“Philosophy is written in this grand book, I mean the universe, which stands continuously open to our gaze, but it cannot be understood unless one first learns to comprehend the language in which it is written. It is written in the language of mathematics and its characters are triangles, circles, and other geometric figures without which it is humanly impossible to understand a single word of it. Without these, one is wondering about in a dark labyrinth.” - Galileo Galilei

Axioms, The Mathematical Foundations

If you have ever seen a mathematical proof, or even if you haven't, you will know that they all have something in common. A beginning. Of course all proofs must start somewhere, with some information which is already known. Logic can be used to show that, given that the initial information is true, the end conclusion must also be true. The initial condition may come a theorem which was itself proved, but the proof of that theorem must also have had a starting point. You may see the problem here. Where does this chain end, what is at the very foundation of mathematics, the seed from which all of mathematics grows, the initial assumptions of the most simple theorems. The answer: axioms.

The world of mathematics is one of human creation, which means that we, mathematians, define the rule which govern this universe. This is not science, where hypotheses must be tested and the more evidence gathered to support an idea, the more likely that idea is to be true. No. This is mathematics, in which our assumptions are true because we say they are, any consequences of these assumptions, these axioms simply become what we observe. I believe that the answer to the question, is mathematics invented or discovered is thus: we invent axioms and we discover their consequences.

How should we decide which axioms to define? Typically it makes sense to create axioms based on intuitive information because otherwise there may be disputes as to whether or not an axiom should be accepted. Before creating any axioms though, it would be usedul to create some definitions.

A simple definition to define might incrementation. To increase a number by one. Starting at 0, we get 1, then 2 then 3, 4, 5 and so on.

Next defining addition. I will define $a + b$ as meaning "increment from b times, starting at a".

So to find $2 + 3$, start at two and increase by one thrice.

2, 3 (that's once), 4 (that's twice), 5 (that's three times), so $2 + 3$ is 5

This may all seem rather basic, but of course we must start somewhere.

Next I will define multiplication as repeated addition: $a \times b$ is $a + a + a + a + a + \cdots$ such that there are b as written down. e.g., 4×6 is $4 + 4 + 4 + 4 + 4 + 4$, which six fours being added together in total which is 24.

I will define one more, exponentiation, as repeated multiplication: a^b is $a \times a \times a \times a \times$ ... such that there are b as written down. e.g. 4^3 is $4 \times 4 \times 4$ which is three fours being multiplied together which is 64.

It is clear that addition can be used to find the total number of things given different sets of things. For example, if I have a apples, and b bananas, how many fruit (assuming bananas are counted as fruit) do I have in total? $a + b$ of course (it is rather intuitive why). But why not $b + a$? It seems obvious that these two quantities $a + b$ and $b + a$ should be equivalent.

To write this more concisely I will define a new notation: $=$ (equals). This symbol essentially means, "what is on the left of the symbol is equivalent to (the same as) what is on the right of the symbol". So $a+b$ is equivalent to $b+a$ can be rewritten as $a+b=b+a$.

But how to prove it? Have I already proved it? Do I even need to prove it? Just because $a+b$ is something and $b+a$ is the same thing, doesn't mean that $a+b$ is the same as $b+a$. It may seem like they should be, but they aren't, not yet. That's because there is no axiom to state that such a thing should be true. So now, I declare axiomatically that if two things are both equal to the same thing then those two things are equal to each other. Now, using this axiom I can say that:

$$a+b=b+a$$

This property has a name, addition is "commutative".

Before moving on, I would like to point out a certain property of equality. If what is on the left of the equals sign is the same thing as what is on the right, then it reasons that if you do an operation to both sides, then you are doing something to a number on one side and then doing the same thing to the same number on the other side, meaning that the equality should be maintained. I will now create an axiom for this, stating that: if an operation is performed on both sides of an equation, then the equality is maintained. This axiom lies at the heart of algebra.

Imagine a grid of squares, width a and height b. The total number of squares (known as the area of the grid) could be found by counting all squares in the grid, or more simply by adding the number of squares in each row together. The number of squares in each row is a as the width of the grid is a, the number of rows is b as the height of the grid is b, so the total number of squares is $a+a+a+a+a+\cdots$ with b as being added together. This is just $a\times b$, so the total number of squares in the grid is $a\times b$. If the grid was rotated so that it was now on its side, the total number of squares would be the same, but the height and width would be switched so the total number of squares is now $b\times a$ which means that, if we call the total number of squares n, then $n=a\times b$ and $n=b\times a$ and so:

$$a\times b=b\times a$$

If addition and multiplication are both commutative, it makes sense that exponentiation should be as well. Should I declare that as an axiom? Well, be careful, axioms should only be used sparingly and only when you are confident that it does not:

- Contradict other axioms directly.
- Lead to conclusions which contradict other axioms directly.
- Contradict with conclusions led to by other axioms.
- Lead to conclusions which contradict the conclusions led to by other axioms.
- Lead to conclusions which contradict itself.
- Lead to conclusions which contradict conclusions led to by itself.

Or worst of all:

- Directly contradict itself.

To put it simply, no contradictions.

Let's see what happens if I declare axiomatically that $a^\wedge b=b^\wedge a$. Are any contradictions created? As it turns out, no. If I let $a=3$ and $b=2$ then we get $3^\wedge 2=2^\wedge 3$ meaning $3\times 3=2\times 2\times 2$ and so $9=8$. This seems ridiculous, but there is no contradiction here. There is no rule against such a conclusion, so let's create one. I axiomatically declare that if a is some number and b is some number which is not the same as a, then $a \neq b$. ($\neq$ means "not equal to"). Now the axiom which states that $a\hat{}b = b\hat{}a$ does lead to a contradiction and so this axiom must be discarded. This may seem like a simple axiom which doesn't even need to be stated, but in the world of mathematics everything must either be an axiom or derived from them. To make it more clear in the future that exponentiation is not commutative I will change the notation from $a\hat{}b$ to a^b.

There are some more operations to create before moving on, the inverse operations. In other words, doing an operation in reverse. The inverse of incrementation will be called decrementation, which can be thought of as a question. What number could I increment to get this number. Decrementing 5, for example, is asking, which number could I increase by 1 to get 5? This is of course 4. It is not hard to see that decrementation simply means finding the previous number.

Next, I will define the inverse of addition as subtraction. $a - b$ is asking the question, b added to what number gives a? In other words: if $a - b = c$ then $c + b = a$. If addition is repeated incrementation, and subtraction is the inverse of that, then subtraction is repeated decrementation. To find $5 - 3$, decrement five three times, 5 (zero times), 4 (once), 3 (twice), 2 (three times), so $5 - 3 = 2$. Alternatively, $5 - 3 = c$, $5 = 3 + c$ so $c = 2$ because $3 + 2 = 5$ so $5 - 3 = 2$.

Note that if $a - a = c$ then $a = a + c$ so $c = 0$ because $a + 0 = a$ meaning $a - a = 0$ for all a.

Next is the inverse of multiplication which I will call division. $a \div b = \frac{a}{b} = a\ divided\ by\ b$. $\frac{a}{b}$ means what number multiplied by b equals a, in other word, if $\frac{a}{b} = c$ then $a = bc$. (bc means the same as $b \times c$). E.g., $\frac{12}{4} = c$, $12 = 4c$, so $c = 3$ because $4 \times 3 = 12$ so $\frac{12}{4} = 3$.

Note that if $\frac{a}{a} = c$ then $a = ac$ so $c = 1$ because $a = a \times 1$ meaning $\frac{a}{a} = 1$ for all a.

The inverse of exponentiation, I will get to much later.

The next order of business is to extent these operations to work for all values as currently they do not work for all numbers.

First, decrementation which works for all numbers except 0 because "what number can you add 1 to and get 0" has no answers. Writing this as an equation where x is the unknown number:

$$x + 1 = 0$$

$$x + 1 - 1 = 0 - 1$$

$$x + 0 = 0 - 1$$

$$x = 0 - 1$$

So $x = 0 - 1$, which seems to be a nonsense statement. How could a number be less than 0. To solve this problem, I will define a new type of number, negative numbers. Numbers can either be positive $(0 + a)$ or negative $(0 - a)$ or, so that I don't have to write $0 + \cdots$ or $0 - \cdots$ each time, I will write a for a positive number, and $-a$ for a negative number.

To decrement -1:

$$x + 1 = -1$$

$$2 - 2 = 0$$

$$1 - 2 = 0 - 1$$

$$1 + (0 - 2) = -1$$

$$(0 - 2) + 1 = -1$$

$$-2 + 1 = -1$$

$$x + 1 = -1$$

$$x = -2$$

Decrementing -1 gives -2, -2 gives -3, then -4, -5 and so on.

This new type of number, “negative numbers” has now been defined with the property that:

$$0 - a = -a$$

where a is a positive number and $-a$ is a negative number.

Next, subtraction only works when the first number $\geq$ (is greater than or equal to) the second. e.g., $3 - 5$ has no solutions because $3 < 5$ (three is less than five).

This problem can be solved by using negative numbers. $3 - 5$ means decrement three five times. 3, 2, 1, 0, -1, -2. So $3 - 5 = -2$.

Division also only works for certain numbers, $e.g., \frac{12}{4} = 3$, but $\frac{5}{2}$ does not make any sense.

$$x = \frac{5}{2}$$

$$2x = 5$$

has no solutions as there is no number satisfying the equation. To solve this problem, I will create a new type of number: fractions. Simply let $x = \frac{5}{2}$ be the solution.

I will next move on to defining multiplication of fractions.

How should $a \times \frac{1}{b}$ be defined?

If $c = \frac{1}{b}$ then $1 = bc$, so $a = abc, \frac{a}{b} = ac$

$$a \times \frac{1}{b} = a \times c = \frac{a}{b}$$

$$a \times \frac{1}{b} = \frac{a}{b}$$

As for multiplying two fractions together:

$$Let\ x = \frac{a}{b} \times \frac{c}{d}$$

$$x = a \times \frac{1}{b} \times c \times \frac{1}{d}$$

$$x \times b \times d = a \times b \times \frac{1}{b} \times c \times d \times \frac{1}{d}$$

$$xbd = a \times \frac{b}{b} \times c \times \frac{d}{d}$$

$$xbd = ac$$

$$x = \frac{ac}{bd}$$

$$\frac{a}{b} \times \frac{c}{d} = \frac{ac}{bd}$$

This also means that $\frac{ak}{bk} = \frac{a}{b} \times \frac{k}{k} = \frac{a}{b} \times 1 = \frac{a}{b}$, so $\frac{ak}{bk} = \frac{a}{b}$ meaning you can multiply the top and bottom of a fraction by some constant and the fraction will be unchanged.

Back to division momentarily: What happens if we divide by a fraction?

$$x = a \div \frac{1}{b}$$

$$x \times \frac{1}{b} = a$$

$$\frac{x}{b} = a$$

$$x = ab$$

$$a \div \frac{1}{b} = ab$$

What about dividing one fraction by another?

$$x = \frac{a}{b} \div \frac{c}{d}$$

$$x = \frac{\left(\frac{a}{b}\right)}{\left(\frac{c}{d}\right)}$$

$$x \times \frac{c}{d} = \frac{a}{b}$$

$$x \times c \times \frac{1}{d} = \frac{a}{b}$$

$$xcb \times \frac{1}{d} = a$$

$$\frac{bcx}{d} = a$$

$$bcx = ad$$

$$x = \frac{ad}{bc}$$

$$\frac{a}{b} \div \frac{c}{d} = \frac{ad}{bc} = \frac{a}{b} \times \frac{d}{c}$$

$$\frac{a}{b} \div \frac{c}{d} = \frac{a}{b} \times \frac{d}{c}$$

Next, multiplying negatives. e.g., $a \times -b$. Remember multiplying is repeated addition, so $a \times -b$ means add $-b$ as together which doesn't make any sense. We could look at it the other way however: adding a $(-b)$s together, so $a \times -b = (0 - b) + (0 - b) + (0 - b) + \cdots = -b - b - b - b - \cdots$.

An example: $a \times -1 = -1 - 1 - 1 - 1 - \cdots$ which decrements a times, starting from 0, meaning $a \times -1 = -a$

Now imagine a grid of squares, height x, width a and another grid height x, width b.
$total\ number\ of\ squares = number\ of\ squares\ in\ the\ first\ grid + number\ of\ squares\ in\ the\ second\ grid = ax + bx$

If you put these two grids together side by side, the total number of squares has not changed, and the grid now has height x and width $(a + b)$, so $total\ number\ of\ squares = x(a + b)$

So $x(a + b) = ax + bx$

$$a \times -b = -b - b - b - b - \cdots = (-1 \times b) + (-1 \times b) + (-1 \times b) + \cdots$$

$$= -1 \times (b + b + b + b + b + \cdots)$$

$$= -1 \times ab = -ab$$

$$so\ a \times -b = -ab$$

Note:

$$x(a + b) = ax + bx$$

$$(a + b)(c + d) = a(c + d) + b(c + d) = ac + ad + bc + bd$$

$$(a + b)^2 = (a + b)(a + b) = aa + ab + ba + bb = a^2 + 2ab + b^2$$

$$So\ (a + b)^2 = a^2 + 2ab + b^2$$

What about a negative multiplied by a negative?

$$0 = 0^2$$

$$0 = (1 - 1)^2$$

$$0 = 1^2 + 2(1)(-1) + (-1)^2$$

$$0 = 1 - 2 + (-1)(-1)$$

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

$$(-1)(-1) = 1$$

More generally:

$$-a \times -b = -1 \times a \times -1 \times b$$

$$-a \times -b = -1 \times -1 \times a \times b$$

$$-a \times -b = 1 \times a \times b$$

$$-a \times -b = ab$$

So, the negatives cancel each other out.

Also:

$$-(a - b) = -1 \times (a + (-1) \times b) = -1 \times a + (-1) \times (-1) \times b = -a + b = b - a$$

$$So\ -(a - b) = b - a$$

Next, exponentiation only work when the exponent is a positive integer, (for a^b, a is the base and b is the exponent).

When multiplying two exponentials of the same base, watch what happens:

$$a^b \times a^c = (a \times a \times a \times a \times \dots) \times (a \times a \times a \times a \times \dots)$$

Where the left bracket has b as and the right bracket has c as

$$a^b \times a^c = a \times a \times a \times a \times a \times \dots$$

for a total of $(b + c)$ as, so:

$$a^b \times a^c = a^{b+c}$$

So, multiplying exponentials of the same base adds together their exponents.

$$\left(a^b\right)^c = (a \times a \times a \times a \times \dots)^c$$

The bracket has b as in it.

$$\left(a^b\right)^c = (a \times a \times a \times a \times \dots) \times (a \times a \times a \times a \times \dots) \times (a \times a \times a \times a \times \dots) \times \dots$$

each bracket has b as in it & there are c brackets, so the total number of as is bc so:

$$\left(a^b\right)^c = a^{bc}$$

To ensure that exponentials are as convenient to use as possible, I will attempt to ensure that for negative and fractional definitions of the operation, these qualities are preserved.

