$\Rightarrow \int_a^b \big(f(x) - g(x)\big)\mathrm{d}x = \int_a^b f(x)\mathrm{d}x - \int_a^b g(x)\mathrm{d}x \geq 0$$

$$\Rightarrow \int_a^b f(x)\mathrm{d}x \geq \int_a^b g(x)\mathrm{d}x$$

There is also the triangle inequality for integrals:

Theorem 7.6 (Triangle inequality for integrals):

$$\left|\int_a^b f(x)\mathrm{d}x\right| \leq \int_a^b |f(x)|\mathrm{d}x$$

pf:

$$|f(x)| \geq f(x) \text{ and } |f(x)| \geq -f(x)$$

(this is just a propery of the modulus funciton)

so by monotonicity:

$$I(|f|) \geq I(f), \qquad \text{and,} \qquad I(|f|) \geq I(-f) = -I(f)$$

Then since $|I(f)|$ either equals $I(f)$ or $-I(f)$ (and $I(|f|)$ is $\geq$ both of these things)

$$I(|f|) \geq |I(f)|$$

$$\text{i.e.,} \qquad \int_a^b |f(x)|\mathrm{d}x \geq \left|\int_a^b f(x)\mathrm{d}x\right|$$

If $b < a$ we define $\int_a^b f(x)\mathrm{d}x := -\int_b^a f(x)\mathrm{d}x$.

Finally, we have the property of additivity:

Theorem 7.7 (Additivity):

$$\int_a^b f(x)\mathrm{d}x + \int_b^c f(x)\mathrm{d}x = \int_a^c f(x)\mathrm{d}x$$

The idea is that we can add the area under a curve on multiple adjacent intervals to find the area under the curve on the whole interval (though this still works even if we don't have $a < b < c$).

pf:

First assuming that $a < b < c$:

Let f_n be a sequence of step functions converging uniformly to f on $[a, c]$. f_n has a partition on $[a, c]$, if this partition does not include b as an endpoint, add the endpoint b.

Let the partition of f_n on $[a, b]$ be $a = x_0 < \cdots < x_M = b$, and the partition on $[b, c]$ be $b = x_M < \cdots < x_{M+N} = c$. Then the partition on $[a, c]$ is $a = x_0 < \cdots < x_{M+N} = c$.

$$\int_a^b f_n(x)\mathrm{d}x + \int_b^c f_n(x)\mathrm{d}x = \sum_{k=0}^{M-1}(x_{k+1}-x_k)f_n(x_k^*) + \sum_{k=M}^{M+N-1}(x_{k+1}-x_k)f_n(x_k^*)$$

$$= \sum_{k=0}^{M+N-1}(x_{k+1}-x_k)f_n(x_k^*) = \int_a^c f_n(x)\mathrm{d}x$$

Take the limit $n \to \infty$ on both sides:

$$\int_a^b f(x)\mathrm{d}x + \int_b^c f(x)\mathrm{d}x = \int_a^c f(x)\mathrm{d}x$$

If we don't have $a < b < c$, you can use $\int_a^b f(x)\mathrm{d}x = -\int_b^a f(x)\mathrm{d}x$ and go case by case (problem sheet).

We are now ready for the all-important mean value theorem:

Theorem 7.8 (MVT):

If f is continuous on $[a,b]$ then there exists a $c \in [a,b]$ with

$$f(c) = \frac{1}{b-a}\int_a^b f(x)\mathrm{d}x$$

pf:

$$\text{Let } d = \frac{1}{b-a}\int_a^b f(x)\mathrm{d}x$$

We wish to find a $c \in (a,b)$ with $f(c) = d$.

By the extreme value theorem there exist some $m \leq M$ such that $m \leq f(x) \leq M$ for all $x \in [a,b]$.

Then by monotonicity,

$$\int_a^b m\mathrm{d}x \leq \int_a^b f(x)\mathrm{d}x \leq \int_a^b M\mathrm{d}x$$

Since m and M are constants, they are step functions, and their integrals are trivial:

$$m(b-a) \leq \int_a^b f(x)\mathrm{d}x \leq M(b-a)$$

$$m \leq \frac{1}{b-a}\int_a^b f(x)\mathrm{d}x \leq M$$

$$m \leq d \leq M$$

Since m and M are attained (by the extreme value theorem) there exist some x_1, x_2 such that $f(x_1) = m, f(x_2) = M$:

$$f(x_1) \leq d \leq f(x_2)$$

Then (since f is continuous) by the intermediate value theorem, there exists some c between x_1 and x_2 with $f(c) = d$. Since $x_1, x_2 \in [a, b]$ and c is between x_1 and x_2, we have $c \in [a, b]$.

$$\text{Hence we have some } c \in [a, b] \text{ with } f(c) = \frac{1}{b-a}\int_a^b f(x)\mathrm{d}x$$

We are done!

The Mean Value Theorem is about to be of great importance.

The Fundamental Theorem of Calculus

Every calculus student should be familiar with the Fundamental Theorem of Calculus (FTC). There are a few ways to present the intuition for it, e.g., consider a distance time graph. $\text{velocity} = v$ $\text{displacement (signed distance)} = s$, $\text{time} = t$. $v = \frac{s}{t}$. A small change in distance divided by a small change in time (the derivative of displacement with respect to time) gives us velocity. On a velocity time graph, $s = vt$ so the displacement is width (t) times height (v). This is the area under the curve (integral of velocity with respect to time). Differentiating displacement gives velocity, integrating velocity gives displacement, suggesting the differentiation and integration are inverse operations.

Another intuition comes from the method of differences. The discrete version of $\frac{\mathrm{d}}{\mathrm{d}x}\big(f(x)\big)$ could be considered $\Delta f(k) = f(k+1) - f(k)$. The discrete version of an integral is a sum. $\sum_{k=0}^{n} \Delta\big(f(k)\big) = f(n+1) - f(0)$ (method of differences). The continuous analogue to this is that $\int_a^b \frac{\mathrm{d}}{\mathrm{d}x}\big(f(x)\big)\mathrm{d}x = f(b) - f(a)$.

However you think about it, or even if you just take the intuition for granted, the fundamental theorem of calculus says:

$$\text{Fundamental Theorem of Calculus (FTC part a) (7.9):}$$

$$\text{If } f \text{ is continuous on on } [a, b]$$

$$\text{Let } F(x) = \int_a^x f(t)\mathrm{d}t\,, \qquad x \in [a, b]$$

$$\text{Then } F'(x) = f(x)$$

pf:

We use the definition of the derivative:

$$F'(c) = \lim_{x\to c} \frac{F(x) - F(c)}{x - c} = \lim_{x\to c} \frac{\int_a^x f(t)\mathrm{d}t - \int_a^c f(t)\mathrm{d}t}{x - c} = \lim_{x\to c} \frac{\int_a^x f(t)\mathrm{d}t + \int_c^a f(t)\mathrm{d}t}{x - c}$$

Then use additivity

$$= \lim_{x\to c} \frac{\int_c^x f(t)\mathrm{d}t}{x - c}$$

Then use the Mean Value Theorem. If $x < c$ there is a $\xi \in [x, c]$ such that $f(\xi) = \frac{1}{c-x}\int_x^c f(t)\mathrm{d}t = \frac{1}{x-c}\int_c^x f(t)\mathrm{d}t$. If $x > c$ there is a $\xi \in [c, x]$ such that $f(\xi) = \frac{1}{x-c}\int_c^x f(t)\mathrm{d}t$. This ξ depends on x, so we will write $\xi(x)$. Either way, there is a $\xi(x)$ between x and c so that:

$$F'(c) = \lim_{x\to c} f\big(\xi(x)\big)$$

Since $\xi(x)$ is between x and c, $\lim_{x\to c}\xi(x) = c$ (squeezing theorem). Also, since f is continuous at c, $\lim_{x\to c} f(\xi\,(x) = f\left(\lim_{x\to c}\xi(x)\right) = f(c)$.

$$\text{so } F'(c) = f(c) \text{ (for every } c)$$

$$\text{i.e., } F'(x) = f(x)$$

We can then write the usual version:

$$\text{Fundamental Theorem of Calculus (FTC part b) (7.9):}$$

$$\text{Let } f \text{ be continuously differentiable on } [a, b]$$

$$\text{Then } \int_a^b f'(x)\mathrm{d}x = f(b) - f(a)$$

Continuously differentiable means differentiable and the derivative is continuous.

pf:

$$\text{Let } F(x) = \int_a^x f'(t)\mathrm{d}t$$

$$\text{Then } F'(x) = f'(x) \text{ (FTC part a)}$$

$$\Rightarrow F(x) = f(x) + C \text{ for some constant } C \text{ (Lemma 6.10)}$$

$$\text{Also } F(a) = \int_a^a f'(t)\mathrm{d}t = 0$$

$$F(a) = f(a) + C \Rightarrow 0 = f(a) + C$$

$$\Rightarrow C = -f(a)$$

$$\text{And } \int_a^b f'(x)\mathrm{d}x = \int_a^b f'(t)\mathrm{d}t = F(b) = f(b) + C = f(b) - f(a)$$

And we are done!

$$\text{e.g., } \int_0^b x\mathrm{d}x = \frac{1}{2}b^2 - \frac{1}{2}0^2 = \frac{b^2}{2}$$

$$\text{(makes sense since the area under the line creates a triangle with width} = \text{height} = b)$$

$$\text{e.g., } \int_0^{\pi} \sin x\,\mathrm{d}x = (-\cos\pi) - (-\cos 0) = -(-1) - (-1) = 1 + 1 = 2$$

Methods of Integration

We call $\int_a^b f(x)\mathrm{d}x$ a definite integral. We define the indefinite integral:

$$\int f(x)\mathrm{d}x \coloneqq \int_a^x f(t)\mathrm{d}t$$

$$\text{Where } a \text{ is an arbitrary constant.}$$

A definite integral gives a constant, and indefinite integral gives a function which (by FTC) is the antiderivative of f (at least in the case that f is continuous between a and x).

The arbitrary constant a gives rise to the arbitrary constant of integration $+C$. When we differentiate the $+C$ vanishes.

We also define the notation: $[F(x)]_a^b \coloneqq F(b) - F(a)$, so $\int_a^b f(x)\mathrm{d}x = [F(x)]_a^b$ (if f is continuous on $[a, b]$).

There are 2 methods that we will look at which will help us to evaluate integrals.

For both of these methods, we assume continuity of all functions involved.

Integration by Parts

If f and g are continuous, the Integration by Parts formula is:

$$\text{Theorem 7.10 (Integration by Parts):}$$

$$\int f(x)g(x)\mathrm{d}x = f(x)\int g(x)\mathrm{d}x - \int f'(x)\int g(x)\mathrm{d}x\,\mathrm{d}x$$

This comes from the product rule for differentiation.

pf:

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(f(x)\int g(x)\mathrm{d}x\right) = f'(x)\int g(x)\mathrm{d}x + f(x)g(x)$$

Integrate both sides (use FTC for the LHS):

$$f(x)\int g(x)\mathrm{d}x = \int f'(x)\int g(x)\mathrm{d}x\,\mathrm{d}x + \int f(x)g(x)\mathrm{d}x$$

$$\Rightarrow \int f(x)g(x)\mathrm{d}x = f(x)\int g(x)\mathrm{d}x - \int f'(x)\int g(x)\mathrm{d}x\,\mathrm{d}x$$

Integration by Substitution

The idea of integration by substitution is (first very informally):

$$\text{Let } g \text{ be continuously differentiable on } [a, b]$$

$$\text{and } f \text{ be continuous on } I = \mathrm{Ran}(g)$$

$\mathrm{Ran}(g)$ is the range of the function g which is an interval since continuous functions map intervals to intervals by IVT.

$$\text{I want to evaluate } \int_a^b f\big(g(x)\big)g'(x)\mathrm{d}x$$

$$\text{Let } u = g(x), \qquad \text{then } \frac{\mathrm{d}u}{\mathrm{d}x} = g'(x) \Rightarrow \mathrm{d}u = g'(x)\mathrm{d}x$$

$$\text{when } x = a, u = g(a), \qquad \text{when } x = b, u = g(b)$$

$$\Rightarrow \int_a^b f\big(g(x)\big)g'(x)\mathrm{d}x = \int_{g(a)}^{g(b)} f(u)\mathrm{d}u$$

The actual proof comes from the chain rule for differentiation.

Theorem 7.11 (Integration by Substitution):

$$\text{Let } g \text{ be continuously differentiable on } [a, b]$$

$$\text{and } f \text{ be continuous on } I = \mathrm{Ran}(g)$$

$$\text{Then } \int_a^b f\big(g'(x)\big)g(x)\mathrm{d}x = \int_{g(a)}^{g(b)} f(u)\mathrm{d}u$$

pf:

$$\text{Let } F(x) = \int f(x)\mathrm{d}x$$

$$F'(x) = f(x) \text{ so } F \text{ is differentiatble, hence } F \text{ is continuous}$$

$$\text{Let } h(x) = F\big(g(x)\big) \text{ (so } h \text{ is continuous)}$$

$$h'(x) = f\big(g(x)\big)g'(x) \text{ (chian rule)}$$

$$\text{By FTC: } \int_a^b f\big(g(x)\big)g'(x)\mathrm{d}x = \big[F\big(g(x)\big)\big]_a^b = F\big(g(b)\big) - F\big(g(a)\big)$$

$$\text{Also by FTC } \int_{g(a)}^{g(b)} f(u)\mathrm{d}u = [F(u)]_{g(a)}^{g(b)} = F\big(g(b)\big) - F\big(g(a)\big)$$

$$\text{So } \int_a^b f\big(g(x)\big)g'(x)\mathrm{d}x = \int_{g(a)}^{g(b)} f(u)\mathrm{d}u \text{ as required}$$

We are done.

We will leave it here for integration techniques.

Improper Integrals

We now deal with integrals to ∞ (and $-\infty$).

We simply define:

$$\int_a^\infty f(x)\mathrm{d}x \coloneqq \lim_{t\to\infty}\int_a^t f(x)\mathrm{d}x$$

$$\int_{-\infty}^b f(x)\mathrm{d}x \coloneqq \lim_{t\to\infty}\int_{-t}^b f(x)\mathrm{d}x$$

$$\text{e.g.,} \int_1^\infty \frac{1}{x^2}\mathrm{d}x = \lim_{t\to\infty}\int_1^t x^{-2}\mathrm{d}x = \lim_{t\to\infty}\left[-\frac{1}{x}\right]_1^t = \lim_{t\to\infty}\left(-\frac{1}{t}+1\right) = 1$$

$$\text{but} \int_1^\infty \frac{1}{x}\mathrm{d}x = \lim_{t\to\infty}[\log x]_1^t = \lim_{t\to\infty}(\log t - \log 1) = \lim_{t\to\infty}\log t = \infty \text{ diverges}$$

We also define:

$$\int_{-\infty}^\infty f(x)\mathrm{d}x \coloneqq \int_{-\infty}^0 f(x)\mathrm{d}x + \int_0^\infty f(x)\mathrm{d}x$$

Assuming both of these integrals converge. By additivity we could use $\int_{-\infty}^a f(x)\mathrm{d}x + \int_a^\infty f(x)\mathrm{d}x$ and get the same thing.

You may wonder why we don't define:

$$\int_{-\infty}^\infty f(x)\mathrm{d}x \coloneqq \int_{-t}^t f(x)\mathrm{d}x \ \ \text{(incorrect)}$$

This definition gives the same thing so long as the $-\infty$ and ∞ integrals converge, but there are some "$\infty - \infty$" cases where we get strange results. E.g., $\int_{-\infty}^\infty x\mathrm{d}x = \lim_{t\to\infty}\int_{-t}^t x\mathrm{d}x = \lim_{t\to\infty} 0 = 0$. We call this the Cauchy principle value of the improper integral centred at 0.

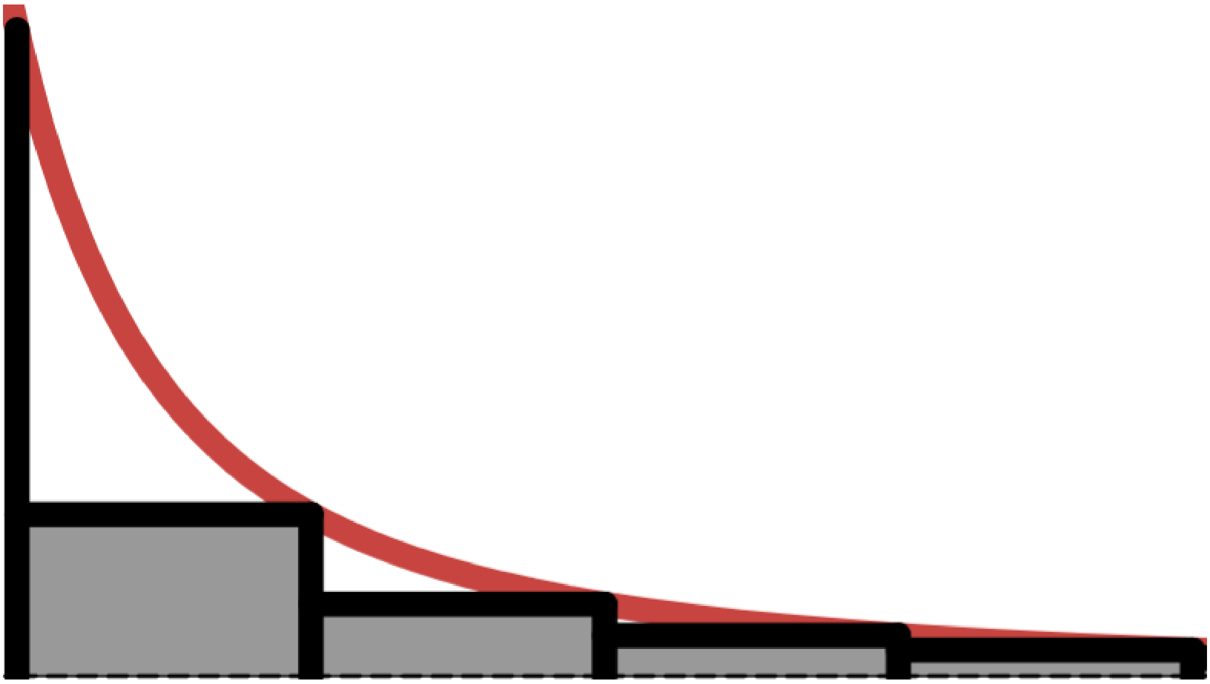
Even more weirdly, if we shift along and use, for example, $\lim_{t\to\infty}\int_{1-t}^{1+t} x\mathrm{d}x$, we would get divergence to $\to\infty$, (this would be the cauchy principle value of the integral, centred at 1). We will not discuss Cauchy princple values further.