First, what is a^0, or to be more precise, how should a^0 be defined.

$$a^b \times a^0 = a^{b+0} = a^b$$

$$a^b \times a^0 = a^b$$

$$a^0 = 1$$

So, I now define a^0 as 1.

$$a^b \times a^{-b} = a^{b+(-b)} = a^{b-b} = a^0 = 1 = \frac{a^b}{a^b}$$

$$a^b \times a^{-b} = \frac{a^b}{a^b}$$

$$a^{-b} = \frac{1}{a^b}$$

Exponentials have now been defined for negative numbers.

$$\left(a^{\frac{1}{b}}\right)^b = a^{\frac{b}{b}} = a^1 = a$$

$$\left(a^{\frac{1}{b}}\right)^b = a$$

$a^{\frac{1}{b}}$ is a number such that raising it to the power of b gives you a. I will create a new notation for this: $\sqrt[b]{a}$.

$$a^{\frac{1}{b}} = \sqrt[b]{a}$$

Addition of negative numbers has already been accounted for by the axiom $a + (-b) = a - b$

Addition of fractions is shown here.

$$\frac{a}{b} + \frac{c}{b}$$

$$= a \times \frac{1}{b} + c \times \frac{1}{b}$$

$$= \frac{1}{b}(a + c)$$

$$= \frac{a + c}{b}$$

$$\frac{a}{b} + \frac{c}{b} = \frac{a + c}{b}$$

To add an integer to a fraction:

$$a + \frac{b}{c}$$

$$= \frac{ac}{c} + \frac{b}{c}$$

$$= \frac{ac + b}{c}$$

Before finishing this chapter, I think there is one more thing worth discussing: inequalities. An inequality is like an equation, but instead of saying that one thing is equal to another, it says that one thing is greater than another. To introduce the symbols; $<$ means "is less than", $>$ means "is greater than", $\leq$ means "is less than or equal to" and $\geq$ means "is greater than or equal to".

Like with equations, a number can be added to both sides of an inequality. If you imagine both number on a number line, adding and subtracting moves both numbers to the left or the right by the same amount. Whichever number was the furthest to the right on the number line (whichever was the greatest number) will still be further to the right than the other afterwards (is still the greatest of the two). As for multiplying and dividing, multiplying both numbers by some constant will simply scale them about the origin on the number line, moving them further apart. Dividing does the same but the numbers would move closer together instead. In either case it is intuitively obvious whichever number was greater before will be the greatest of the two after.

This is all unless the number by which you are multiplying or dividing is negative. If both numbers are positive or both negative, they will be reflected on the number line meaning the one which was the furthest right before, is now furthest left and vice-versa. If one number is positive and the other is negative, the positive number $>$ the negative number, but when both are multiplied or divided by a negative number the positive one becomes negative and the negative one becomes positive. Either way, if multiplying or dividing both side of an inequality by a negative number, you must change the sign. $<$ changes to $>$, $>$ changes to $<$, $\leq$ changes to $\geq$ and $\geq$ changes to $\leq$.

So numbers can be added to and subtracted from both sides of an inequality, both sides can be multiplied or divided by some positive number. Both sides can also be multiplied or divided by some negative number, but the sign must then be changed.

This chapter probably seemed quite annoying to read because most of what was done seemed intuitive and it seems as though all of this has overcomplicated what could have been so simple. However, as I said before, we must start somewhere, and luckily the absolute basics have now been laid out and operations defined. We will now be able to move on to more interesting things.

To summarise everything thus far: axioms must be defined, and from them, theorems can arise. At first many axioms needed to be created as something cannot be proved from nothing, but as we went on, less axioms needed to be defined.

To summarise just some of what was covered here:

$$1 \div \frac{1}{b} = b$$

$$\frac{a}{b} \times \frac{c}{d} = \frac{ac}{bd}$$

$$\frac{a}{b} \div \frac{c}{d} = \frac{a}{b} \times \frac{d}{c}$$

$$\frac{ak}{bk} = \frac{a}{b}$$

$$\frac{a}{c} + \frac{b}{c} = \frac{a+b}{c}$$

$$-a \times b = -ab$$

$$-a \times -b = ab$$

$$-(a-b) = b - a$$

$$a^b \times a^c = a^{b+c}$$

$$\left(a^b\right)^c = a^{bc}$$

$$a^{-b} = \frac{1}{a^b}$$

$$a^{\frac{1}{b}} = \sqrt[b]{a}$$

As well as the definitions of the incrementation, addition, multiplication, exponentiation, decrementation, subtraction and division operations.

Finally, if an operation is done to both sides of an equation, the equality shall hold.

The purpose of this chapter has been to demonstrate the purpose of axioms and to lay out some foundations of algebra which can later be used. This chapter has also shown how operations can be extended to accept new types of inputs, even if this requires the invention of new types of numbers.

In summary, as I said before:

We invent axioms and we discover their consequences.

Euclidean Geometry & Trigonometry

Euclidean geometry is geometry on a flat surface, a piece of paper for example. This means that Euclidean geometry is not concerned with curved surfaces such as the surface of the Earth, or spacetime.

Euclidean geometry is the likely the simplest and most intuitive form of geometry and is also the easiest to visualise as it is geometry on flat surfaces.

As I discussed in the previous chapter, axioms must be defined before anything else can be done and so we must define some axioms for Euclidean geometry.

When Euclid himself created Euclidean geometry, these were his five axioms:

1. A straight line may be drawn between any two points.

2. Any terminated straight line may be extended indefinitely.

3. A circle may be drawn with any given point as centre and any given radius.

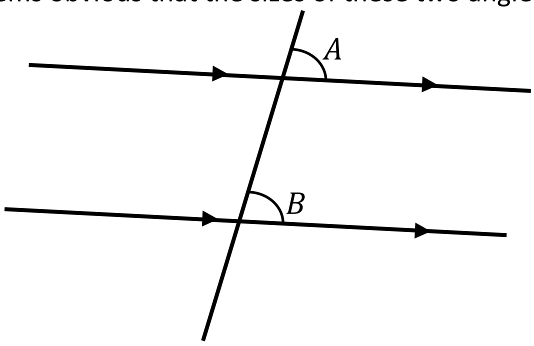
4. All right angles are equal.

5. If two straight lines in a plane are met by another line, and if the sum of the internal angles on one side is less than two right angles, then the straight lines will meet if extended sufficiently on the side on which the sum of the angles is less than two right angles.

(A right angle is a quarter of a full rotation, if two lines have a right angle between them, then those lines are perpendicular.)

The last of these axioms is a bit more complicated than the others. I believe that there is simpler axiom which can be defined instead, and from that I can derive what was his 5th axiom.

In the above diagram, A and B are two angles which have been labelled (These are referred to as corresponding angles). It seems obvious that the sizes of these two angles are equal. This could be

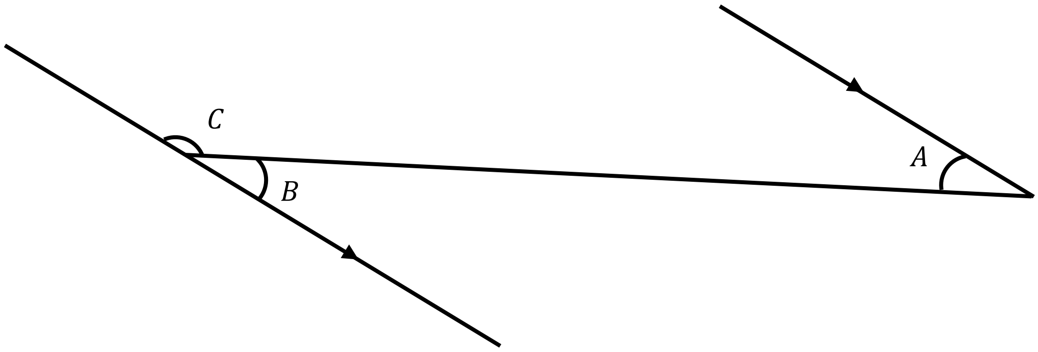


further justified by sliding one of the parallel lines up and down. As this is done, the angle does not change, showing that $A = B$. I will let this be an axiom: corresponding angles are equal.

Using Euclid's fourth axiom and the definition of a right angle, all right angles are equal, and a right angle is a quarter of a full rotation so all full rotations are equal to four right angles and so are equal to each other. Also, all half rotations are equal to two right angles and so are equal to each other. To assign units to angles, I will define a degree as being a unit of measurement for angles such that there are $360°$ (° means degrees) in a full rotation. I use this number because 360 has a lot of factors meaning it can be divided by many different integers and give an integer result. A half rotation (sum angles on a straight line) should therefore be $\frac{360°}{2} = 180°$. A quarter rotation (A right angle) should be $\frac{360°}{4} = 90°$.

In this diagram, angles C and B are co-interior angles. $A = B$ because they are corresponding angles (previous axiom), $A + C = 180°$ because they are angles on a straight line. So $B + C = 180°$, meaning co-interior angles add to $180°$.

Here, I have defined an axiom that corresponding angles are equal and used this to show that co-interior add to $180°$, Euclid instead defined as an axiom that co-interior angles add to $180°$ and used that to show that corresponding angles are equal.

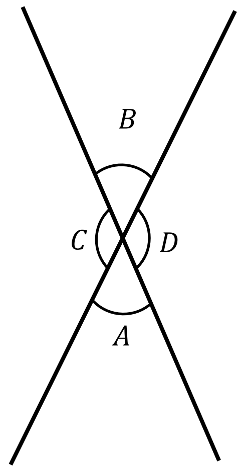


In this diagram, A and B are alternate angles. $A + C = 180°$ because they are co-interior angles. $C + B = 180°$ because they are angles on a straight line.

$$A + C = C + B$$

$$A = B$$

So alternate angles are equal.

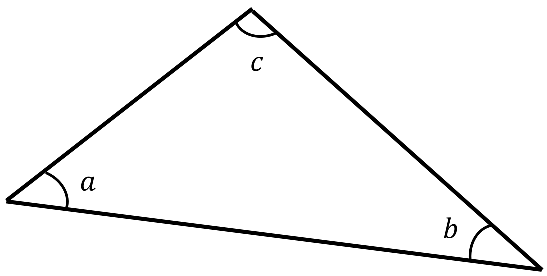


In this diagram, A and B are called vertically opposite angles. C and D are also called vertically opposite angles. It seems relatively intuitive that vertically opposite angles are equal so I could have started by defining that as an axiom and proving the others from there.

$A + D = 180°$ because they are angles on a straight line. $B + D = 180°$ because they are angles on a straight line. So, $A + D = B + D$ meaning $A = B$ so vertically opposite angles are equal.

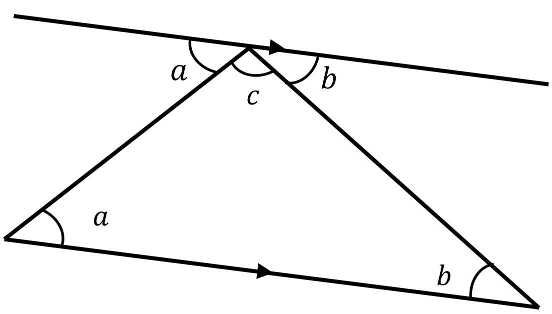
I will now move on to one of the most important parts of geometry, trigonometry, the study of triangles.

Here is a triangle, angles a, b and c.



$$sum\ of\ angles\ in\ triangle = a + b + c$$

If you draw a line through one of the vertices (corners) which is parallel to the side opposite that vertex (corner), you get the following diagram:



The new angles are labelled a and b because they are alternate angles to a and b meaning they are equal to them.

$$a + b + c = 180°$$

because they are angles on a straight line.

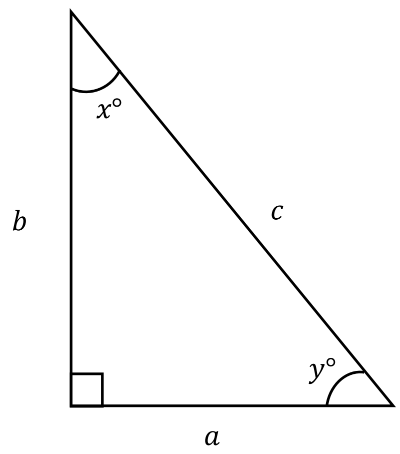
$$sum\ of\ angles\ in\ triangle = a + b + c = 180°$$

$$sum\ of\ angles\ in\ triangle = 180°$$

Angles in a triangle sum to $180°$ or two right angles. This is one of the most basic facts about triangles.

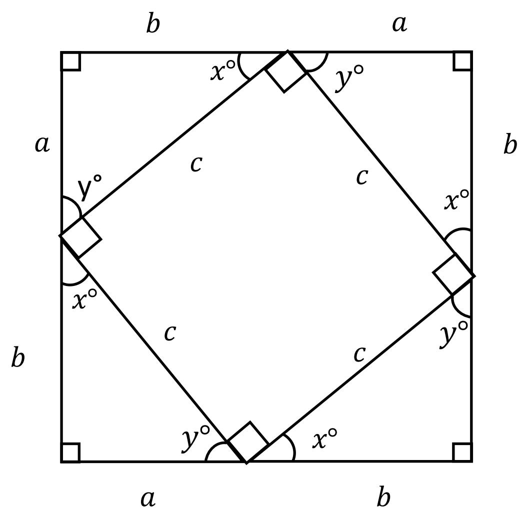
Next I will show arguably the most infamous theorem in all of mathematics, the Pythagorean Theorem.

A right triangle is a triangle with a right angle in it. The side opposite the right angle is known as the triangle's hypotenuse, with side length here labelled as c and the other two side lengths are labelled a and b.



Note that $x + y + 90 = 180$ because angles in a triangle add to $180°$ this will be important later.

If I take four copies of this diagram and arrange them like so:



The resulting diagram is what appears to be a large square with a smaller square inscribed in it. The larger shape is clearly a square because it has four sides, each corner is a right angle and each side is the same length ($a+b$). The smaller shape also has four sides of equal length (c), but does how do we know it has four right angles?

If we take one corner and label its angle as z, then $x+y+z=180$ because they are angles on a straight line.

$x+y+90=180$, and $x+y+z=180$ so:

$$z=90$$

This argument could be applied to all four corners of the smaller shape meaning the smaller shape is a square.

There are two ways to find the area of the smaller square (remember that the area of a rectangular grid is its width multiplied by its height).