Integral Test for Convergence

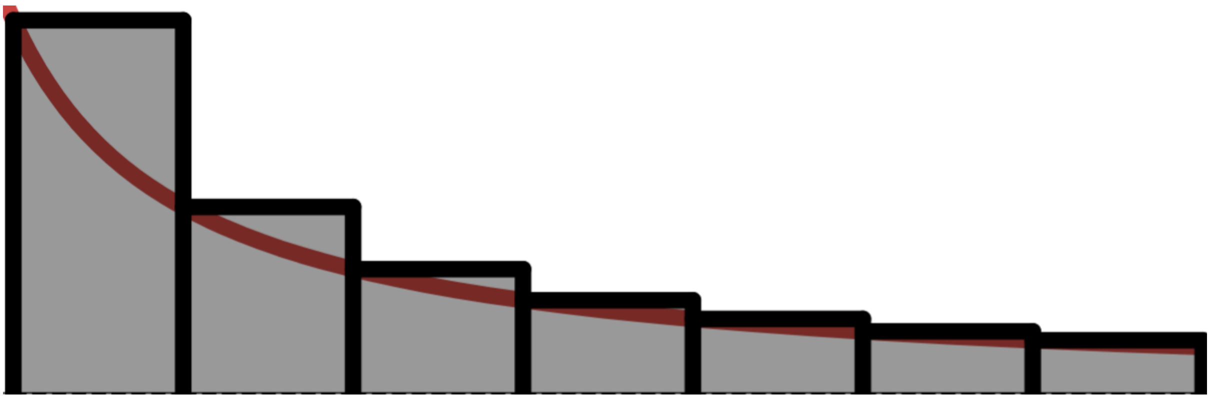
There is one test for convergence of infinite series which I mentioned once during chapter 3. We were not yet ready for it, but now we are. The idea is to compare an infinite series with an improper integral.

We think of our infinite series as an improper integral of step functions and compare it to another integral whose convergence status we know.

To show convergence, we compare against a function (the red curve) which is always above the discrete sum (represented by the black step function) and show that the improper integral of the function converges.



To show divergence, we instead compare against a divergent function which is always below the step function.



Amazingly we can use the same continuous function for both cases (we will see how).

This picture only works if $f(x) \geq 0$ and is decreasing.

Theorem 7.12 (Integral test for Convergence):

If $f(x) \geq 0$ is monotone decreasing,

$$\text{Then } \sum_{n=1}^{\infty} f(k) \text{ converges iff } \int_1^{\infty} f(x)\mathrm{d}x \text{ converges}$$

(Of course we don't need to start at $n = 1$, any natural number would do, but this will give the idea.)

pf:

Define $a_k = f(k) \geq 0$, then we are considering the series $\sum_{k=1}^{\infty} a_k$.

We find the area under the red curve on each interval $(k, k+1)$ as $b_k := \int_k^{k+1} f(x)\mathrm{d}x$

Then by additivity $\int_1^{\infty} f(x)\mathrm{d}x = \sum_{k=1}^{\infty} b_k$.

Since f is decreasing for $x \in (k, k+1)$, we have $0 \leq f(k+1) \leq f(x) \leq f(k)$,

$$\text{i.e.}, 0 \leq a_{k+1} \leq f(x) \leq a_k$$

$$\Rightarrow 0 \leq \int_k^{k+1} a_{k+1}\mathrm{d}x \leq \int_k^{k+1} f(x)\mathrm{d}x \leq \int_k^{k+1} a_k\mathrm{d}x$$

$$\Rightarrow 0 \leq a_{k+1} \leq b_k \leq a_k$$

Then by the comparison test, if $\sum_{k=1}^{\infty} a_k$ converges, then $\sum_{k=1}^{\infty} b_k$ converges, meaning $\int_1^{\infty} f(x)\mathrm{d}x$ converges.

Likewise, if $\sum_{k=1}^{\infty} a_k$ diverges, then $\sum_{k=1}^{\infty} a_{k+1}$ diverges, meaning $\sum_{k=1}^{\infty} b_k$ diverges, hence $\int_1^{\infty} f(x)\mathrm{d}x$ diverges.

We are done!

We can use the integral test to determine whether $\sum_{k=1}^{\infty} \frac{1}{k^p}$ converges (we already saw this is for $p > 1$ but we now have another way to do it).

for $p < 0, f$ is increasing for $x \geq 1$ so we cannot use the integral test but we

clearly have divergence in this case anyway (since each term of the series does not $\rightarrow 0$).

Also clear divergence for $p = 0$.

$$\frac{1}{x^p} \text{ is decreasing for } x \geq 1 \text{ and } p > 0$$

$$\sum_{k=1}^{\infty} \frac{1}{k^p} \text{ converges iff } \int_1^{\infty} \frac{1}{x^p} \mathrm{d}x \text{ converges}$$

$$\text{for } p \neq 1: \int_1^{\infty} \frac{1}{x^p} \mathrm{d}x = \lim_{t \to \infty} \left[\frac{1}{1-p} x^{1-p}\right]_1^t = \frac{1}{1-p}\left(\lim_{t\to\infty} t^{1-p} - 1\right)$$

$$\lim_{t\to\infty} t^{1-p} \text{ converges for } 1 - p < 0 \Leftrightarrow p > 1$$

$$\lim_{t\to\infty} t^{1-p} \text{ diverges for } 1 - p > 0 \Leftrightarrow p < 1$$

$$\text{For } p = 1: \int_1^{\infty} \frac{1}{x^p} \mathrm{d}x = \int_1^{\infty} \frac{1}{x} \mathrm{d}x \text{ diverges (as we saw before)}$$

$$\text{So } \sum_{k=1}^{\infty} \frac{1}{k^p} \text{ converges iff } p > 1 \text{ (which we already knew)}$$

This also means we can use all of our tests for convergence of infinite series to test whether certain improper integrals converge.

$$\text{e.g., } \int_0^{\infty} e^{-x^2} \mathrm{d}x \text{ converges iff } \sum_{k=0}^{\infty} e^{-k^2} \text{ converges}$$

$$0 \leq e^{-k^2} \leq e^{-k} \text{ for } (k \geq 1) \text{ and } \sum_{k=1}^{\infty} e^{-k} \text{ converges (geometric series)}$$

$$\text{so by comparison, } \sum_{k=1}^{\infty} e^{-k^2} \text{ converges and then } \sum_{k=0}^{\infty} e^{-k^2}$$

$$\text{hence by the integral test, } \int_0^{\infty} e^{-x^2} \mathrm{d}x \text{ converges}$$

($\int e^{-x^2} \mathrm{d}x$ is "non-elementary" meaning (loosely) that it cannot be written in terms of functions we are familiar with (like sine, cosine, roots, logairthms, exponentials). Non-elementary also means it cannot be written in terms of algebraic functions which are functions which are the roots of polynomials, but this is far, far beyond the scope of this course. $\int_0^{\infty} e^{-x^2} \mathrm{d}x$ turns out to be $\frac{\sqrt{\pi}}{2}$, which is (only slightly) outside the scope of this course.)

Term-Wise Differentiation and Integration

There are occasionaly situations where we might wish to swap the integral and sum symbols to make a problem easier,

$$\text{e.g.}, \int_0^{\frac{1}{2}} \sum_{k=0}^{\infty} \frac{1}{2^k} \cos(3^k \pi x)\, \mathrm{d}x = \sum_{k=0}^{\infty} \int_0^{\frac{1}{2}} \frac{1}{2^k} \cos(3^k \pi x)\, \mathrm{d}x = \sum_{k=0}^{\infty} \frac{1}{2^k} \left[\frac{1}{3^k \pi} \sin(3^k \pi x) \right]_0^{\frac{1}{2}}$$

$$= \frac{1}{\pi} \sum_{k=0}^{\infty} \frac{1}{6^k} \left(\sin\left(3^k \frac{\pi}{2}\right) - \sin 0 \right) = \frac{1}{\pi} \sum_{k=0}^{\infty} \frac{\sin\left(3^k \frac{\pi}{2}\right)}{6^k} = \frac{1}{\pi} \sum_{k=0}^{\infty} \left(-\frac{1}{6}\right)^k$$

$$= \frac{1}{\pi\left(1 + \frac{1}{6}\right)} = \frac{6}{7\pi}$$

You can check by looking at the graph of $\sin x$ (or more rigorously by induction and the triple angle formula) that indeed $\sin\left(3^k \frac{\pi}{2}\right) = (-1)^k$.

For this result to be valid, we need to make sure that this integral actually exists and that the first step (swapping the integral and the infinite series) is valid. Essentially, (when) can we integrate term-wise?

Since an infinite series is just a limit of partial sums, it will be sufficient to answer the question,

$$\text{if } f(x) = \lim f_n(x)$$

$$\text{when is } \int_a^b f(x) \mathrm{d}x = \lim \int_a^b f_n(x) \mathrm{d}x$$

That is, when can we take a limit outside of a definite integral.

We have already seen (in fact by definition) that this is true when f_n is a sequence of step functions and $f_n \to f$ uniformly. Does this work more generally? Yes!

Since regulated functions can be approximated arbitrarily well by step functions $(\text{Theorem } 7.1)$, f_n can be any regulated function and this will still work!

Theorem 7.13:

If f_n is a sequence of regulated functions converging uniformly to f on $[a, b]$

$$\text{Then } \int_a^b f(x)\mathrm{d}x = \lim \int_a^b f_n(x)\mathrm{d}x$$

pf:

We know that (since each f_n is regulated) $\int_a^b f_n(x)\mathrm{d}x$ exists. To show that $\int_a^b f(x)\mathrm{d}x$ exists, we need to show that f is regulated.

Since $f_n \to f$ uniformly:

$$\text{Let } \varepsilon > 0, \qquad \text{there exists an } N \text{ such that } n \geq N \Rightarrow |f(x) - f_n(x)| < \varepsilon$$

$$\text{this } N \text{ is independent of } x$$

For each $f_n(x)$ there is a step function $g_n(x)$ which approximates $f_n(x)$ well (Theorem 7.1). More precisely,

$$|f_n(x) - g_n(x)| < \varepsilon \text{ for every } x \in [a, b]$$

$$\text{Then } |f(x) - g_n(x)| \leq |f(x) - f_n(x)| + |f_n(x) - g_n(x)| < 2\varepsilon\ \forall_x$$

$$\text{Hence } g_n(x) \to f(x) \text{ uniformly meaning } f \text{ is regulated on } [a, b] \text{ and } \int_a^b f(x)\mathrm{d}x \text{ exists}$$

$$\text{Now to show that } \int_a^b f(x)\mathrm{d}x = \lim \int_a^b f_n(x)\mathrm{d}x:$$

$$\left|\int_a^b f(x)\mathrm{d}x - \int_a^b f_n(x)\mathrm{d}x\right| = \left|\int_a^b \big(f(x) - f_n(x)\big)\mathrm{d}x\right| \leq \int_a^b |f(x) - f_n(x)|\mathrm{d}x < \int_a^b \varepsilon \mathrm{d}x$$

$$= (b-a)\varepsilon$$

$$\text{Hence } \lim \int_a^b f_n(x)\mathrm{d}x = \int_a^b f(x)\mathrm{d}x$$

And we are done.

Recalling the example from before:

$$\int_0^{\frac{1}{2}} \sum_{k=0}^{\infty} \frac{1}{2^k}\cos(3^k \pi x)\,\mathrm{d}x$$

$$\text{Let } f_n(x) \coloneqq \sum_{k=0}^{n} \frac{1}{2^k}\cos(3^k \pi x), \qquad f(x) = \lim f_n(x) = \sum_{k=0}^{\infty} \frac{1}{2^k}\cos(3^k \pi x)$$

$\frac{1}{2^k}\cos(3^k \pi x)$ is continuous, hence $f_n(x)$ is continuous, hence it is regulated.

Using the Weierstrass M-Test:

$$\left|\frac{1}{2^k}\cos(3^k\pi x)\right| \leq \frac{1}{2^k}\text{, so let } M_k = \frac{1}{2^k}$$

$$\text{Then } \sum_{k=0}^{\infty} M_k = \sum_{k=0}^{\infty} \frac{1}{2^k} = 2 \text{ (converges)}$$

$$\text{Meaning } f_n(x) \text{ converges uniformly.}$$

(Meaning f is continuous since uniform convergence preserves continuity. This means f is regulated and so the integral actually exists.)

$$\text{Hence } \int_0^{\frac{1}{2}} \sum_{k=0}^{\infty} \frac{1}{2^k}\cos(3^k\pi x)\,\mathrm{d}x = \int_0^{\frac{1}{2}} f(x)\mathrm{d}x = \lim \int_0^{\frac{1}{2}} f_n(x)\mathrm{d}x$$

$$= \lim \int_0^{\frac{1}{2}} \sum_{k=0}^{n} \frac{1}{2^k}\cos(3^k\pi x)\,\mathrm{d}x$$

By linearity of integrals, we can always swap an integral with a finite sum,

$$= \lim \sum_{k=0}^{n} \int_0^{\frac{1}{2}} \frac{1}{2^k}\cos(3^k\pi x)\,\mathrm{d}x = \sum_{k=0}^{\infty} \int_0^{\frac{1}{2}} \frac{1}{2^k}\cos(3^k\pi x)\,\mathrm{d}x$$

The swapping of $\sum_{k=0}^{\infty}$ and $\int$ will always work as long as you can show that the infinite series converges uniformly and each term is regulated.

We now revisit a question from a while ago, when can we differentiate term-wise? We already know that you can differentiate power series term-wise, but we want something more general.

Since we can integrate term-wise if the series converges uniformly, perhaps we can differentiate if the sum of the derivatives converges uniformly. Since this theorem is so closely related to the previous one, we will almost certainly want to use the previous theorem in the proof.

We will again consider the sequence of partial sums f_n.

We want to relate the derivatives f_n' to the original sequence f_n, we can do this using FTC (so long as the f_n' are continuous, i.e., f_n is continuously differentiable).

Then we will have $\int_a^x f_n'(t)\mathrm{d}t = f_n(x) - f_n(a)$. We can then take the limit on both sides to get our result.

Theorem 7.14

$$\text{If } f_n \text{ converges to } f \text{ on } [a,b], \qquad \text{and each } f_n \text{ is continuously differentiable,}$$

$$\text{and } f_n' \text{ converges uniformly,}$$

$$\text{Then } f \text{ is differentiable with } f' = \lim f_n'$$

pf:

$$\text{Define } g(x) \coloneqq \lim f_n'(x)$$

$$\text{Since } f_n' \text{ is continuous, we can use FTC:}$$

$$\int_a^x f_n'(t)\mathrm{d}t = f_n(x) - f_n(a)$$

Since f_n' is continuous, it is regulated, also f_n' converges uniformly. Hence we can use Theorem 7.13 taking the limit as $n \to \infty$ on both sides.

$$\int_a^x g(t)\mathrm{d}t = f(x) - f(a)$$

Then since uniform convergence preserves continuity, g is continuous.

Then by FTC, we can differentiate both sides to get $g(x) = f'(x) \Rightarrow f'(x) = \lim f_n'(x)$.

(FTC guarantees that $\int_a^x g(t)\mathrm{d}t$ is differentiable and hence also that f is differentiable.)

Corollary 7.15 (termwise differentiation):

$$\text{If } f(x) = \sum_{k=0}^{\infty} f_k(x) \text{ for } x \in [a,b], \qquad \text{and each } f_k \text{ is continuously differentiable,}$$

$$\text{and } \sum_{k=0}^{\infty} f_k'(x) \text{ converges uniformly,}$$

$$\text{Then } f \text{ is differentiable with } f'(x) = \sum_{k=0}^{\infty} f_k'(x)$$

pf:

$$\text{Let } g_n(x) = \sum_{k=0}^{n} f_k(x)$$

$$\text{Then } \lim g_n(x) = f(x)$$

$$g_n'(x) = \sum_{k=0}^{n} f_k'(x) \text{ is continuous (since each } f_k \text{ is continuously differentiable)}$$

Hence each g_n is continuously differentiable

We are given that $g_n'(x)$ converges uniformly

So by Theorem 7.14, f is differentiable with

$$f'(x) = \lim g_n'(x) = \sum_{k=0}^{\infty} f_k'(x)$$

Final Word

Thank you for joining me on this fascinating, and sometimes grueling, joruney through the real numbers. I hoped you learned as much reading it as I did writing it. Make sure to check out the problem sheet if you haven't already. If you have any questions, don't hesitate to get in touch at c.m.math@outlook.com (check the YouTube channel for my email address in case this one changes). Also let me know if you have any suggestions for future courses.