The smaller square has side lengths of c so

$$small\ square\ area = c \times c = c^2$$

Or it could be found by subtracting the areas of the triangles from the area of the larger square.

$$small\ square\ area = large\ square\ area - sum\ of\ triangle\ areas$$

sA right triangle is simply a rectangle cut diagonally in half meaning its area is half of that of the rectangle. For a rectangle, height h and width w, $area = hw$, so for a triangle, height h and width w, $area = \frac{hw}{2}$

$$Area\ of\ each\ triangle = \frac{ab}{2}$$

There are four triangles so

$$sum\ of\ triangle\ areas = \frac{ab}{2} \times 4 = 2ab$$

$$large\ square\ area = (a+b)^2 = a^2 + 2ab + b^2$$

$$small\ square\ area = a^2 + 2ab + b^2 - 2ab$$

$$small\ square\ area = a^2 + b^2$$

$$a^2 + b^2 = small\ square\ area = c^2$$

$$a^2 + b^2 = c^2$$

$$Notice\ that\ c = \sqrt{a^2 + b^2}$$

and if I multiply the two sides a and b by some constant, k:

$$(ak)^2 + (bk)^2 = c^2$$

$$a^2k^2 + b^2k^2 = c^2$$

$$k^2(a^2 + b^2) = c^2$$

$$c = \sqrt{k^2(a^2 + b^2)}$$

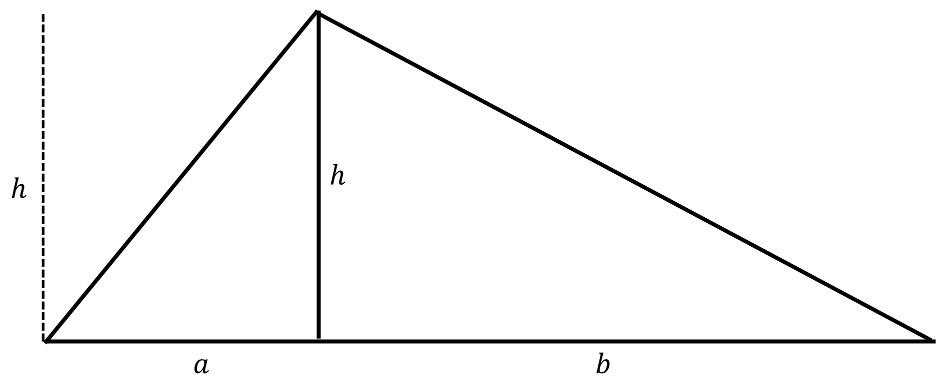
$$c = k\sqrt{a^2 + b^2}$$

This means that the ratio between any two given sides remains the same after is it was before the rescaling. E.g., $\frac{a}{b} = \frac{ak}{bk}$.

This means that resizing a right triangle does not affect the ratios between any two sides. If you draw any polygon on a set of two perpendicular axis, each side will either be on one of those axis or can be expressed as the hypotenuse of a triangle with the other two sides being parallel to those axis, which means that this should also hold true for any shape, which makes intuitive sense too because if you change its size but its shape remains the same then the ratio between any two given sides will be unchanged.

Speaking of areas of triangles, the area of a right triangle, base b and height h can be easily calculated using $Area = \frac{bh}{2}$, but what about a non-right triangle?

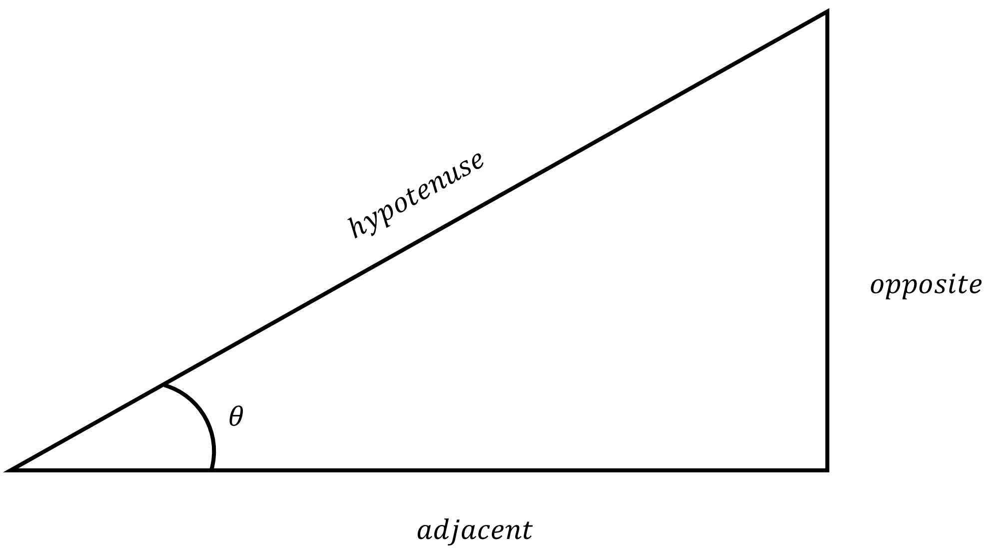
Such a triangle could be split into two right triangles as shown:



The area of this triangle $=$ the sum of the area of the other two triangles $= \frac{ah}{2} + \frac{bh}{2} = \frac{h}{2}(a + b)$ where h is the perpendicular height (as opposed to the slanted height) and $a + b$ is the length of the base. This means that $triangle\ area = \frac{1}{2} \times base \times perpendicular\ height.$

If I told you that I have a right-angled triangle with an angle of $62°$ in it, you would be able to work out the size of the third angle as you know the sizes of two angles ($62°$, and $90°$) and that angles in a triangle add to $180°$. The fact that, from these two pieces of information, you can find out everything about the shape of a triangle, (but not its size), means that you can create a function to compare the ratios of two sides in that triangle, (because ratios between sides remain unchanged by rescaling). This function would take the angle, θ as an input would output the ratio.

The following diagram shows one way in which the different sides could be labelled:



Where the hypotenuse (h), as mentioned before, is opposite the right angle, opposite (o) is the side length opposite the angle θ and adjacent (a) the side length between the angle θ and the right angle.

There are six possible ratios, so I will define six different functions.

$$sine(\theta) = \frac{o}{h}$$

$$cosine(\theta) = \frac{a}{h}$$

$$tangent(\theta) = \frac{o}{a}$$

$$cotangent(\theta) = \frac{a}{o}$$

$$secant(\theta) = \frac{h}{a}$$

$$cosecant(\theta) = \frac{h}{o}$$

I will write these from now on as: $\sin(\theta)$, $\cos(\theta)$, $\tan(\theta)$, $\cot(\theta)$, $\sec(\theta)$, $\csc(\theta)$

From these ratios, the following identities can be derived:

$$\cot(\theta) = \frac{a}{o} = 1 \div \frac{o}{a} = \frac{1}{\tan(\theta)}$$

$$\sec(\theta) = \frac{h}{a} = 1 \div \frac{a}{h} = \frac{1}{\cos(\theta)}$$

$$\csc(\theta) = \frac{h}{o} = 1 \div \frac{o}{h} = \frac{1}{\sin(\theta)}$$

$$\frac{\sin(\theta)}{\cos(\theta)} = \frac{o}{h} \div \frac{a}{h} = \frac{o}{h} \times \frac{h}{a} = \frac{oh}{ha} = \frac{o}{a} = \tan(\theta)$$

$$\frac{\cos(\theta)}{\sin(\theta)} = \frac{a}{h} \div \frac{o}{h} = \frac{a}{h} \times \frac{h}{o} = \frac{ah}{ho} = \frac{a}{o} = \cot(\theta)$$

$$\cos^2(\theta) + \sin^2(\theta) = \left(\frac{a}{h}\right)^2 + \left(\frac{o}{h}\right)^2 = \frac{a^2}{h^2} + \frac{o^2}{h^2} = \frac{o^2 + a^2}{h^2} = 1$$

Because $o^2 + a^2 = h^2$ because of the Pythagorean theorem.

Also, the notation: $f^2(x)$ means $\left(f(x)\right)^2$.

We can use these to obtain some more:

$$\cos^2(\theta) + \sin^2(\theta) = 1$$

$$\frac{\cos^2(\theta) + \sin^2(\theta)}{\cos^2(\theta)} = \frac{1}{\cos^2(\theta)}$$

$$\frac{\cos^2(\theta)}{\cos^2(\theta)} + \frac{\sin^2(\theta)}{\cos^2(\theta)} = \left(\frac{1}{\cos(\theta)}\right)^2$$

$$1 + \left(\frac{\sin(\theta)}{\cos(\theta)}\right)^2 = \sec^2(\theta)$$

$$1 + \tan^2(\theta) = \sec^2(\theta)$$

and

$$\cos^2(\theta) + \sin^2(\theta) = 1$$

$$\frac{\cos^2(\theta) + \sin^2(\theta)}{\sin^2(\theta)} = \frac{1}{\sin^2(\theta)}$$

$$\frac{\cos^2(\theta)}{\sin^2(\theta)} + \frac{\sin^2(\theta)}{\sin^2(\theta)} = \left(\frac{1}{\sin(\theta)}\right)^2$$

$$\left(\frac{\cos(\theta)}{\sin(\theta)}\right)^2 + 1 = \csc^2(\theta)$$

$$\cot^2(\theta) + 1 = \csc^2(\theta)$$

I will create a few basic right angle triangles with known side lengths and will use these to find the $sines$ and $cosines$ of certain angles, the above identities can then be used to find the $tangent$, $cotangent$, $secant$ and $cosecant$.

I will create a symmetrical triangle, meaning that the non-hypotenuse side lengths (a and b) are equal, making the hypotenuse (c):

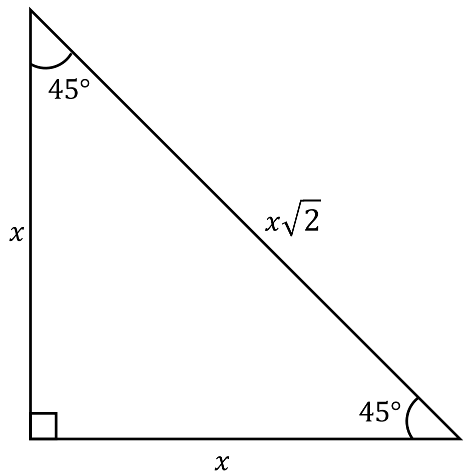
$$c = \sqrt{a^2 + b^2} = \sqrt{a^2 + a^2} = \sqrt{2a^2} = \sqrt{2}\sqrt{a^2} = a\sqrt{2}$$

I will call the length of the non-hypotenuse sides x, meaning the hypotenuse has length $x\sqrt{2}$. Because the triangle is symmetrical, the two angles (which aren't the right angle) will both be equal. I will call this angle θ and because angles in a triangle add to $180°$:

$$\theta + \theta + 90 = 180$$

$$2\theta = 90$$

$$\theta = 45$$



Such a triangle (one which is symmetrical) is known as an isosceles triangle.

Here, the $opposite = x$, $adjacent = x$, $hypotenuse = x\sqrt{2}$, so:

$$\sin(45) = \frac{opposite}{hypotenuse} = \frac{x}{x\sqrt{2}} = \frac{1}{\sqrt{2}} = \frac{1\sqrt{2}}{\sqrt{2}\sqrt{2}} = \frac{\sqrt{2}}{2}$$

$$\cos(45) = \frac{adjacent}{hypotenuse} = \frac{x}{x\sqrt{2}} = \frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2}$$

$$\tan(45) = \frac{\sin(45)}{\cos(45)} = \frac{\sqrt{2}}{2} \div \frac{\sqrt{2}}{2} = 1$$

$$\cot(45) = \frac{1}{\tan(45)} = \frac{1}{1} = 1$$

$$\sec(45) = \frac{1}{\cos(45)} = 1 \div \frac{1}{\sqrt{2}} = \sqrt{2}$$

$$\csc(45) = \frac{1}{\sin(45)} = 1 \div \frac{1}{\sqrt{2}} = \sqrt{2}$$

The next triangle I will draw is an equilateral triangle (a triangle with all sides and angles being equal) which I will then cut in half to create a right angle. I will call each side of the equilateral x. The side which is cut in half will therefore have a length of $\frac{x}{2}$. The right triangle created will have a hypotenuse of x and a base of $\frac{x}{2}$ so the other side I will call b can be found by:

$$\left(\frac{x}{2}\right)^2 + b^2 = x^2$$

$$\frac{x^2}{2^2} + b^2 = x^2$$

$$\frac{x^2}{4} + b^2 = x^2$$

$$b^2 = x^2 - \frac{x^2}{4}$$

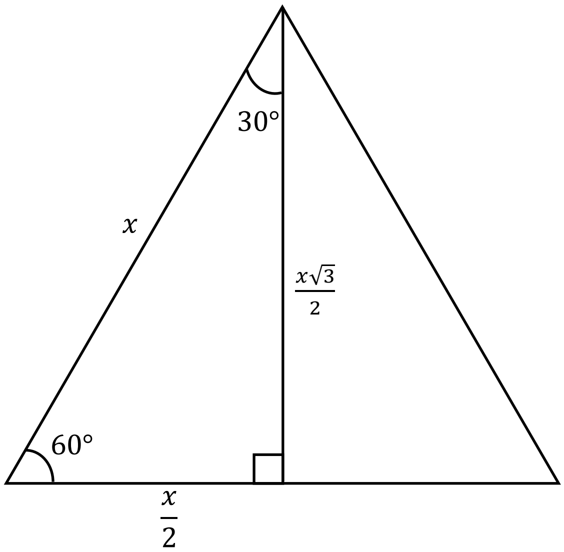
$$b^2 = x^2\left(1 - \frac{1}{4}\right)$$

$$b^2 = x^2\left(\frac{3}{4}\right)$$

$$b = \sqrt{x^2\left(\frac{3}{4}\right)}$$

$$b = x\left(\frac{\sqrt{3}}{2}\right)$$

As for the angles, the angles in an equilateral triangle are all equal and angles in a triangle add to 180° so each angle is $\frac{180°}{3} = 60°$. When one of these angles is bisected as the triangle is cut in half, it becomes 30°. The diagram is shown here:



In this diagram, for 60°, $hypotenuse = x$, $opposite = \frac{x\sqrt{3}}{2}$, $adjacent = \frac{x}{2}$.

$$\sin(60) = \frac{opposite}{hypotenuse} = \frac{x\sqrt{3}}{2} \div x = \frac{\sqrt{3}}{2}$$

$$\cos(60) = \frac{adjacent}{hypotenuse} = \frac{x}{2} \div x = \frac{1}{2}$$

$$\tan(60) = \frac{\sin(60)}{\cos(60)} = \frac{\sqrt{3}}{2} \div \frac{1}{2} = \frac{\sqrt{3}}{2} \times 2 = \sqrt{3}$$

$$\cot(60) = \frac{1}{\tan(60)} = \frac{1}{\sqrt{3}}$$

$$\sec(60) = \frac{1}{\cos(60)} = 1 \div \frac{1}{2} = 2$$

$$\csc(60) = \frac{1}{\sin(60)} = 1 \div \frac{\sqrt{3}}{2} = \frac{2}{\sqrt{3}} = \frac{2\sqrt{3}}{\sqrt{3}\sqrt{3}} = \frac{2\sqrt{3}}{3}$$

For 30°: $hypotenuse = x, opposite = \frac{x}{2}, adjacent = \frac{x\sqrt{3}}{2}$.

$$\sin(30) = \frac{opposite}{hypotenuse} = \frac{x}{2} \div x = \frac{1}{2}$$

$$\cos(30) = \frac{adjacent}{hypotenuse} = \frac{x\sqrt{3}}{2} \div x = \frac{\sqrt{3}}{2}$$

$$\tan(30) = \frac{\sin(30)}{\cos(30)} = \frac{1}{2} \div \frac{\sqrt{3}}{2} = \frac{1}{2} \times \frac{2}{\sqrt{3}} = \frac{1}{\sqrt{3}} = \frac{1\sqrt{3}}{\sqrt{3}\sqrt{3}} = \frac{\sqrt{3}}{3}$$

$$\cot(30) = \frac{1}{\tan(30)} = 1 \div \frac{1}{\sqrt{3}} = \sqrt{3}$$

$$\sec(30) = \frac{1}{\cos(30)} = 1 \div \frac{\sqrt{3}}{2} = \frac{2}{\sqrt{3}} = \frac{2\sqrt{3}}{\sqrt{3}\sqrt{3}} = \frac{2\sqrt{3}}{3}$$

$$\csc(30) = \frac{1}{\sin(30)} = 1 \div \frac{1}{2} = 2$$

You may have noticed that $\sin(30) = \cos(60)$ and that $\cos(30) = \sin(60)$. This is neither a coincidence, nor a surprise because the hypotenuse for both angles is the same, the opposite of one is the adjacent of the other and the adjacent of one is the opposite of the other. The means that the opposites and adjacents are switched around meaning the *cosines* and *sines* are switched, as well as the *secant* and *cosecant* and *tangent* and *cotangent* both become their reciprocals (the reciprocal of $\frac{a}{b}$ is $\frac{b}{a}$). This happens with 30 and 60 because they both exist as different angles in the same right triangle. If two angles θ_1 and θ_2 both exist in the same right triangle, then $\theta_1 + \theta_2 + 90 = 180$ so $\theta_1 + \theta_2 = 90$ meaning $\theta_1 = 90 - \theta_2$ and $\theta_2 = 90 - \theta_1$ so

$$\sin(90 - \theta) = \cos(\theta)$$

$$\cos(90 - \theta) = \sin(\theta)$$

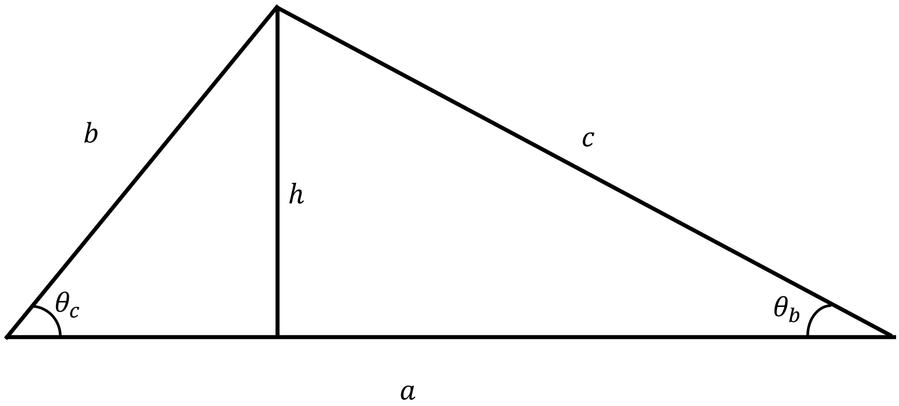
$$\sec(90 - \theta) = \csc(\theta)$$

$$\csc(90 - \theta) = \sec(\theta)$$

$$\tan(90 - \theta) = \cot(\theta)$$

$$\cot(90 - \theta) = \tan(\theta)$$

These formulae which link the trigonometric ratios in these pairs is why their names come in pairs as well, $sine$ with $cosine$, $tangent$ with $cotangent$ and $secant$ with $cosecant$.



Now that we have a firmer understand of what these functions are, how can they be used?

A triangle has side lengths a, b and c, perpendicular height h and angles of θ_a, θ_b and θ_c opposite sides a, b and c respectively as shown in this diagram.

$$\sin(\theta_c) = \frac{h}{b}$$

$$b\sin(\theta_c) = h$$

$$\sin(\theta_b) = \frac{h}{c}$$

$$c\sin(\theta_b) = h$$

$$b\sin(\theta_c) = c\sin(\theta_b)$$

$$\frac{b}{\sin(\theta_b)} = \frac{c}{\sin(\theta_c)}$$

The perpendicular line could have been drawn perpendicular to c and the algebra repeated to show that $\frac{a}{\sin(\theta_a)} = \frac{b}{\sin(\theta_b)}$, So:

$$\frac{a}{\sin(\theta_a)} = \frac{b}{\sin(\theta_b)} = \frac{c}{\sin(\theta_c)}$$

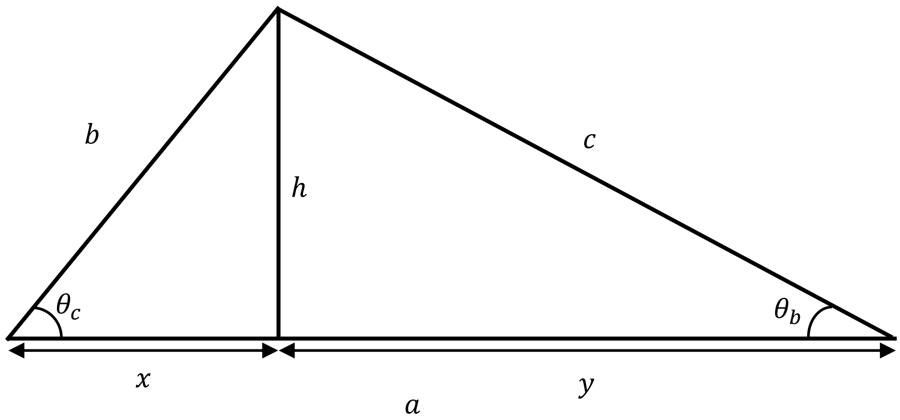
This is known as the sine rule.

The area of the above triangle can be found by $Area = \frac{ah}{2}$.

$$b\sin(\theta_c) = h$$

$$Area = \frac{ah}{2} = \frac{ab\sin(\theta_c)}{2}$$

$$Area = \frac{1}{2}ab\sin(\theta_c)$$



So this formula can be used to find the area of any triangle given an angle and the two side lengths adjacent to it.

By splitting the side a into x and y this diagram is created where $a = x + y$ so $y = a - x$.

$$x^2 + h^2 = b^2$$

$$h^2 = b^2 - x^2$$

$$y^2 + h^2 = c^2$$

$$h^2 = c^2 - y^2$$

$$b^2 - x^2 = c^2 - y^2$$

$$b^2 - x^2 + y^2 = c^2$$

$$b^2 - x^2 + (a - x)^2 = c^2$$

$$b^2 - x^2 + a^2 - 2ax + x^2 = c^2$$

$$c^2 = a^2 + b^2 - 2ax$$

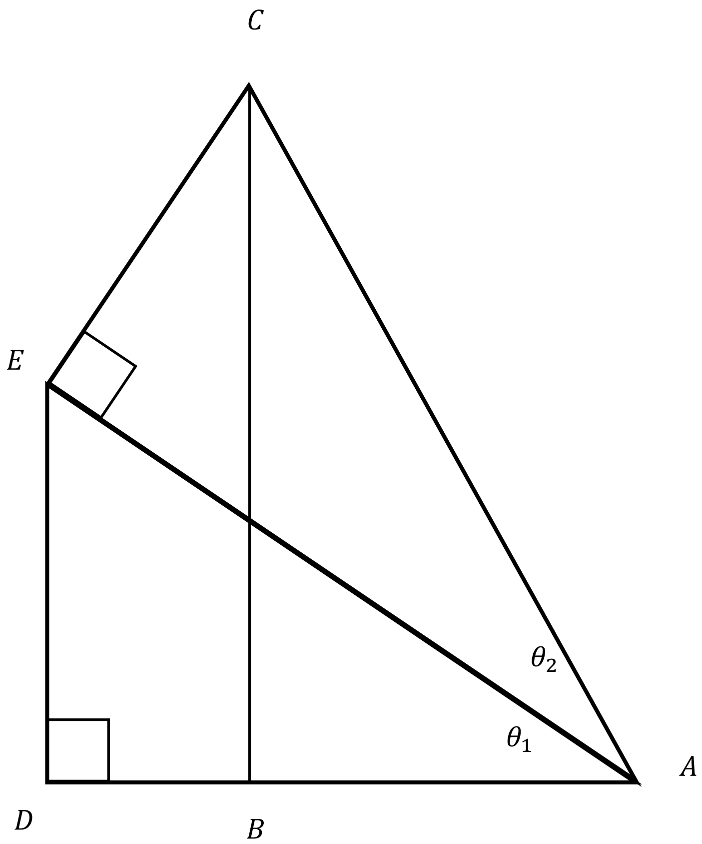
$$\cos(\theta_c) = \frac{x}{b}$$

$$b\cos(\theta_c) = x$$

$$c^2 = a^2 + b^2 - 2ab\cos(\theta_c)$$

This is called the cosine rule.

The next thing I would like to discuss in terms of trigonometry is how I could find the $sine$, $cosine$ etc., of the sum of two values in terms of those values e.g., finding $\sin(\theta_1 + \theta_2)$ in terms of θ_1 and θ_2. To do this I will need to create a triangle with an angle of $\theta_1 + \theta_2$ and work out its side lengths in



terms of θ_1 and θ_2. To do this, I will need angles θ_1 and θ_2 exist in their own triangles. What I mean should be made clear by the diagram:

It is worth noting that when A, B and C represent points on a diagram as they do here, ABC means the triangle with vertices A, B and C, the line segment going from A to B is written as AB and the angle at point B between lines AB and BC is written as $A\hat{B}C$. So AB does not mean $A \times B$, nor does ABC mean $A \times B \times C$ in this context.

In the diagram, ABC is a right triangle with an angle $\theta_1 + \theta_2$, ADE is a right triangle with an angle θ_1 and AEC is a right triangle with an angle θ_2. This diagram proved no sense of scale and so no side lengths can be calculated, but because we are only concerned with the *ratios* between side lengths, I can let and side length equal x, or I could just choose a number as it makes no difference what the value of x is, so I will $Let\ AC = 1$ as 1 is an easy number to work with.

I must now find the lengths of AB, AC and BC as these are the sides of the triangle ABC.

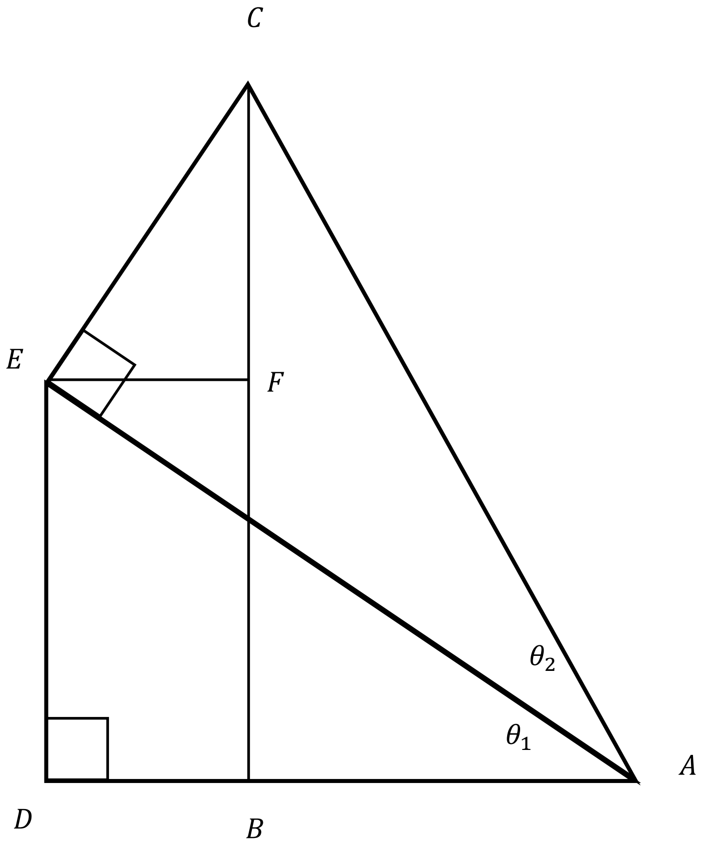
AC is 1, so that one was easy. $AB = AD - BD$.

Looking at triangle ACE: $\sin(\theta_2) = \frac{CE}{AC} = \frac{CE}{1} = CE$, so $CE = \sin(\theta_2)$

and $\cos(\theta_2) = \frac{AE}{AC} = \frac{AE}{1} = AE$, so $AE = \cos(\theta_2)$

Looking at triangle ADE: $\sin(\theta_1) = \frac{DE}{AE} = \frac{DE}{\cos(\theta_2)}$, so $DE = \sin(\theta_1)\cos(\theta_2)$

and $\cos(\theta_1) = \frac{AD}{AE} = \frac{AD}{\cos(\theta_2)}$, so $AD = \cos(\theta_1)\cos(\theta_2)$



$$AB = AD - BD = \cos(\theta_1)\cos(\theta_2) - BD$$

To find BD I will add another line to the diagram.

This new line EF has the same length as BD so I now need to find EF.

Angle $A\hat{E}F = \theta_1$ because they are alternate angles. $A\hat{E}C = 90$ and $C\hat{E}F + A\hat{E}F = A\hat{E}C$ so $C\hat{E}F + \theta_1 = 90$ meaning $C\hat{E}F = 90 - \theta_1$

$$\cos\left(C\hat{E}F\right) = \frac{EF}{CE}$$

$$\cos(90 - \theta_1) = \frac{EF}{\sin(\theta_2)}$$

$$\sin(\theta_1) = \frac{EF}{\sin(\theta_2)}$$

$$EF = \sin(\theta_1)\sin(\theta_2)$$

$$BD = \sin(\theta_1)\sin(\theta_2)$$

$$AB = \cos(\theta_1)\cos(\theta_2) - BD = \cos(\theta_1)\cos(\theta_2) - \sin(\theta_1)\sin(\theta_2)$$

Now I need to find the length of BC.

Notice $BC = BF + CF$ and $BF = DE$ so $BC = DE + CF = \sin(\theta_1)\cos(\theta_1) + CF$.

$$\sin(C\hat{E}F) = \frac{CF}{CE}$$

$$\sin(90 - \theta_1) = \frac{CF}{\sin(\theta_2)}$$

$$\cos(\theta_1) = \frac{CF}{\sin(\theta_2)}$$

$$CF = \sin(\theta_2)\cos(\theta_1)$$

$$BC = \sin(\theta_1)\cos(\theta_1) + CF = \sin(\theta_1)\cos(\theta_1) + \sin(\theta_2)\cos(\theta_1)$$

Now that I know AB, AC and BC in terms of θ_1 and θ_2 I can find the $sines$, $cosines$ etc., of $\theta_1 + \theta_2$ using triangle ABC.

$$\sin(B\hat{A}C) = \frac{BC}{AC}$$

$$\sin(\theta_1 + \theta_2) = \frac{\sin(\theta_1)\cos(\theta_1) + \sin(\theta_2)\cos(\theta_1)}{1}$$

$$\sin(\theta_1 + \theta_2) = \sin(\theta_1)\cos(\theta_1) + \sin(\theta_2)\cos(\theta_1)$$

$$\cos(B\hat{A}C) = \frac{AB}{AC}$$

$$\cos(\theta_1 + \theta_2) = \frac{\cos(\theta_1)\cos(\theta_2) - \sin(\theta_1)\sin(\theta_2)}{1}$$

$$\cos(\theta_1 + \theta_2) = \cos(\theta_1)\cos(\theta_2) - \sin(\theta_1)\sin(\theta_2)$$

These are known as the addition formulae and are more commonly written as:

$$\sin(A + B) = \sin(A)\cos(B) + \sin(B)\cos(A)$$

$$\cos(A + B) = \cos(A)\cos(B) - \sin(A)\sin(B)$$

We can use these to find a formula for $\tan(A + B)$

$$\tan(A + B) = \frac{\sin(A + B)}{\cos(A + B)} = \frac{\sin(A)\cos(B) + \sin(B)\cos(A)}{\cos(A)\cos(B) - \sin(A)\sin(B)}$$

Whilst this is a formula, it doesn't look particularly nice. Perhaps there is way to make it look nicer, maybe making it in terms of $\tan(A)$ and $\tan(B)$.