See you next time. Bye!

Problem Sheet (no hints)

Well done for deciding to brave these problems without any help. If you get stuck, check the problem sheet with hints to help you out.

Proof, Sets and the Real Numbers

1)

a) Let $f: \mathbb{R} \to \mathbb{R}$ with $f(x) = x^2$. Show that f is neither injective nor surjective.

b) We wish to define $\sqrt{x} = f^{-1}(x)$ but we can't since f is not bijective.

We can solve this problem by restricting the domain to non-negative real numbers (to get

injectivity) and restricting the codomain to non-negative real numbers (to get surjectivity).

We can denote the set of non-negative real numbers as $\mathbb{R}_{\geq 0}$

Let $f: \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ with $f(x) = x^2$. Show that f is now bijective.

2) Show that $\{0,1\}$ forms a field if we define addition and multiplication in the usual way except with $1 + 1 = 0$. (Check the axioms of addition, multiplication, and distributivity).

3) Check that the rational numbers, $\mathbb{Q}$ form a field.

4) Let $a < b$ and $c < d$, show that $a + c < b + d$

5) Let $0 < a < b$ and $0 < c < d$, show that $ac < bd$

6) Let $a < b$ and $c < 0$, show that $bc < ac$

7) Let $0 < a < b$, show that $0 < b^{-1} < a^{-1}$

8) Prove Pascal's Identity:

Show that $\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$

9) Show that the supremum is unique. i. e., Let α, β be supremums of S, show that $\alpha = \beta$

10) Show that if f is non-empty and bounded below, that f has an infimum

11) Completing the proof that $\sqrt{2} \in \mathbb{R}$.

Let $S = \{x \in \mathbb{R}: x^2 < 2\}, \alpha = \sup S$. Show that $\alpha^2 > 2$ leads to a contradiction

12) Show that $\big||a| - |b|\big| \leq |a - b|$

13) Prove by induction that $k! \geq 2^{k-1}$ for $k \in \mathbb{N}$

Sequences

14) Show that limits of sequences are unique. That is, if L_1 and L_2 each fit the definition for the limit of a_n, then $L_1 = L_2$.

15) Show that if $b = \lim b_n$ with $b \neq 0$ and $b_n \neq 0$ for any n

Then $\lim \frac{1}{b_n} = \frac{1}{b}$

16) show that if $\lim a_n = \infty$ then $\lim \frac{1}{a_n} = 0$

17) Prove the Montone Convergence Theorem in the case that a_n is decreasing.

18) Show that the second version of the sequeezing theorem:

$$|a_n| \leq b_n \text{ for every } n \text{ and } \lim b_n = 0 \Longrightarrow \lim a_n = 0$$

Follows directly from the first version, which stated:

$$a_n \leq b_n \leq c_n \text{ for every } n \text{ and } \lim a_n = \lim c_n = L \Longrightarrow \lim b_n = L$$

19) Complete the proof of Theorem 2.6

If $x_n \in [a, b]$ is a sequence with $\lim x_n = L$, show that if $L < a$ then there is some $x_n < a$ which is a contradiction.

20) Let n_j be a strictly increasing sequence of natural numbers.

Prove by induction that $n_j \geq j$.

21) Let a_n be a bounded sequence. Show that there exists a subsequence a_{n_j} with $\lim a_{n_j} = \liminf a_n$

22) Let $a_n = \begin{cases} \dfrac{(-1)^n}{5} & n = 3k \text{ for some } k \in \mathbb{N} \\ e^{-n} + \dfrac{n^2 - 2n}{3n^2 - 1} & n = 3k - 1 \text{ for some } k \in \mathbb{N} \\ \dfrac{-2n + 1}{n^2 - 3n} & n = 3k - 2 \text{ for some } k \in \mathbb{N} \end{cases}$

(If you have not yet reached the part where we define the exponential function, you may assume that $\lim e^{-n} = 0$)

Determine $\limsup a_n$ and $\liminf a_n$.

Infinite Series

23) Determine whether the following series converge, state which test(s) you use:

a) $\displaystyle\sum_{n=1}^{\infty}\frac{n^2+3}{n^3-2}$

b) $\displaystyle\sum_{n=1}^{\infty}3\left(-\frac{1}{3}\right)^n$

c) $\displaystyle\sum_{n=1}^{\infty}\frac{n^2-4n}{n^4+2}$

d) $\displaystyle\sum_{n=1}^{\infty}\frac{n!}{3^n}$

e) $\displaystyle\sum_{n=1}^{\infty}\frac{1}{n^{2.5}}$

(Check your answers using at wolframalpha.com).

24) Complete the proof that e^x is bijective (3.14).

Let $y > 1$ and $X = \{x: e^x < y\}$. $\alpha = \sup X$. Suppoes $e^\alpha > y$.

Show that you can choose n so that $\alpha - \dfrac{1}{n}$ is an upper bound for X

(giving contradiction).

25) Prove the following properties of logarithms :

$\log(a^b) = b\log a$ (Recall: $x^\alpha := e^{\alpha \log x}$)

$\log ab = \log a + \log b$

$\log\dfrac{a}{b} = \log a - \log b$

$a^{\log_a b} = b$ $\left(\text{Recall } \log_a b := \dfrac{\log b}{\log a}\right)$

26) Prove the following properties of expressions like x^a [for $x > 0$]

(You will likely want to use the properties of exponentials):

$$x^{a+b} = x^a x^b$$

$$x^{ab} = (x^a)^b$$

$$x^{-\alpha} = \frac{1}{x^{\alpha}}$$

Limits of Functions and Continuity

27) Given $L = \lim_{x\to c} f(x)$ exists,

Show that $\lim_{x\to c^-} f(x)$ and $\lim_{x\to c^+} f(x)$ both exist and are both $= L$

28) Prove directly from the ε-δ definition the following results:

$$\lim_{x\to c}\big(f(x)+g(x)\big) = \lim_{x\to c} f(x) + \lim_{x\to c} g(x)$$

$$\lim_{x\to c}\big(f(x)g(x)\big) = \lim_{x\to c} f(x) \lim_{x\to c} g(x)$$

$\lim_{x\to c}\dfrac{f(x)}{g(x)} = \dfrac{\lim_{x\to c} f(x)}{\lim_{x\to c} g(x)}$ so long as $\lim_{x\to c} g(x) \neq 0$ and $g(x) \neq 0$ sufficiently close to c

29) Prove that limits are unique.

30) Use the negation of the definition of uniform continuity to show that e^x is not uniformly continuous:

i. e., you need to show that

$\exists_{\varepsilon>0}$, s. t., $\forall_{\delta>0}$, $\exists_{x,c\in\mathbb{R}}$, s. t., $|x-c|<\delta$ and $|e^x - e^c| \geq \varepsilon$

Differentiation and Differentiability

31) Use the limit definition of the derivative and the binomial theorem to show that

$\frac{\mathrm{d}}{\mathrm{dx}}(x^n) = nx^{n-1}$ for $n \in \mathbb{W}$.

32) During the proof of Cauchy′s Generalised Mean Value theroem, we used the guess:

$$h(t) = f(t) - \frac{f(b) - f(a)}{g(b) - g(a)} g(t) - f(a) + \frac{f(b) - f(a)}{g(b) - g(a)} g(a)$$

Which represents the vertical difference between the curve $\left(f(t)\right)$ and the secant line joining the endpoints. By determining the equation of the secant line (between points $\left(g(a), f(a)\right)$ and $\left(g(b), f(b)\right)$), derive this guess for $h(t)$.

33) Using L'Hospitals rule (and justifying that you can use it), evaluate:

a) $\lim\limits_{x \to 0} \frac{n^x - 1}{x}$ (your answer should be in terms of n)

b) $\lim\limits_{x \to 0^+} x \ln x$

c) $\lim\limits_{x \to \infty} x^{\frac{1}{x}}$

34) Using L′Hospital′s rule, show that $\lim\limits_{\alpha \to \infty} \left(1 + \frac{x}{\alpha}\right)^{\alpha} = e^x$ (this is a common alternative definition for the exponential function).

35) Determine the maxima and minima of $f(x) = x^3 - \frac{5}{2}x^2 + \frac{3}{4}x$ on $[-1,1]$

Functions as Infinite Series

36) Filling a gap in the proof of the Lagrange Remainder:

$$\text{Given } F(t) = f(x) - \sum_{k=0}^{n} \frac{f^{(k)}(t)}{k!}(x-t)^k$$

$$\text{Show that } F'(t) = \frac{-f^{(n+1)}(t)}{n!}(x-t)^n$$

37) I wish to approximate e^x by $1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \frac{x^5}{120} + \frac{x^6}{720} + \frac{x^7}{5040}$, for $0 < x < 0.5$

(order 7 taylor polynomial of e^x)

Using the Lagrange remainder, show that the error in this approximation is always $< \frac{1}{5000000}$

(You should use that $e < 4$, $8! = 40320 > 40000$, $2^8 = 256 > 250$)

38) Let f_n be a sequence of functions which converge uniformly on a closed interval I, to a limit function f (also with domain I).

If each f_n is continuous at the endpoints of I, show that f is continuous at the endpoints of I.

39) Prove that $x^n - y^n = (x-y)\sum_{k=0}^{n-1} x^{n-1-k}y^k$ for $n \in \mathbb{N}$

a) by induction
b) by considering $\sum_{k=0}^{n-1} x^{n-1-k}y^k$ as a geometric series

40) By using the strong root test, show that the formal derivative, $\sum_{k=1}^{\infty} ka_k x^{k-1}$ has the same radius of convergence as $\sum_{k=0}^{\infty} a_k x^k$.

41) Given $f(x) := \sum_{k=0}^{\infty} a_k(x-c)^k$, show (by induction) that $f^{(n)}(c) = n!\,a_n$

(you may assume that we can differentiate power series term-wise).

42) Use the ratio test to show that the power series for cosine, $\sum_{k=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!}$, converges

for all $x \in \mathbb{R}$

43) Show that the Lagrange Remainder for the natural logarithm $\to 0$ as n
$\to \infty$ for

$x \geq \frac{1}{2}$. As we saw in the notes, the remainder is $R_n = \frac{(-1)^n}{n+1}\left(\frac{x-1}{\xi}\right)^{n+1}$ where ξ is

strictly between x and 1.

Integration

44) Given that $\alpha, \beta \in \mathbb{R}$ and that $\lim f_n = f, \lim g_n = g$ (both uniform limits), show

that $\lim(\alpha f_n + \beta g_n) = \alpha f + \beta g$ converges uniformly (we already know that it

converges to this limit, but we need to check that the convergence is uniform).

45) Given that $f(x)$ and $g(x)$ are regulated, show that $f(x)g(x)$ and $|f(x)|$ are regulated.

46) We have seen in the notes that when $a < b < c$,

$\int_a^b f(x)\mathrm{d}x + \int_b^c f(x)\mathrm{d}x = \int_a^c f(x)\mathrm{d}x$. Using the fact that $\int_a^b f(x)\mathrm{d}x = -\int_b^a f(x)\mathrm{d}x$

$\left(\text{and that } \int_a^a f(x)\mathrm{d}x = 0\right)$, show that the result is true for any $a, b, c \in \mathbb{R}$

(assuming f is regulated on $[\min(a, b, c), \max(a, b, c)]$.

47) Suppose that $f(x) \geq 0$ for every $x \in [a, b]$ and there exists some $c \in [a, b]$ with

$f(c) > 0$ and that f is continuous on $[a, b]$. Show that $\int_a^b f(x)\mathrm{d}x > 0$. This is one of

my favourite proofs from this course. Try it yourself before looking at the hints.

48)

a) Use the quotient rule for differentiation to show that $\frac{\mathrm{d}}{\mathrm{d}x}\tan x = \sec^2 x$

$\left(\text{recall } \tan x \coloneqq \frac{\sin x}{\cos x}\right)$

b) Then use the inverse function rule for differentiation to show that

$$\frac{\mathrm{d}}{\mathrm{d}x}\arctan x = \frac{1}{1+x^2} \text{ (arctan is the inverse of tan)}$$

c) Use integration by substitution to verify that indeed $\displaystyle\int_0^x \frac{1}{1+t^2}\,\mathrm{d}t = \arctan x$

49) Determine whether $\displaystyle\int_0^\infty \frac{\sin x}{2x+e^x}\,\mathrm{d}x$ converges.

50) When evaluating $\displaystyle\int_0^{\frac{1}{2}} \sum_{k=0}^{\infty} \frac{1}{2^k}\cos(3^k \pi x)\,\mathrm{d}x$, at one point, we had to evaluate $\sin\left(3^k \frac{\pi}{2}\right)$.

You can tell by looking at a graph of sine that this is $(-1)^k$, but (for completeness), I want you to prove it by induction. You might want to use that $\sin(3x) = 3\sin x - 4\sin^3 x$.

Show (by induction) that $\sin\left(3^k \frac{\pi}{2}\right) = (-1)^k$ for every integer $k \geq 0$.

51) $\phi(x)$ $\displaystyle := \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{1}{2}t^2}\,\mathrm{d}t$ is a very important integral in probability and statistics (used to find probabilities for a normal distribution). We cannot evaluate this integral directly by finding an antiderivative (we will not discuss why not here)but we can evaluate it as an infinite series.

Find a power series for $\phi(x)$, justifying each step. You may assume without proof that

$\phi(0) = \frac{1}{2}$.

Problem Sheet (with hints)

I recommend you use the *no hints* version of the problem sheet until you get stuck, then look at this one for the hints.

[Hint: Hints appear like this]

Proof, Sets and the Real Numbers

1a) Let $f: \mathbb{R} \to \mathbb{R}$ with $f(x) = x^2$. Show that f is neither injective nor surjective.

b) We wish to define $\sqrt{x} = f^{-1}(x)$ but we can't since f is not bijective.

We can solve this problem by restricting the domain to non-negative real numbers (to get

injectivity) and restricting the codomain to non-negative real numbers (to get surjectivity).

We can denote the set of non-negative real numbers as $\mathbb{R}_{\geq 0}$

Let $f: \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ with $f(x) = x^2$. Show that f is bijective.

[Hint 1: for injectivity show that $f(a) = f(b) \Rightarrow a = b$]

[Hint 2: For surjectivity, you can simply note that the domain and codomain are the same ($\mathbb{R}_{\geq 0}$)]

2) Show that $\{0,1\}$ forms a field if we define addition and multiplication in the usual way except with $1 + 1 = 0$. (Check the axioms of addition, multiplication, and distributivity).

3) Check that the rational numbers, $\mathbb{Q}$ form a field.

4) Let $a < b$ and $c < d$, show that $a + c < b + d$

5) Let $0 < a < b$ and $0 < c < d$, show that $ac < bd$

6) Let $a < b$ and $c < 0$, show that $bc < ac$

7) Let $0 < a < b$, show that $0 < b^{-1} < a^{-1}$

8) Pascal's Identity:

Show that $\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$

[Hint: If you're struggling to get from LHS to RHS, you can simplify the LHS and RHS into the same expression]

9) Show that the supremum is unique. i. e., Let α, β be supremums of S, show that $\alpha = \beta$

10) Show that if f is non-empty and bounded below, that f has an infimum

[Hint: Consider the function $-f(x)$ and use the axiom of completeness]

11) Completing the proof that $\sqrt{2} \in \mathbb{R}$.

Let $S = \{x \in \mathbb{R}: x^2 < 2\}, \alpha = \sup S$. Show that $\alpha^2 > 2$ leads to a contradiction

[Hint: aim to find an upper bound which is $< \alpha$, try to find an n such that $\left(\alpha - \frac{1}{n}\right)^2 > 2$]

12) Show that $\big||a| - |b|\big| \leq |a - b|$

[Hint: you could do this by cases, but it might be easier to observe that $|a| = |b + a - b| \leq |b| + |a - b|$ if you can similarly show that $|b| - |a - b| \leq |a|$, you can use Theorem 1.12 and you will be done]

13) Prove by induction that $k! \geq 2^{k-1}$ for $k \in \mathbb{N}$

Sequences

14) Show that limits of sequences are unique. That is, if L_1 and L_2 each fit the definition for the limit of a_n, then $L_1 = L_2$.

[Hint 1: use proof by contradiction, suppose $L_1 \neq L_2$ and let $\varepsilon = \frac{|L_1 - L_2|}{2}$]

[Hint 2: show that you can find an n such that $|a_n - L_1| < \varepsilon$ and $|a_n - L_2| < \varepsilon$]

[Hint 3: use the triangle inequality: $|L_1 - L_2| = |L_1 - a_n + a_n - L_2|$

$\leq |a_n - L_1| + |a_n - L_2| < 2\varepsilon = |L_2 - L_2|$,

but $|L_2 - L_2| < |L_1 - L_2|$ is a contradiction]

15) Show that if $b = \lim b_n$ with $b \neq 0$ and $b_n \neq 0$ for any n

Then $\lim \frac{1}{b_n} = \frac{1}{b}$

[Hint: show that $|b_n|$ is has a *lower* bound $C > 0$ (hinges on $b \neq 0$)

$\left|\frac{1}{b_n} - \frac{1}{b}\right| = \frac{|b_n - b|}{|b||b_n|}$, then make $|b_n - b| < \varepsilon C|b|$]

16) show that if $\lim a_n = \infty$ then $\lim \frac{1}{a_n} = 0$

17) Prove the Montone Convergence Theorem in the case that a_n is decreasing.