$\sin(A)\cos(B)$ could be divided by $\cos(B)$ to get rid of that part and could be divided by $\cos(A)$ to turn $\sin(A)$ into $\tan(A)$. Similarly, $\sin(B)\cos(A)$ could be divided by $\cos(A)$ to get rid of that part and then divided by $\cos(B)$ to turn $\sin(B)$ into $\tan(B)$. Dividing by $\cos(A)\cos(B)$ is the same as multiplying by $\frac{1}{\cos(A)\cos(B)}$ and so if the top of the quotient (fraction) is multiplied by this, the bottom must be as well. $\frac{\cos(A)\cos(B)}{\cos(A)\cos(B)} = 1$ and $\frac{\sin(A)\sin(B)}{\cos(A)\cos(B)} = \frac{\sin(A)}{\cos(A)} \times \frac{\sin(B)}{\cos(B)} = \tan(A)\tan(B)$, so:

$$\tan(A+B) = \frac{\sin(A)\cos(B) + \sin(B)\cos(A)}{\cos(A)\cos(B) - \sin(A)\sin(B)}$$

$$\tan(A+B) = \frac{\left(\frac{\sin(A)\cos(B) + \sin(B)\cos(A)}{\cos(A)\cos(B)}\right)}{\left(\frac{\cos(A)\cos(B) - \sin(A)\sin(B)}{\cos(A)\cos(B)}\right)}$$

$$\tan(A+B) = \frac{\left(\frac{\sin(A)\cos(B)}{\cos(A)\cos(B)}\right) + \left(\frac{\sin(B)\cos(A)}{\cos(A)\cos(B)}\right)}{\left(\frac{\cos(A)\cos(B)}{\cos(A)\cos(B)}\right) - \left(\frac{\sin(A)\sin(B)}{\cos(A)\cos(B)}\right)}$$

$$\tan(A+B) = \frac{\tan(A) + \tan(B)}{1 - \tan(A)\tan(B)}$$

The $cotangent$, $secant$ and $cosecant$ versions of these formulae can be found by taking the reciprocals of the $tangent$, $cosine$ and $sine$ versions respectively.

From these addition formulae, we can derive the "double angle formulae" where A and B are the same.

$$\sin(2x) = \sin(x+x) = \sin(x)\cos(x) + \sin(x)\cos(x) = 2\sin(x)\cos(x)$$

$$\cos(2x) = \cos(x+x) = \cos(x)\cos(x) - \sin(x)\sin(x) = \cos^2(x) - \sin^2(x)$$

$$\tan(2x) = \tan(x+x) = \frac{\tan(x) + \tan(x)}{1 - \tan(x)\tan(x)} = \frac{2\tan(x)}{1 - \tan^2(x)}$$

Again, the cot, sec and csc can be found by using the reciprocals of these.

These trigonometric functions can take any input between 0° and 90° (exclusive) because these are the only values which an angle in a right triangle can take. (Exclusive means not including 0° and 90°).

These formulae can be used to extent these trigonometric functions beyond this range. For example:

$$\sin(90) = \sin(2 \times 45) = 2\sin(45)\cos(45) = 2 \times \frac{\sqrt{2}}{2} \times \frac{\sqrt{2}}{2} = 2 \times \frac{2}{4} = 2 \times \frac{1}{2} = 1$$

This does not "prove" that $\sin(90) = 1$ necessarily, but it does provide a justification as to why it should be defined that way.

Likewise:

$$\cos(90) = \cos(2 \times 45) = \cos^2(45) - \sin^2(45) = \left(\frac{\sqrt{2}}{2}\right)^2 - \left(\frac{\sqrt{2}}{2}\right)^2 = 0$$

We can take this even further:

$$\sin(180) = \sin(2 \times 90) = 2\sin(90)\cos(90) = 2 \times 1 \times 0 = 0$$

$$\cos(180) = \cos(2 \times 90) = \cos^2(90) - \sin^2(90) = 0^2 - 1^2 = -1$$

$$\sin(270) = \sin(180 + 90) = \sin(180)\cos(90) + \sin(90)\cos(180) = 0 \times 0 + 1 \times -1 = -1$$

$$\cos(270) = \cos(180 + 90) = \cos(180)\cos(90) - \sin(180)\sin(90) = -1 \times 0 - 0 \times 1 = 0$$

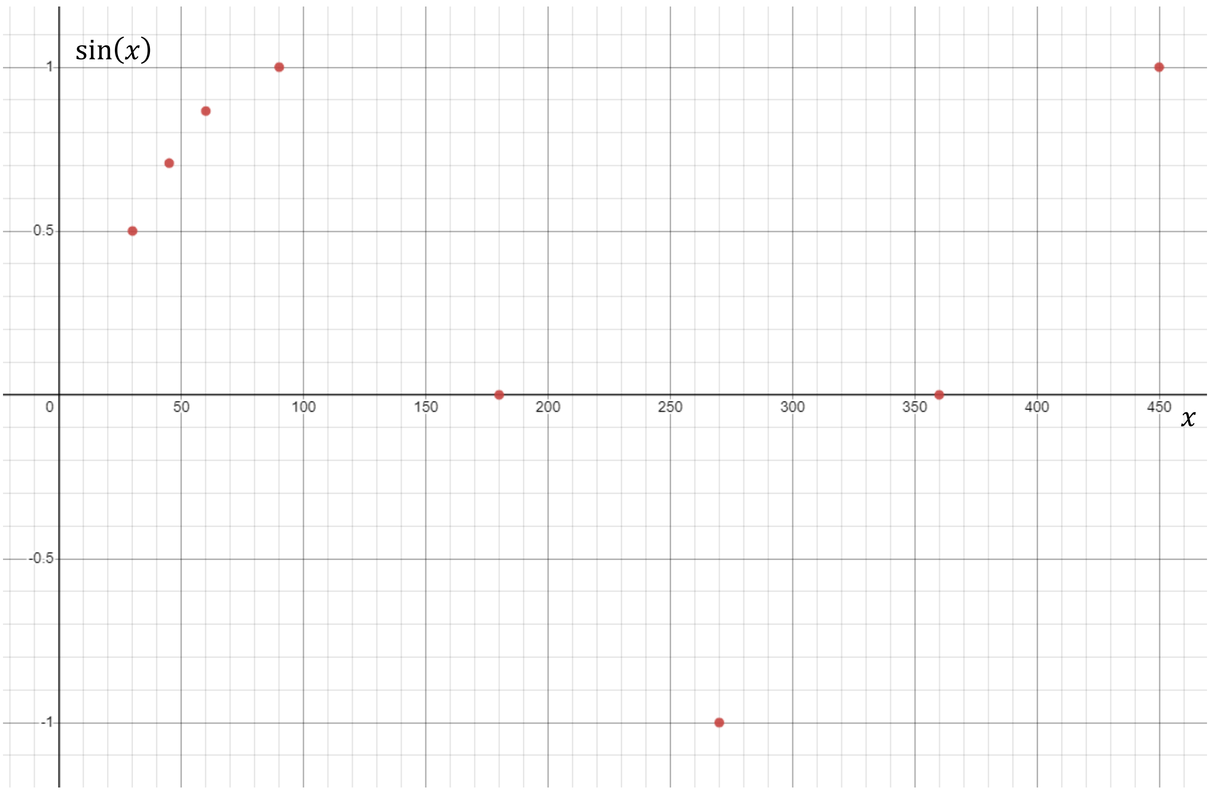
$$\sin(360) = \sin(2 \times 180) = 2\sin(180)\cos(180) = 2 \times 0 \times -1 = 0$$

$$\cos(360) = \cos(2 \times 180) = \cos^2(180) - \sin^2(180) = (-1)^2 - 0^2 = 1$$

$$\sin(450) = \sin(360 + 90) = \sin(360)\cos(90) + \sin(90)\cos(360) = 0 \times 0 + 1 \times 1 = 1$$

$$\cos(450) = \cos(360 + 90) = \cos(360)\cos(90) - \sin(360)\sin(90) = 1 \times 0 - 0 \times 1 = 0$$

Plotting the values we have for $sine$:



$$\sin(30) = \frac{1}{2}, \qquad \sin(45) = \frac{\sqrt{2}}{2}, \qquad \sin(60) = \frac{\sqrt{3}}{2}, \qquad \sin(90) = 1, \qquad \sin(180) = 0,$$
$$\sin(270) = -1, \qquad \sin(360) = 0, \qquad \sin(450) = 1$$

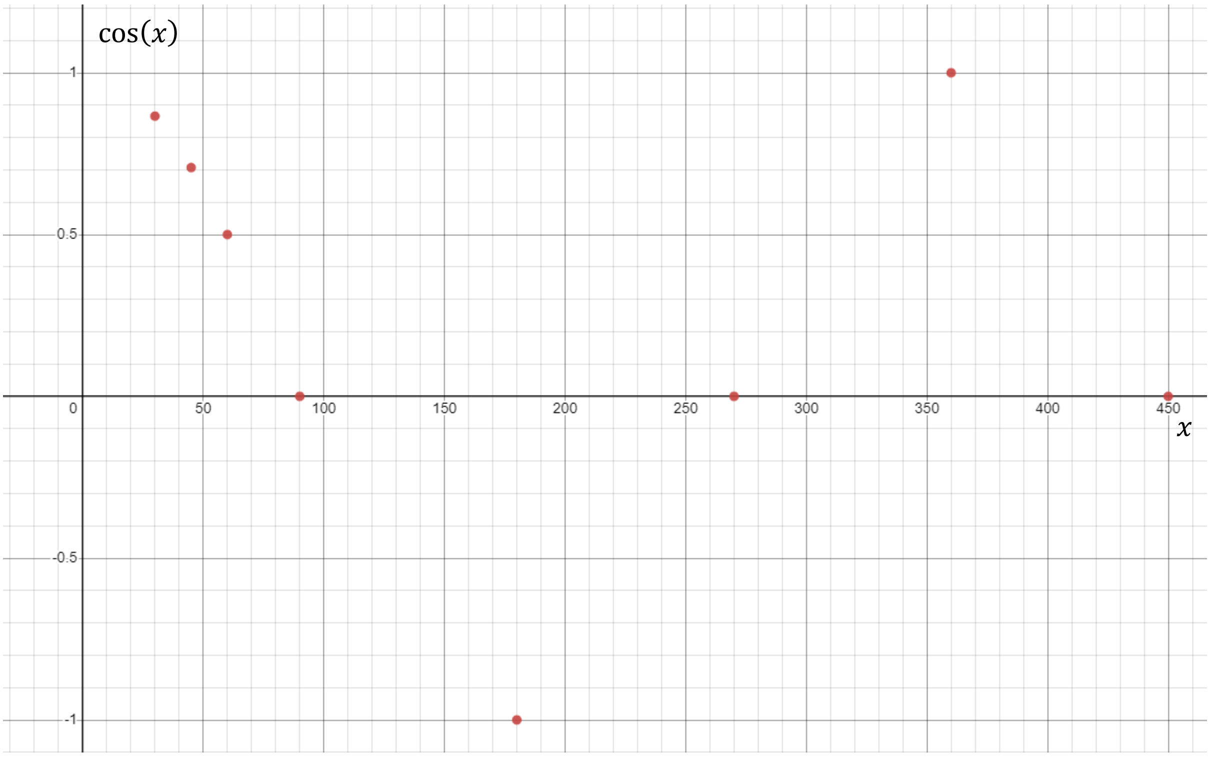
This each point has a horizontal (θ-coordinate in this case) representing a value for θ and a vertical position which represents the value of $\sin(\theta)$ for that value of θ. If we knew the value of $\sin(\theta)$ for every possible value of θ we could draw a curve which describes every point. It seems as though the curve is flipped after $180°$ and is then flipped back at $360°$. It seems that every $180°$, the curve is reflected in the x-axis. If this is true, then $\sin(x + 180) = -\sin(x)$. We can easily work out if this is true or not by using the addition formulae.

$$\sin(x + 180) = \sin(x)\cos(180) + \sin(180)\cos(x) = \sin(x) \times -1 + 0 \times \cos(x) = -\sin(x)$$

So $\sin(x + 180) = -\sin(x)$ meaning the curve is reflected in the x-axis (is flipped upside-down) every $180°$.

Plotting the values we have for $cosine$:

$$\cos(30) = \frac{\sqrt{3}}{2}, \qquad \cos(45) = \frac{\sqrt{2}}{2}, \qquad \cos(60) = \frac{1}{2}, \qquad \cos(90) = 0, \qquad \cos(180) = -1,$$
$$\cos(270) = 0, \qquad \cos(360) = 1, \qquad \cos(450) = 0$$



It also seems here that the curve is reflected once every $180°$, meaning $\cos(x + 180) = \cos(x)$. To verify this:

$$\cos(x + 180) = \cos(x)\cos(180) - \sin(x)\sin(180) = \cos(x) \times -1 + \sin(x) \times 0 = -\cos(x)$$

So, the curve is indeed reflected every $180°$.

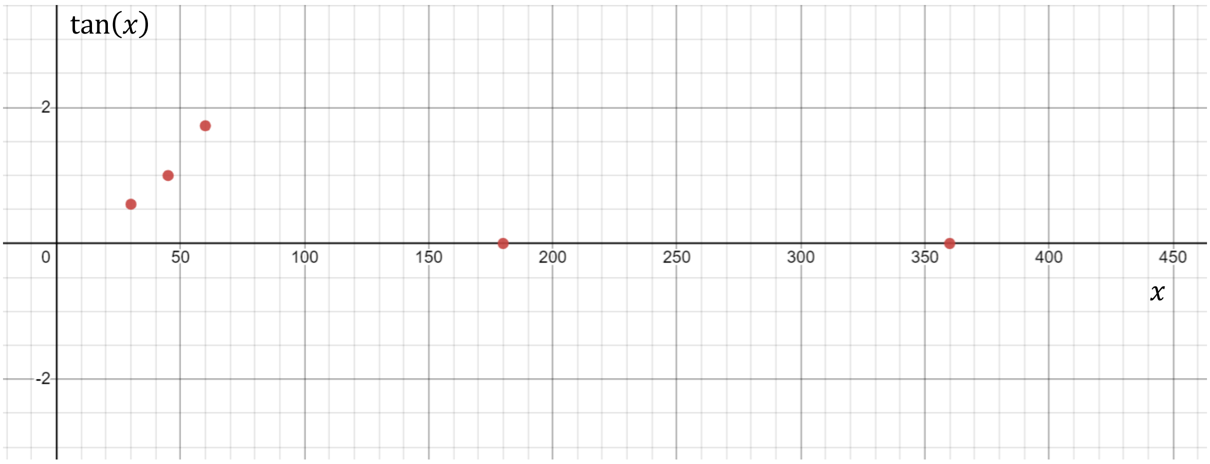
Finally, plotting the value for $tangent$:

$$\tan(30) = \frac{\sqrt{3}}{3}, \qquad \tan(45) = 1, \qquad \tan(60) = \sqrt{3}, \qquad \tan(90) = \frac{\sin(90)}{\cos(90)} = \frac{1}{0},$$

$$\tan(180) = \frac{\sin(180)}{\cos(180)} = \frac{0}{-1} = 0, \qquad \tan(270) = \frac{\sin(270)}{\cos(270)} = \frac{-1}{0},$$

$$\tan(360) = \frac{\sin(360)}{\cos(360)} = \frac{0}{1} = 0, \qquad \tan(450) = \frac{\sin(450)}{\cos(450)} = \frac{1}{0}$$

I will ignore the points with division by 0 for now but I will talk more about them later on.