[Hint: Shortcut, consider the sequence $-a_n$ and apply the increasing case]

18) Show that the second version of the sequeezing theorem:

$$|a_n| \leq b_n \text{ for every } n \text{ and } \lim b_n = 0 \Longrightarrow \lim a_n = 0$$

Follows directly from the first version, which stated:

$$a_n \leq b_n \leq c_n \text{ for every } n \text{ and } \lim a_n = \lim c_n = L \Longrightarrow \lim b_n = L$$

[Hint: $|a_n| \leq b_n \Leftrightarrow -b_n \leq a_n \leq b_n$]

19) Complete the proof of Theorem 2.6

If $x_n \in [a, b]$ is a sequence with $\lim x_n = L$, show that if $L < a$ then there is some $x_n < a$ which is a contradiction.

20) Let n_j be a strictly increasing sequence of natural numbers.

Prove by induction that $n_j \geq j$.

21) Let a_n be a bounded sequence. Show that there exists a subsequence a_{n_j}

$$\text{with } \lim a_{n_j} = \liminf a_n$$

[Hint: See proof of 2.17a]

22) Let $a_n = \begin{cases} \dfrac{(-1)^n}{5} & n = 3k \text{ for some } k \in \mathbb{N} \\ e^{-n} + \dfrac{n^2 - 2n}{3n^2 - 1} & n = 3k - 1 \text{ for some } k \in \mathbb{N} \\ \dfrac{-2n + 1}{n^2 - 3n} & n = 3k - 2 \text{ for some } k \in \mathbb{N} \end{cases}$

(If you have not yet reached the part where we define the exponential function, you may assume that $\lim e^{-n} = 0$)

Determine $\limsup a_n$ and $\liminf a_n$.

Infinite Series

23) Determine whether the following series converge, state which test(s) you use:

a) $\displaystyle\sum_{n=1}^{\infty} \frac{n^2+3}{n^3-2}$

b) $\displaystyle\sum_{n=1}^{\infty} 3\left(-\frac{1}{3}\right)^n$

c) $\displaystyle\sum_{n=1}^{\infty} \frac{n^2-4n}{n^4+2}$

d) $\displaystyle\sum_{n=1}^{\infty} \frac{n!}{3^n}$

e) $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^{2.5}}$

(Check your answers using at wolframalpha.com).

24) Complete the proof that e^x is bijective (3.14).

Let $y > 1$ and $X = \{x: e^x < y\}$. $\alpha = \sup X$. Suppoes $e^\alpha > y$.

Show that you can choose n so that $\alpha - \dfrac{1}{n}$ is an upper bound for X

(giving contradiction).

25) Prove the following properties of logarithms :

$\log(a^b) = b \log a$ (Recall: $x^\alpha := e^{\alpha \log x}$)

$\log ab = \log a + \log b$

$\log \dfrac{a}{b} = \log a - \log b$

$a^{\log_a b} = b$ $\left(\text{Recall } \log_a b := \dfrac{\log b}{\log a}\right)$

26) Prove the following properties of expressions like x^a [for $x > 0$]

(You will likely want to use the properties of exponentials):

$x^{a+b} = x^a x^b$

$x^{ab} = (x^a)^b$

$x^{-\alpha} = \dfrac{1}{x^\alpha}$

Limits of Functions and Continuity

27) Given $L = \lim_{x \to c} f(x)$ exists,

Show that $\lim_{x \to c^-} f(x)$ and $\lim_{x \to c^+} f(x)$ both exist and are both $= L$

28) Prove directly from the ε-δ definition the following results:

$$\lim_{x \to c}\big(f(x) + g(x)\big) = \lim_{x \to c} f(x) + \lim_{x \to c} g(x)$$

$$\lim_{x \to c}\big(f(x)g(x)\big) = \lim_{x \to c} f(x) \lim_{x \to c} g(x)$$

$$\lim_{x \to c} \frac{f(x)}{g(x)} = \frac{\lim_{x \to c} f(x)}{\lim_{x \to c} g(x)}$$ so long as $\lim_{x \to c} g(x) \neq 0$ and $g(x) \neq 0$ sufficiently close to c

[Hint: these proofs are similar to the analogous proofs for sequences]

29) Prove that limits are unique.

[Hint: see hint for Q14]

30) Use the negation of the definition of uniform continuity to show that e^x is not uniformly continuous:

i. e., you need to show that

$$\exists_{\varepsilon > 0}, \text{s.t.}, \quad \forall_{\delta > 0}, \quad \exists_{x, c \in \mathbb{R}}, \text{s.t.}, \quad |x - c| < \delta \text{ and } |e^x - e^c| \geq \varepsilon$$

[Hint 1: Any choice of $\varepsilon > 0$ will end up working, so let $\varepsilon = 1$, then let $\delta > 0$.

Your task is now to find any pair of real numbers x, c so that $|x - c| < \delta$ but $|e^x - e^c| \geq 1$.

(you should expect these x, c to depend on δ)]

[Hint 2: let $x = c - \frac{\delta}{2}$ (ensures $|x - c| < \delta$) then

rearrang $|e^x - e^c| = 1$ to find c.

If there is a solution, you now have x, c such that

$|x - c| < \delta$ and $|e^x - e^c| \geq 1 (= \varepsilon)$ and you are done.

(You should get $c = \log\left(\frac{1}{1-e^{\frac{-\delta}{2}}}\right)$)]

Differentiation and Differentiability

31) Use the limit definition of the derivative and the binomial theorem to show that $\frac{\mathrm{d}}{\mathrm{dx}}(x^n) = nx^{n-1}$ for $n \in \mathbb{W}$. Remember $\mathbb{W} = \{0,1,2,3,\cdots\}$

32) During the proof of Cauchy's Generalised Mean Value theroem, we used the guess:

$$h(t) = f(t) - \frac{f(b) - f(a)}{g(b) - g(a)} g(t) - f(a) + \frac{f(b) - f(a)}{g(b) - g(a)} g(a)$$

Which represents the vertical difference between the curve $(f(t))$ and the secant line joining the endpoints. By determining the equation of the secant line (between points $(g(a), f(a))$ and $(g(b), f(b))$), derive this guess for $h(t)$.

[Hint: When you find the height of the secant line in the form $mx + c$, make sure to substitute $x = g(t)$ to get it in terms of t and not x]

33) Using L'Hospitals rule (and justifying that you can use it), evaluate:

a) $\lim_{x \to 0} \frac{n^x - 1}{x}$ (your answer should be in terms of n)

b) $\lim_{x \to 0^+} x \ln x$

[Hint: $x \ln x = \frac{\ln x}{1/x}$]

c) $\lim_{x \to \infty} x^{\frac{1}{x}}$

[Hint: Write $x^{\frac{1}{x}} = e^{\frac{1}{x} \ln x} = e^{\frac{\ln x}{x}}$]

34) Using L'Hospital's rule, show that $\lim_{\alpha \to \infty} \left(1 + \frac{x}{\alpha}\right)^{\alpha} = e^x$ (this is a common alternative definition for the exponential function).

[Hint: write $\lim\left(1+\frac{x}{\alpha}\right)^{\alpha} = e^{\lim \alpha \log\left(1+\frac{x}{\alpha}\right)}$ then use L'Hopital's rule to evaluate $\exp\left(\lim_{\alpha\to\infty} \alpha\, log\left(1+\frac{x}{\alpha}\right)\right) = \exp\left(\lim_{\alpha\to\infty} \frac{log\left(1+\frac{x}{\alpha}\right)}{1/\alpha}\right)$]

35) Determine the maxima and minima of $f(x) = x^3 - \frac{5}{2}x^2 + \frac{3}{4}x$ on $[-1,1]$

[Hint 1: The only places where maxima and minima can possibly occur are; where the derivative is 0; where the function is not differentiable; endpoints (since the theorem which says minima and axima have derivative 0 only work for interior points)]

Functions as Infinite Series

36) Filling a gap in the proof of the Lagrange Remainder:

$$\text{Given } F(t) = f(x) - \sum_{k=0}^{n} \frac{f^{(k)}(t)}{k!}(x-t)^k$$

$$\text{Show that } F'(t) = \frac{-f^{(n+1)}(t)}{n!}(x-t)^n$$

[Hint: Method of differences]

37) I wish to approximate e^x by $1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \frac{x^5}{120} + \frac{x^6}{720} + \frac{x^7}{5040}$, for $0 < x < 0.5$

(order 7 taylor polynomial of e^x)

Using the Lagrange remainder, show that the error in this approximation is always $< \frac{1}{5000000}$

(You should use that $e < 4, 8! = 40320 > 40000, 2^8 = 256 > 250$)

[Hint: $0 < \xi < 0.5$]

38) Let f_n be a sequence of functions which converge uniformly on a closed interval I, to a limit function f (also with domain I).

If each f_n is continuous at the endpoints of I, show that f is continuous at the endpoints of I.