This graph gives much less information than the other two, as some of the points could not be plotted. Perhaps this graph is also reflected in the x-axis once every 180°.

$$\tan(180 + x) = \frac{\sin(180 + x)}{\cos(180 + x)} = \frac{-\sin(x)}{-\cos(x)} = \frac{\sin(x)}{\cos(x)} = \tan(x)$$

Meaning this graph actually repeats once every 180° and does not reflect.

I will fill the gaps in these graphs later, when I have the tools required to calculate the value of $sine$, $cosine$ and $tangent$ of any values. I will then use these graphs to find out what the graphs of $cotangent$, $secant$ and $cosecant$ should look like.

How should we choose to define the $sine$, $cosine$ and $tangent$ function for negative inputs?

We know that $\sin(90 - x) = \cos(x)$ and

$$\sin(90 - x) = \sin\big(90 + (-x)\big)$$

$$= \sin(90)\cos(-x) + \sin(-x)\cos(90)$$

$$= 1 \times \cos(-x) + \sin(-x) \times 0$$

$$= \cos(-x)$$

So

$$\cos(x) = \cos(-x)$$

We also know that $\cos(90 - x) = \sin(x)$ and

$$\cos(90 - x) = \cos\big(90 + (-x)\big)$$

$$= \cos(90)\cos(-x) - \sin(90)\sin(-x)$$

$$= 0 \times \cos(-x) - 1 \times \sin(-x) =$$

$$-\sin(-x)$$

$$\sin(x) = -\sin(-x)$$

$$-\sin(x) = \sin(-x)$$

For $tangent$:

$$\tan(-x) = \frac{\sin(-x)}{\cos(-x)} = \frac{-\sin(x)}{\cos(x)} = -\tan(x)$$

So this is how $sines$ and $cosines$ of negative angles should be defined. Let's think about what this would look like on the graphs.

$$\cos(-x) = \cos(x)$$

This means that for some negative number x, the its $cosine$ equals the $cosine$ of its positive counterpart. This goes for all negative values of x meaning the graph should be mirrored in the y-axis.

$$\sin(-x) = -\sin(x)$$

This means that for some negative number x, its $sine$ should be the negative of the $sine$ of that numbers positive counterpart. In other words, the left of the y-axis should be the same as what is on its right, but upside-down. The same can be said for $tangent$ because $\tan(-x) = -\tan(x)$.

These negative angles now allow the addition formulae to be used as subtraction formulae.

$$\sin(A - B)$$

$$= \sin\big(A + (-B)\big)$$

$$= \sin(A)\cos(-B) + \sin(-B)\cos(A)$$

$$= \sin(A)\cos(B) - \sin(B)\cos(A)$$

$$\cos(A - B)$$

$$= \cos(A + (-B))$$

$$= \cos(A)\cos(-B) - \sin(A)\sin(-B)$$

$$= \cos(A)\cos(B) + \sin(A)\sin(B)$$

$$\tan(A - B)$$

$$= \tan(A + (-B))$$

$$= \frac{\tan(A) + \tan(-B)}{1 - \tan(A)\tan(-B)}$$

$$= \frac{\tan(A) - \tan(B)}{1 + \tan(A)\tan(B)}$$

These formulae are almost identical to the addition formulae, the only difference being plusses and minuses.

$$\sin(A + B) = \sin(A)\cos(B) + \sin(B)\cos(A)$$

$$\sin(A - B) = \sin(A)\cos(B) - \sin(B)\cos(A)$$

These can be written as one formula using $\pm$ notation:

$$\sin(A \pm B) = \sin(A)\cos(B) \pm \sin(B)\cos(A)$$

Which means replace all occurrences of $\pm$ with either $+$ or – (they must all be replaced with the same one).

$$\cos(A + B) = \cos(A)\cos(B) - \sin(A)\sin(B)$$

$$\cos(A - B) = \cos(A)\cos(B) + \sin(A)\sin(B)$$

$$\cos(A \pm B) = \cos(A)\cos(B) \mp \sin(A)\sin(B)$$

Which means either, replace all occurrences of $\pm$ with $+$ and all occurrences of $\mp$ with $-$, or replace all occurrences of $\pm$ with $-$ and all occurrences of $\mp$ with $+$.

$$\tan(A + B) = \frac{\tan(A) + \tan(B)}{1 - \tan(A)\tan(B)}$$

$$\tan(A - B) = \frac{\tan(A) - \tan(B)}{1 + \tan(A)\tan(B)}$$

$$\tan(A \pm B) = \frac{\tan(A) \pm \tan(B)}{1 \mp \tan(A)\tan(B)}$$

These can be used to find the values of the trigonometric functions at 0.

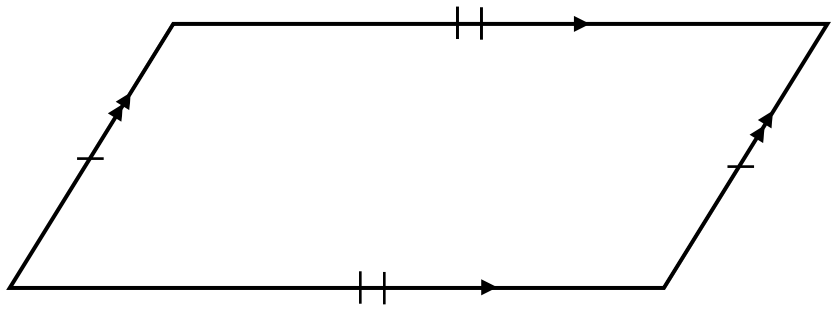
$$\sin(0) = \sin(x - x) = \sin(x)\cos(x) - \sin(x)\cos(x) = 0$$

$$\cos(0) = \cos(x - x) = \cos(x)\cos(x) + \sin(x)\sin(x) = \cos^2(x) + \sin^2(x) = 1$$

$$\tan(0) = \frac{\sin(0)}{\cos(0)} = \frac{0}{1} = 0$$

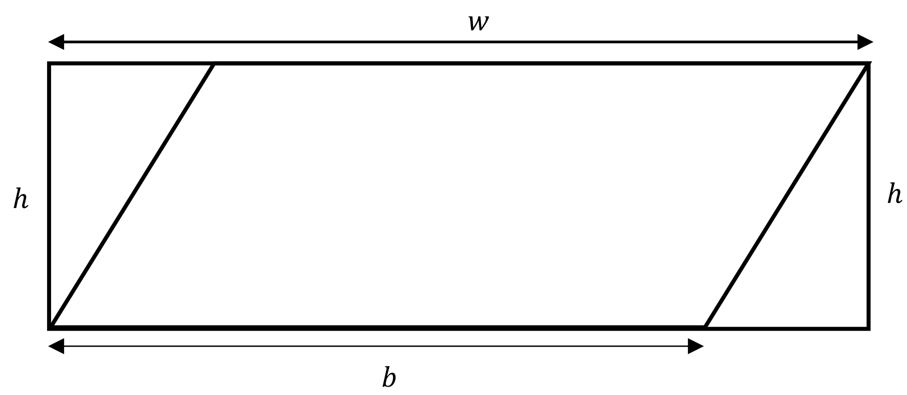
That is all from trigonometry for now. Next, I would like to discuss how to find the area of some other shapes, starting with parallelograms.

A parallelogram is a four-sided shape such that opposite sides are parallel. For example, this is a parallelogram:



The vertical and horizontal lines going through the sides show which side lengths are equal to each other and the arrows show which sides are parallel.

To find the area of this shape, we can inscribe it in a rectangle, labelling the width and height of this rectangle as w and h respectively. The base of the parallelogram, I will call b.



$$parallelogram's\ area = rectangle's\ area - triangles'\ areas$$

The $rectangle's\ area = wh$. The $base\ of\ each\ triangle = w - b$ so the $area\ of\ each\ triangle = \frac{1}{2}(w - b)h$ so the combined $triangles'areas = (w - b)h = wh - bh$

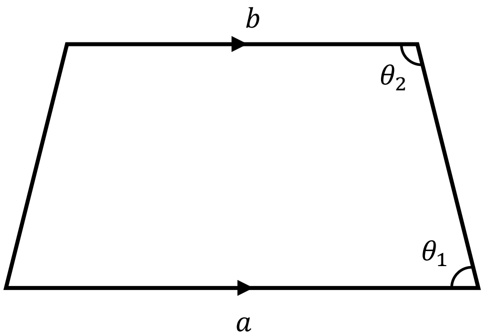
So $parallelogram's\ area = wh - (wh - bh) = wh - wh + bh = bh$

$$parallelogram's\ area = base \times perpendicular\ height$$

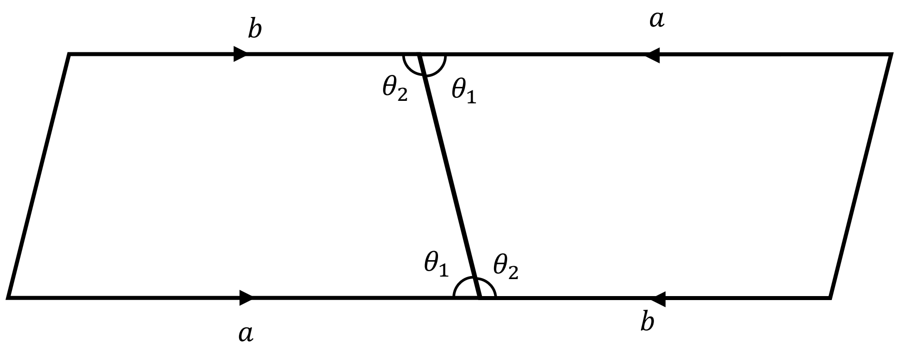
Next, I will look at trapeziums (or trapezoids as the Americans call them.)

A trapezium is like a parallelogram, but only two of its sides need to be parallel, this means that all parallelograms are also trapeziums, but not all trapeziums are parallelograms.

Here is an example of a trapezium:



I have called the lengths of the two parallel sides a and b. Notice that the two labelled angles θ_1 and θ_2 are co-interior and so add to $180°$. I will create a copy of the above trapezium and arrange the two copies as shown:

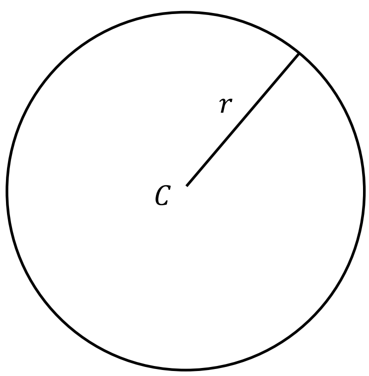


The two pieces fit together like this because the angles θ_1 and θ_2 form a straight line because they add to $180°$. Two top and bottom sides are clearly still parallel. The top and bottom sides are also of the same length $a + b$ meaning this shape is a parallelogram meaning its $Area = h(a + b)$ where h is its perpendicular height. This is the area of the parallelogram, which is two copies of the trapezium, meaning the $trapezium\ area = \frac{1}{2}h(a + b)$.

The last topic in Euclidean Geometry which I would like to discuss is circles.

First, what is a circle? A circle is defined as a shape such that each point on its perimeter is equal to a given point. The perimeter of a circle is known as its circumference. this distance is known as its radius and the given point is known as the circle's centre.

This diagram shows a circle, radius r and centre C:



The diameter of a circle is the length going from one side to the other, cutting through the centre. This means the diameter cuts the circle in half and has a length of twice the radius so $d = 2r$.

As we said before, the ratio of two lengths does not change if the whole shape is just scaled, this means that for all ratios between a circle's circumference and its radius is always the same for every circle because each circle is the same shape.

I will define this as a constant:

$$Let\ \tau = \frac{c}{r}$$

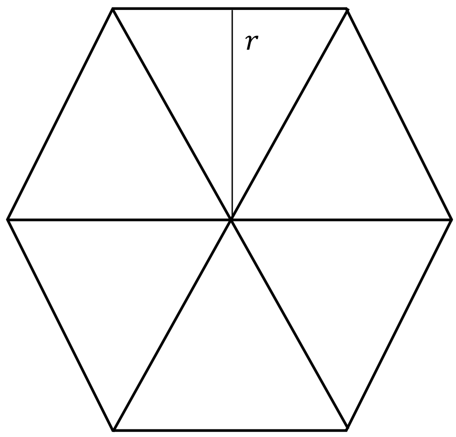
For some inexplicable reason, when first discovering facts about circles, mathematicians decided to instead use the much less natural ratio between the circumference and the diameter and used that as the "circle constant" and called it π, so $\pi = \frac{c}{d}$.

Because $\pi = \frac{c}{d}$ and $d = 2r$ then $\pi = \frac{c}{2r} = \frac{1}{2} \times \frac{c}{r} = \frac{1}{2} \times \tau = \frac{\tau}{2}$ so $\tau = 2\pi$.

Whether or not π should be replaced with τ is strongly debated among mathematicians, but I will be using π throughout this work though I might occasionally refer to τ.

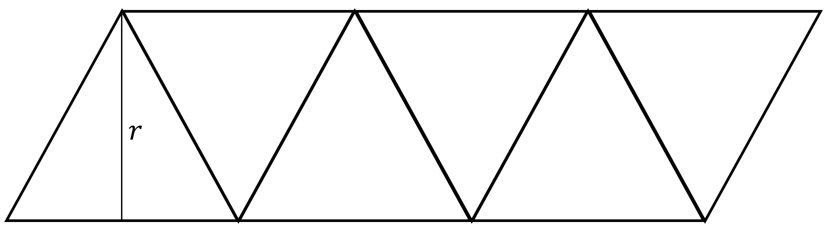
The equation $\pi = \frac{c}{d}$ can be rearranged into $c = \pi d = 2\pi r$ and so that is a formula for the circumference of a circle. If we were using τ instead then: $\tau = \frac{c}{r}$ so $c = \tau r$ which is a slightly simpler formula than using π.

If I draw a regular (all sides and angles are the same) hexagon with a "radius" r as shown:



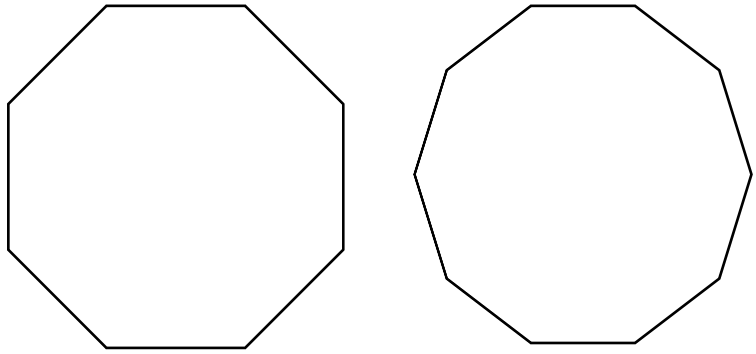
(This isn't actually a radius because it isn't a circle, but I will label it r anyway.)

The area of this hexagon could be found by adding together the areas of these triangles. The triangles could be arranged into this parallelogram:



The parallelogram has a base which is half the perimeter of the hexagon and a perpendicular height r. The area of this parallelogram, and thus the $hexagon\ area = \left(\frac{perimeter}{2}\right) r$. This trick only works if the regular polygon has an even number of sides as otherwise it could not be rearranged into a parallelogram like this.

Here is an octagon and a decagon:



As you can clearly see the more sides that a regular polygon has, the closer its area is to that of a circle. Either of these could be split into triangles which could be rearranged into a parallelogram and its area would equal $\left(\frac{perimeter}{2}\right)r$. A regular polygon with $1{,}000{,}000$ sides would have an area incredibly close to that of a circle. We could split a regular polygon with $1{,}000{,}000$ sides into $1{,}000{,}000$ triangles and form the same trapezium as before. The formula for the area is still the same, so we can say that a regular polygon of infinite sides (a circle) should have an area given by $\left(\frac{circumference}{2}\right)r$ and since $circumference = \pi d = 2\pi r$, $Area = \left(\frac{2\pi r}{2}\right)r = \pi rr = \pi r^2$ so

$$area = \pi r^2$$

Using τ instead of π: $area = \pi r^2 = \left(\frac{\tau}{2}\right)r^2 = \frac{\tau r^2}{2}$. In this case, I will admit that π certainly has a nicer formula than τ does.

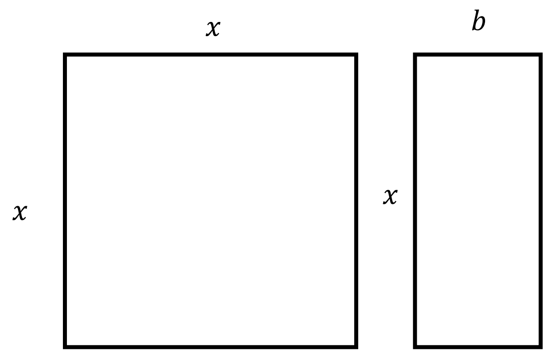
Quadratics

A polynomial is an expression consisting of a sum of monomials and a constant, where a monomial is written in the form ax^n where n is a positive integer. The order or degree of a polynomial is the highest power of x in the polynomial. For example, the polynomial $x^5 + 12x^3 - 4x^2 + 16x - 4$ has order 5 and $x^2 - 6x + 4$ has order 2. A first order (order 1) polynomial is also called a linear expression and equations involving these are very easy to solve. A general linear equation:

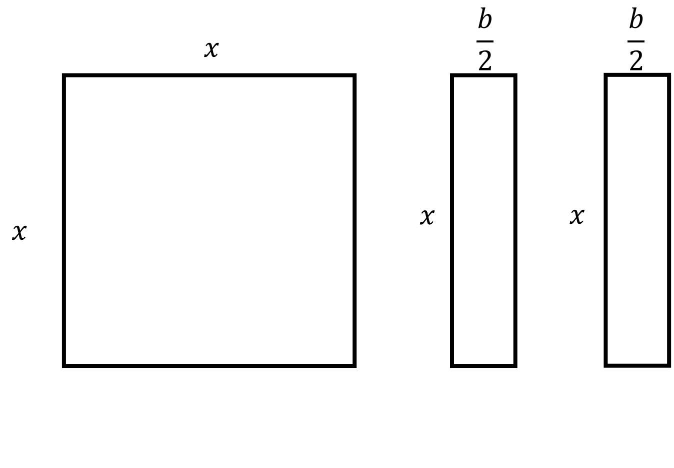
$ax + b = 0$ has a solution, $x = -\frac{b}{a}$ which can be obtained by subtracting b from both sides of the equation and then dividing both sides by a. Rearranging an equation into the form $x = \cdots$ is referred to as making x the subject of the equation. Quadratic equations (equations involving an order 2 polynomial) are more difficult to solve. The general quadratic equation looks like this:

$$ax^2 + bx + c = 0$$

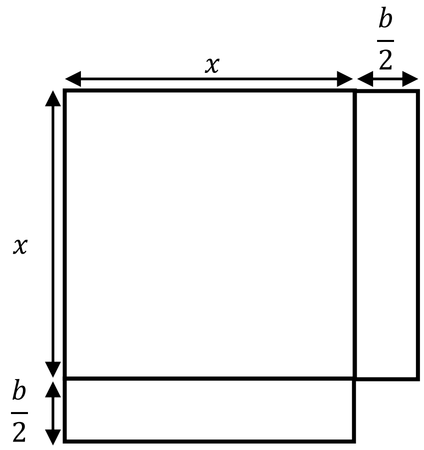
Finding a value for x for which the above equation is true will require some out of the box thinking. The problem here is that the above equation has x in two different places and it is not clear how to isolate it (get it on its own to make it the subject). To solve this problem, I will need to rewrite the quadratic expression $ax^2 + bx + c$ in a different form. It is not clear how to do this so I will start with a simpler case: $x^2 + bx$. How could I rewrite this so that there is only one x and the equation can be rearranged? One way is to think about the expression geometrically, as the sum of the areas of two shapes: a square side length x and a rectangle with side lengths x and b. The total area would be $x^2 + bx$.



These shapes can be cut up, rotated, or rearranged as none of these affect their areas. The right rectangle can be cut into two identical rectangles with height x and width $\frac{b}{2}$ as shown:



These can then be rearranged into a new shape:



This shape is just a square with a corner missing so its area is that of the larger square minus that of the smaller square. The larger square has side lengths $x+\frac{b}{2}$ and the smaller square has side lengths $\frac{b}{2}$ so the area of the shape $=\left(x+\frac{b}{2}\right)^2-\left(\frac{b}{2}\right)^2=\left(x+\frac{b}{2}\right)^2-\frac{b^2}{4}$. This area is equal to the area we started with which was x^2+bx meaning $x^2+bx=\left(x+\frac{b}{2}\right)^2-\frac{b^2}{4}$. This is known as "completing the square and can be used to solve our general quadratic equation by getting an expression in this form.

$$ax^2+bx+c=0$$

$$ax^2+bx=-c$$

$$x^2+\frac{b}{a}x=-\frac{c}{a}$$

Now complete the square on the left-hand side:

$$\left(x+\frac{b}{2a}\right)^2-\frac{\left(\frac{b}{a}\right)^2}{4}=-\frac{c}{a}$$

$$\left(x+\frac{b}{2a}\right)^2-\frac{b^2}{a^2}\div 4=-\frac{c}{a}$$

$$\left(x+\frac{b}{2a}\right)^2-\frac{b^2}{4a^2}=-\frac{c}{a}$$

$$\left(x+\frac{b}{2a}\right)^2=\frac{b^2}{4a^2}-\frac{c}{a}=\frac{b^2}{4a^2}-\frac{4ac}{4a^2}=\frac{b^2-4ac}{4a^2}$$

$$\left(x+\frac{b}{2a}\right)^2=\frac{b^2-4ac}{4a^2}$$

$$x+\frac{b}{2a}=\sqrt{\frac{b^2-4ac}{4a^2}}=\frac{\sqrt{b^2-4ac}}{\sqrt{4a^2}}=\frac{\sqrt{b^2-4ac}}{\sqrt{4}\sqrt{a^2}}=\frac{\sqrt{b^2-4ac}}{2a}$$

$$x+\frac{b}{2a}=\frac{\sqrt{b^2-4ac}}{2a}$$

$$x = -\frac{b}{2a} + \frac{\sqrt{b^2 - 4ac}}{2a} = \frac{-b + \sqrt{b^2 - 4ac}}{2a}$$

$$x = \frac{-b + \sqrt{b^2 - 4ac}}{2a}$$

($\sqrt{x}$ means $\sqrt[2]{x}$ which as previously defined means, the number which, when raised to the power of 2, gives x.)

There is however a problem. With the current definition of the square root function $\sqrt{4}$, for example, means "the number which multiplies by itself to get 4. The problem is that there is more than one answer to this question, 2 *and* -2 which means that $2 = \sqrt{4}$ and $-2 = \sqrt{4}$ and using the axiom that if two things both equal the same thing then they both equal each other, $-2 = 2$. To avoid this contradiction, we must avoid one-to-many functions (functions which have multiple outputs for a single input). To avoid this, I will redefine the square root function, as well as any other even root functions, as they all have this problem. $\sqrt{x}$ is the positive number such that when squared, gives x. (squared means raised to the power of 2). The word "positive" means that the square root function now produces only one output for a given input. We must now revisit the step after this one:

$$\left(x + \frac{b}{2a}\right)^2 = \frac{b^2 - 4ac}{4a^2}$$

There are two different number which square to give $\left(x + \frac{b}{2a}\right)^2$. One is positive, $\sqrt{\frac{b^2-4ac}{4a^2}}$the other negative $-\sqrt{\frac{b^2-4ac}{4a^2}}$. To represent both of these solutions, I will use the $\pm$ notation.

$$x + \frac{b}{2a} = \pm\sqrt{\frac{b^2 - 4ac}{4a^2}}$$

From this point, the algebra can continue as before:

$$x + \frac{b}{2a} = \pm\sqrt{\frac{b^2 - 4ac}{4a^2}} = \pm\frac{\sqrt{b^2 - 4ac}}{\sqrt{4a^2}} = \pm\frac{\sqrt{b^2 - 4ac}}{\sqrt{4}\sqrt{a^2}} = \pm\frac{\sqrt{b^2 - 4ac}}{2a}$$

$$x = -\frac{b}{2a} \pm \frac{\sqrt{b^2 - 4ac}}{2a}$$

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$If\ ax^2 + bx + c = 0$$

$$Then\ x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Where the $\pm$ notation indicates that there are two possible answers.

It is important to note that there are not always 2 possible answers. The number of solutions can be found by looking at the bit inside the square root, $b^2 - 4ac$, which is called the discriminant.

If $b^2 - 4ac > 0$ so the discriminant is positive, then there will be a positive and a negative root, meaning that there are two possible solutions.

If $b^2 - 4ac = 0$ so the discriminant is zero, then there will be only one root because $\sqrt{0} = 0$ and positive and negative zero are both just zero.

If $b^2 - 4ac < 0$ so the discriminant is negative, then you are trying to square root a negative number. A negative number squared is positive, a positive number squared is positive and $0^2 = 0$ so there is no number x such that x^2 is negative. This means that when the discriminant is zero, there are no solutions.

Logarithms

In chapter 1, I promised that I would later get to the inverse of exponentiation, that is the logarithm.

To define the logarithm function:

$$if\ a^x = b$$

$$then\, x = \log_a b$$

Examples:

$$4^2 = 16$$

$$so \log_4(16) = 2$$

$$b^1 = b$$

$$so \log_b(b) = 1$$

$$\&\ b^0 = 1$$

$$so \log_{\mathrm{b}}(1) = 0$$

The rules for multiplying, dividing and exponentiating exponentials can be used to derive the laws of logarithms.

$$Let\ b = a^x$$

$$Let\ c = a^y$$

$$bc = a^x \times a^y = a^{x+y}$$

$$x + y = \log_a bc$$

$$x = \log_a b$$

$$y = \log_a c$$

$$x + y = \log_a b + \log_a c$$

$$\log_a bc = x + y = \log_a b + \log_a c$$

$$So \log_a bc = \log_a b + \log_a c$$

Similarly for division:

$$Let\ b = a^x$$

$$Let\ c = a^y$$

$$\frac{b}{c} = a^x \div a^y = a^{x-y}$$

$$x - y = \log_a\left(\frac{b}{c}\right)$$

$$x = \log_a b$$

$$y = \log_a c$$

$$x - y = \log_a b - \log_a c$$

$$\log_a\left(\frac{b}{c}\right) = x - y = \log_a b - \log_a c$$

$$So\ \log_a\left(\frac{b}{c}\right) = \log_a b - \log_a c$$

For exponentiation.

$$Let\ b = a^x$$

$$x = \log_a b$$

$$b^n = (a^x)^n = a^{xn}$$

$$\log_a(b^n) = xn = n\log_a b$$

$$So\ \log_a(b^n) = n\log_a b$$

There is one more:

$$Let\ x = \log_a b$$

$$a^x = b$$

$$\log_c(a^x) = \log_c(b)$$

$$x\log_c(a) = \log_c(b)$$

$$x = \frac{\log_c(b)}{\log_c(a)}$$

$$\log_a(b) = x = \frac{\log_c(b)}{\log_c(a)}$$

$$So \log_a(b) = \frac{\log_c(b)}{\log_c(a)}$$

These are known as the laws of logarithms.

Calculus Part 1: The Differential

If you managed to survive the heart attack which you received after reading the word "Calculus", welcome to calculus. This chapter will focus on limits, a key part of calculus. A limit, to put it simply is when a variable approaches some number. You can think of this variable as being a number very close to what it is approaching. Infinitely close in fact. To give an example, if $f(x)$ approaches 0 as x approaches infinity, this would mean that you could get $f(x)$ to be any number arbitrarily close to 0 if by letting x be big enough (close enough to infinity). In this case, no matter what value you chose for x, you will not get $f(x)$ to be 0 because x cannot actually be infinity. This limit would be written as $\lim_{x\to\infty} f(x) = 0$. If x was approaching some constant c and as x approaches c, $f(x)$ approaches b (written as $\lim_{x\to c} f(x) = b$) then $f(c) = b$, but only if two conditions are met. $f(c)$ exists, and $f(x)$ is continuous on the interval $\lim_{h\to 0}[c-h,\ c+h]$, in other words, the function is continuous as x gets arbitrarily close to c. The first of these conditions is obvious, if $f(c)$ does not exists ($f(x)$ is not defined when $x = c$) then the function at this point has no value. The second condition is also important because I could define a function $f(x)$ such that $f(x) = x^2$ except when $x = 2$ and $f(2) = 45$. This function would approach 4 as x approaches 2, but $f(2) \neq 4$.