[Hint: this is finishing the proof of theorem 6.3 and almost identical to the proof for interior points]

39) Prove that $x^n - y^n = (x-y)\sum_{k=0}^{n-1} x^{n-1-k}y^k$ for $n \in \mathbb{N}$

a) by induction

b) by considering $\sum_{k=0}^{n-1} x^{n-1-k}y^k$ as a geometric series

40) By using the strong root test, show that the formal derivative, $\displaystyle\sum_{k=1}^{\infty} ka_k x^{k-1}$

has the same radius of convergence as $\displaystyle\sum_{k=0}^{\infty} a_k x^k$.

[Hint 1: $\sum_{k=1}^{\infty} ka_k x^{k-1} = \frac{1}{x}\sum_{k=1}^{\infty} ka_k x^k$ for $x \neq 0$, so you should consider the radius of convergence of $\sum_{k=1}^{\infty} ka_k x^k$ instead]

[Hint 2: You will need to show that $\limsup k^{\frac{1}{k}} = 1$. You can do this by considering logarithms: $\lim k^{\frac{1}{k}} = e^{\lim \frac{1}{k}\log k} = e^{\lim \frac{\log k}{k}} = e^0 = 1$].

41) Given $f(x) := \displaystyle\sum_{k=0}^{\infty} a_k (x-c)^k$, show (by induction) that $f^{(n)}(c) = n!\, a_n$

(you may assume that we can differentiate power series term-wise).

[Hint 1: instead of applying induction directly, you should use induction to prove a general form for $f^{(n)}(x)$. Then use this to find $f^{(n)}(c)$]

[Hint 2: You should try to show that $f^{(n)}(x) := \sum_{k=n}^{\infty} \frac{k!}{(k-n)!} a_k (x-c)^{k-n}$]

42) Use the ratio test to show that the power series for cosine, $\displaystyle\sum_{k=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!}$, converges

for all $x \in \mathbb{R}$

43) Show that the Lagrange Remainder for the natural logarithm $\to 0$ as $n \to \infty$ for

$x \geq \frac{1}{2}$. As we saw in the notes, the remainder is $R_n = \frac{(-1)^n}{n+1}\left(\frac{x-1}{\xi}\right)^{n+1}$ where ξ is strictly between x and 1.

[Hint 1: consider $x \in \left[\frac{1}{2}, 1\right)$ separately from $x \in (1,2]$ (we can ignore $x = 1$ as $R_n = 0$ for $x = 1$)]

[Hint 2: $x \in (1,2]$ is the easier case. Not that $\xi \in (1,2)$, and $x - 1 \in (0,1]$, meaning $0 < x - 1 < \xi \Rightarrow 0 < \frac{x-1}{\xi} < 1$. Then $\left(\frac{x-1}{\xi}\right)^{n+1}$ becomes a geometric sequence with limit 0]

[Hint 3: $x \in \left[\frac{1}{2}, 1\right)$ is harder. Note that $\frac{1}{2} < \xi < 1$ and $\frac{1}{2} \leq x < 1 \Rightarrow -\frac{1}{2} \leq x - 1 < 0 \Rightarrow 0 < 1 - x \leq \frac{1}{2} < \xi$. then consider $|R_n| = \frac{1}{n+1}\left(\frac{|x-1|}{\xi}\right)^{n+1}$ and notice that again we have a geometric sequence which $\to 0$]

Integration

44) Given that $\alpha, \beta \in \mathbb{R}$ and that $\lim f_n = f, \lim g_n = g$ (both uniform limits), show

that $\lim(\alpha f_n + \beta g_n) = \alpha f + \beta g$ converges uniformly (we already know that it converges to this limit, but we need to check that the convergence is uniform).

[Hint: the proof is literally identical to the proof that $\lim(\alpha x_n + \beta y_n) = \alpha \lim x_n + \beta \lim y_n$ where x_n, y_n are sequences of constants. This is because the choices of N_1, N_2 are independent of x (because uniform convergence).]

45) Given that $f(x)$ and $g(x)$ are regulated, show that $f(x)g(x)$ and $|f(x)|$ are regulated.

46) We have seen in the notes that when $a < b < c$,

$\int_a^b f(x)\mathrm{d}x + \int_b^c f(x)\mathrm{d}x = \int_a^c f(x)\mathrm{d}x$. Using the fact that $\int_a^b f(x)\mathrm{d}x = -\int_b^a f(x)\mathrm{d}x$

$\left(\text{and that } \int_a^a f(x)\mathrm{d}x = 0\right)$, show that the result is true for any $a, b, c \in \mathbb{R}$

(assuming f is regulated on $[\min(a,b,c), \max(a,b,c)]$.

[Hint: In the case that $a < b$ you have 5 cases:

$a < b < c$ (already done), $a < b = c, a < c < b, c = a < b, c < a < b$

In the case $b < a$ you have another 5 cases:

$b < a < c, b < a = c, b < c < a, c = b < a, c < b < a$

In the case $a = b$ there are 3 more cases:

$a = b < c, a = b = c, c < a = b$

You can do all cases separately if you like (or skip the question once you've got the idea or you might be able to spot some symmetries and write “WLOG” to skip some cases.]

47) Suppose that $f(x) \geq 0$ for every $x \in [a, b]$ and there exists some $c \in [a, b]$ with

$f(c) > 0$ and that f is continuous on $[a, b]$. Show that $\int_a^b f(x)\mathrm{d}x > 0$. This is one of

my favourite proofs from this course. Try it yourself before looking at the hints.

[Hint: use the ε-δ definition of continuity and set $\varepsilon = \frac{f(c)}{2}$

to find an interval $(c - \delta, c + \delta)$ in which $|f(x) - f(c)| < \frac{f(c)}{2}$.]

[Hint 2: Then $\frac{f(c)}{2} \leq f(x)$ on this interval so use monotonicty:

$0 < \frac{f(c)}{2}(b - a) \leq \int_a^b f(x)\mathrm{d}x$ and you are done!]

48)

a) Use the quotient rule for differentiation to show that $\frac{\mathrm{d}}{\mathrm{d}x} \tan x = \sec^2 x$

$\left(\text{recall } \tan x := \frac{\sin x}{\cos x}\right)$

b) Then use the inverse function rule for differentiation to show that

$\frac{\mathrm{d}}{\mathrm{d}x} \arctan x = \frac{1}{1 + x^2}$ (arctan is the inverse of tan)

c) Use integration by substitution to verify that indeed $\int_0^x \frac{1}{1 + t^2}\mathrm{d}t = \arctan x$

[Hint: You will need to use the identity $1 + \tan^2 x = \sec^2 x$, which can be easily obtained

by taking $\sin^2 x + \cos^2 x = 1$ and dividing both sides by $\cos^2 x$.]

49) Determine whether $\int_0^\infty \frac{\sin x}{2x + e^x}\mathrm{d}x$ converges.

[Hint 1: Consider $\sum_{n=0}^{\infty}\frac{\sin n}{2n+e^n}$ and use the integral test for convergence.]

[Hint 2: Compare to a geometric series. For $n \geq 0$, $\left|\frac{\sin n}{2n+e^n}\right| \leq \frac{1}{2n+e^n} \leq \frac{1}{e^n}$]

50) When evaluating $\int_0^{\frac{1}{2}}\sum_{k=0}^{\infty}\frac{1}{2^k}\cos(3^k\pi x)\,\mathrm{d}x$, at one point, we had to evaluate $\sin\left(3^k\frac{\pi}{2}\right)$.

You can tell by looking at a graph of sine that this is $(-1)^k$, but (for completeness), I want you to prove it by induction. You might want to use that $\sin(3x) = 3\sin x - 4\sin^3 x$.

Show (by induction) that $\sin\left(3^k\frac{\pi}{2}\right) = (-1)^k$ for every integer $k \geq 0$.

51) $\phi(x) := \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{1}{2}t^2}\mathrm{d}t$ is a very important integral in probability and statistics (used to find probabilities for a normal distribution). We cannot evaluate this integral directly by finding an antiderivative (we will not discuss why not here)but we can evaluate it as an infinite series.

Find a power series for $\phi(x)$, justifying each step. You may assume without proof that $\phi(0) = \frac{1}{2}$.

[Hint 1: Write $e^{-\frac{t^2}{2}}$ by using the power series definition of e^x, replacing x by $-\frac{t^2}{2}$.]

[Hint 2: (By additivity) $\phi(x) = \phi(0) + \int_0^x e^{-t^2}\mathrm{d}t = \frac{1}{2} + \int_0^x e^{-t^2}\mathrm{d}t$]

[Hint 3: Use M-test to show that you may swap $\int$ and $\sum$.

$\left|\frac{(-1)^k t^{2k}}{2^k k!}\right| = \frac{t^{2k}}{2^k k!} \leq \frac{x^{2k}}{2^k k!} = M_k$ (this is fine since x is fixed)]

[Hint 4: to show that $\sum_{k=0}^{\infty} M_k$ converges, either use the ratio test, or note that it equals $e^{\frac{x^2}{2}}$.]

[Ans: you should end up with $\frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{2^k k!(2k+1)}$ (or equivalent)]