To provide a more rigorous definition of the limit, limits are defined thusly:

$$\lim_{x\to c} f(x) = L\ means:$$

$$\forall\, \varepsilon > 0, \qquad \exists\, \delta > 0$$

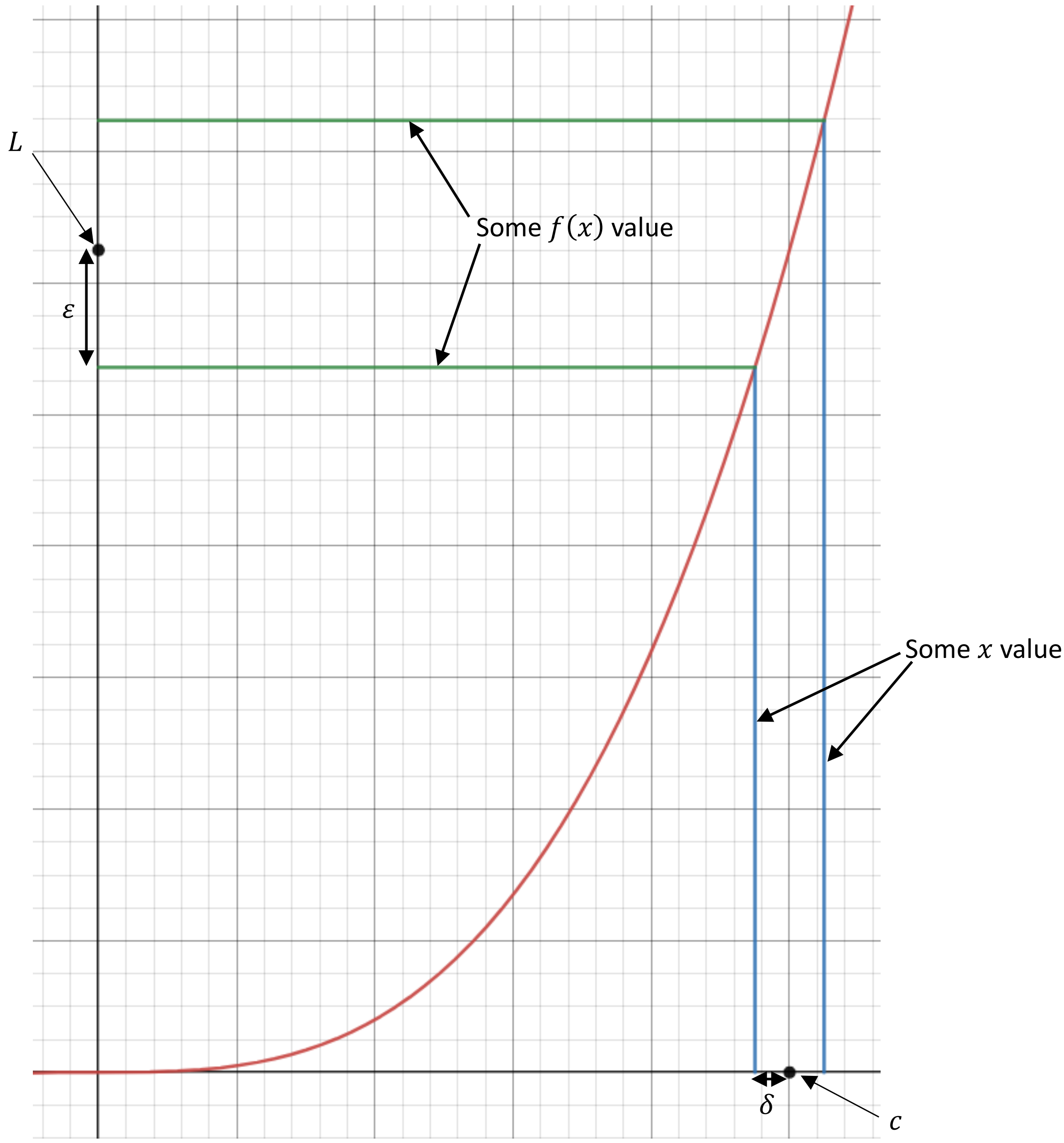
$$s.t.\ \ 0 < |x-c| < \delta \to |f(x) - L| < \varepsilon$$

This is called the epsilon-delta definition of the limit. To write it in English:

if $\lim_{x\to c} f(x) = L$ then: for all $\varepsilon > 0$, there exists some $\delta > 0$ such that the difference between x and c is positive and is less than δ and the difference between $f(x)$ and L is less than ε.

I think it is easiest to interpret this definition graphically.

Some arbitrary graph $y = f(x)$: [NOT TO SCALE!]



The definition says for all positive ε meaning I can choose any positive value of ε and this will hold true. To demonstrate this, I will choose an arbitrarily small value for ε. $Let\ \varepsilon = 0.0001$. The above states that the difference between $f(x)$ and L is less than ε and so is less than 0.0001 meaning they are very close together. In the above diagram, this means that the value of $f(x)$ lies between the two horizontal green lines. For these $f(x)$ values, there is a set of x values which are those which fall between the two blue lines on the diagram. The distance between the blue lines and c is labelled δ because the difference between the c and x is less than δ.

The true purpose of showing the diagram is to help you to visualise the following. If the two blue lines were moved arbitrarily close to c, (the value of δ becomes arbitrarily small) then imagine what would happen to the green lines. They would close in on a single value for $f(x)$. This value is the one labelled L. It is worth noting that in the case of the graph above, $f(c) = L$. This is because the graph is continuous. It would be entirely possible to define a function for example as $f(x) = x^2$ except when $x = 2$ and $f(2) = 12$. This seems arbitrary but it is just a simple example. In this case,

$\lim_{x \to 2} f(x) = 4$, but $f(2) \neq 4$. This is because such a function would not be continuous. A continuous function, to put it simply, is one with a value for every x value (meaning there are no values of x for which $f(x)$ has no value), there is no point at which the $f(x)$ “jumps” from one value to another. Essentially, this means that you could draw the curve $y = f(x)$ without needing to take your pencil off the page at any point. Even having a single point missing (though not visible on a graph unless explicitly labelled) means that the function is not continuous at that point (or to be more precise, on any interval which contains that point).

Before finishing this chapter, I would like to address its title: the differential. What is a differential? A differential is an infinitesimal (infinitely small/close to zero) change in some variable. dx means infinitesimal change in x, dy means infinitesimal change in y, $d\theta$ means infinitesimal change in θ and so on. It is important to note that, whilst all differentials approach zero, that does not mean that they are all equal to each other. Differentials can be thought of as just being numbers, very small numbers still, but numbers nonetheless. This means that all the regular rules apply to them (so long as we define them in this way which I will be doing.)

If you are a bit shook by the epsilon-delta definition of the limit, then don’t be too concerned because that is the by far the most difficult that calculus will get for a while.

Sequences & Series

A sequence is a set of numbers in a specific order, for example: $1, 2, 3, 4$ and $4, 23, 16, -23, 56, 192$. A series is a set of numbers in a specific order being added together, for example: $1 + 2 + 3 + 4$ and $4 + 23 + 16 + (-23) + 56 + 192$. One type of sequence is called an arithmetic sequence where each term has a common difference between itself and the one before it. For example, $10, 15, 20, 25$ is an arithmetic sequence with common difference 5 and first term 10. A general arithmetic sequence, first term a and common difference d with n terms, looks like this:

$$a, a + d, a + 2d, a + 3d, a + 4d, a + 5d, \dots a + (n - 1)d$$

The first term is a, the second term is $a + d$, the third is $a + 2d$, the hundredth is $a + 99d$. In general, the n^{th} term is $a + (n - 1)d$.

An arithmetic series is the sum of an arithmetic sequence. A general arithmetic sequence, with first term a and common difference d with n terms, looks like this:

$$(a) + (a + d) + (a + 2d) + (a + 3d) \dots + (a + (n - 3)d) + (a + (n - 2)d) + (a + (n - 1)d)$$

To find a way to calculate this sum more easily, I will start with a simpler example: the sum of the first 100 natural numbers (the numbers from 1 to 100).

$$1 + 2 + 3 + 4 + \cdots + 97 + 98 + 99 + 100$$

To make this an easier sum to work out, we can change the order in which the terms are added together.

$$1 + 100 + 2 + 99 + 3 + 98 + 4 + 97 + \cdots + 49 + 52 + 50 + 51$$

Now, each pair of numbers adds to 101 and there are 50 pairs of numbers, so the total is $101 \times 50 = 5050$. This only works if the number of terms being added is even. If it was odd, 101 for example, then all terms would not be put into pairs as one would be left over. To fix this, simply pretend that the last term isn't there and add it on at the end.

I will now attempt to apply this trick to the general series:

If n is even:

$$(a) + (a + d) + (a + 2d) + (a + 3d) \dots + (a + (n - 3)d) + (a + (n - 2)d) + (a + (n - 1)d)$$

$$= (a + a + (n - 1)d) + (a + d + a + (n - 2)d) + (a + 2d + a + (n - 3)d) + \cdots$$

$$= (2a + (n - 1)d) + (2a + (n - 1)d) + (2a + (n - 1)d) + \cdots$$

$$= \frac{n}{2}(2a + (n - 1)d)$$

If n is odd:

Ignore the last term,

$$(a) + (a + d) + (a + 2d) + (a + 3d) \dots + (a + (n - 3)d) + (a + (n - 2)d)$$

This series has $n - 1$ terms.

$$= (a + a + (n - 2)d) + (a + d + a + (n - 3)d) + (a + 2d + (n - 4)d) + \cdots$$

$$= (2a + (n-2)d) + (2a + (n-2)d) + (2a + (n-2)d) + \cdots$$

$$= \frac{n-1}{2}(2a + (n-2)d)$$

Add the last term back on:

$$\frac{n-1}{2}(2a + (n-2)d) + (a + (n-1)d)$$

$$\frac{1}{2}(n-1)(2a + (n-2)d) + (a + (n-1)d)$$

$$= \frac{1}{2}(n-1)(2a + (n-2)d) + \frac{1}{2} \times 2(a + (n-1)d)$$

$$= \frac{1}{2}[(n-1)(2a + (n-2)d) + 2(a + (n-1)d)]$$

$$= \frac{1}{2}[n(2a + (n-2)d) - (2a + (n-2)d) + 2(a + (n-1)d)]$$

$$= \frac{1}{2}[2an + n(n-2)d - 2a - (n-2)d + 2a + 2(n-1)d]$$

$$= \frac{1}{2}[2an + n^2d - 2nd - 2a - nd + 2d + 2a + 2nd - 2d]$$

$$= \frac{1}{2}[2an + n^2d - nd]$$

$$= \frac{n}{2}(2a + nd - d)$$

$$\frac{n}{2}(2a + (n-1)d)$$

Which is the same as when n is even.

So, the sum of an arithmetic series, first term a and common difference d is:

$$\frac{n}{2}(2a + (n-1)d)$$

The next type of sequence I will discuss is geometric sequences.

A geometric sequence is one where each term has a common ration between itself and the previous. In other words, each term is the previous term multiplied by some common ratio r. A geometric sequence, first term a and common ratio r with n terms, looks like this:

$$a, ar, ar^2, ar^3, \ldots ar^{n-1}$$

An arithmetic series is the sum of a geometric sequence. A geometric series, first term a, common ratio r and first term a with n terms is:

$$a + ar + ar^2 + ar^3 + \cdots + ar^{n-1}$$

To find the sum of this series, I will call it S.

$$Let\ S = a + ar + ar^2 + ar^3 + \cdots + ar^{n-1}$$

Notice that multiplying each term by r, gives the same series except the first and last terms are changed.

$$Sr = ar + ar^2 + ar^3 + ar^4 + \cdots + ar^{n-1} + ar^n$$

This series is the same as the original, except there is no a term and there is an extra term at the end ar^n.

So $Sr = S - a + ar^n$. This could also have been found by subtracting one series from the other and then rearranging.

$$Sr - S = ar^n - a$$

$$S(r-1) = a(r^n - 1)$$

$$S = \frac{a(r^n - 1)}{r - 1}$$

So, the sum of a geometric series, first term a and common ratio r with n terms is given by:

$$\frac{a(r^n - 1)}{r - 1}$$

The next thing I will investigate is infinite series, where an infinite number of terms are added. It may at first seem that adding an infinite number of terms together would always give an infinite result. Even if this was not the case, how would you even calculate such a thing, you can't add an infinite number of terms together as that would take in infinite amount of time.

We can solve this problem by using limits, instead of saying n is infinity, we can ask what happens to the sum as n approaches infinity.

For the sum of an infinite arithmetic series:

$$S = \lim_{n \to \infty} \frac{n}{2}(2a + (n-1)d)$$

As n approaches ∞, $\frac{n}{2}$ also approaches ∞, as does $n - 1$.

$$\infty(2a + \infty d)$$

If d is positive, then ∞d approaches ∞, If d is negative, then ∞d approaches $-\infty$, and if d is 0, then ∞d approaches 0. This is why it is important that we use limits instead of simply "letting $n = \infty$", because the value of $0 \times \infty$ depends on whether this means $\lim_{h \to 0} \infty h = \infty$ or $\lim_{n \to \infty} 0n = 0$. In this case, if d was positive, then $2a + \infty d$ will approach ∞. If d was negative, then $2a + \infty d$ will approach $-\infty$. When either of these is multiplied by ∞, they remain as they were. If d was 0, the $2a + \infty d$ approaches $2a$, which when multiplied by ∞, will approach ∞ if a is positive, $-\infty$ if a is negative and 0 if $a = 0$.

To summarise, an infinite arithmetic series will approach ∞ if the common difference is positive and will approach $-\infty$ if the common difference is negative. If the common difference is 0, then the series will approach ∞ if the first term was positive and ∞ if the first term was negative. If the common difference and first term were both 0, the series approaches 0, which makes sense because $0 + 0 + 0 + 0 + 0 + \cdots = 0$.

Infinite geometric series, however, are more interesting.

$$S = \lim_{n\to\infty} \frac{a(r^n - 1)}{r - 1}$$

Multiplying by a number which is greater than 1 increases its size, multiplying by a number which is less than 1 decreases its size (assuming r is positive) and multiplying a number by 1 does nothing, so r^{10000} will be a very big number if $r > 1$ and will be a very small number if $r < 1$, it will equal 1 if $r = 1$. This means that $\lim_{n\to\infty} r^n = \infty$ when $r > 1$, $\lim_{n\to\infty} r^n = 0$ when $0 \leq r < 1$ and $\lim_{n\to\infty} r^n = 1$ when $r = 1$. As for negative values of r, $\lim_{n\to\infty} r^n$ does not exist when $r < -1$ because as n increases the value of the limit changes between negative and positive, the limit approaches both $-\infty$ and ∞ so we say that the limit diverges does not exist (it does not converge to a single finite value and also does not approach ∞ or $-\infty$). When $r = -1$, the limit switches between -1 and 1 so again the limit does not exist. When $-1 < r \leq 0$, the limit approaches 0, because whilst it does still switch between positive and negative, because $-0 = +0$, this does not really matter.

So, when $r = 1$:

$$S = \lim_{n\to\infty} \frac{a(1^n - 1)}{1 - 1} = \frac{a(1 - 1)}{1 - 1} = \frac{a0}{0} = \frac{0}{0}$$

Which is indeterminant (meaning the limit may exist but we would need to do more work/have more information to work it out).

When $r > 1$:

$$S = \frac{a(\infty - 1)}{r - 1} = \frac{\infty a}{r - 1}$$

$r > 1$ so $r - 1 > 0$ so whether this is positive or negative infinity depends on the value of a.

When $r \leq -1$: the limit does not exist.

When $-1 < r < 1$, or $|r| < 1$ (which means that if r is made positive, it is less than 1).

$$S = \frac{a(0 - 1)}{r - 1} = \frac{-a}{r - 1} = \frac{a}{1 - r}$$

So, when $|r| < 1$, the sum of an infinite geometric series, first term a, common ratio r is given by:

$$\frac{a}{1 - r}$$

There is a more compact and objective way of writing series called Sigma notation.

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

An arithmetic series with a total of m terms and an n^{th} term of $a + (n-1)d$ could be written in sigma notation as:

$$\sum_{n=1}^{m} (a + (n-1)d) = \frac{m}{2}(2a + (m-1)d)$$

Or by considering the first term as being when $n = 0$:

$$\sum_{n=0}^{m-1}(a+nd)=\frac{m}{2}(2a+(m-1)d)$$

For a geometric series with a total of m terms and an n^{th} term of ar^{n-1}:

$$\sum_{n=1}^{m} ar^{n-1}=\frac{a(r^m-1)}{r-1}$$

Or, again by considering the first term as being when $n=0$:

$$\sum_{n=0}^{m-1} ar^{n}=\frac{a(r^m-1)}{r-1}$$

Finally, for an infinite geometric series:

$$\lim_{m\to\infty}\left(\sum_{n=0}^{m-1}(ar^n)\right)=\frac{a}{1-r}$$

As $m\to\infty$, $(m-1)\to\infty$, so they above could also be written as:

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

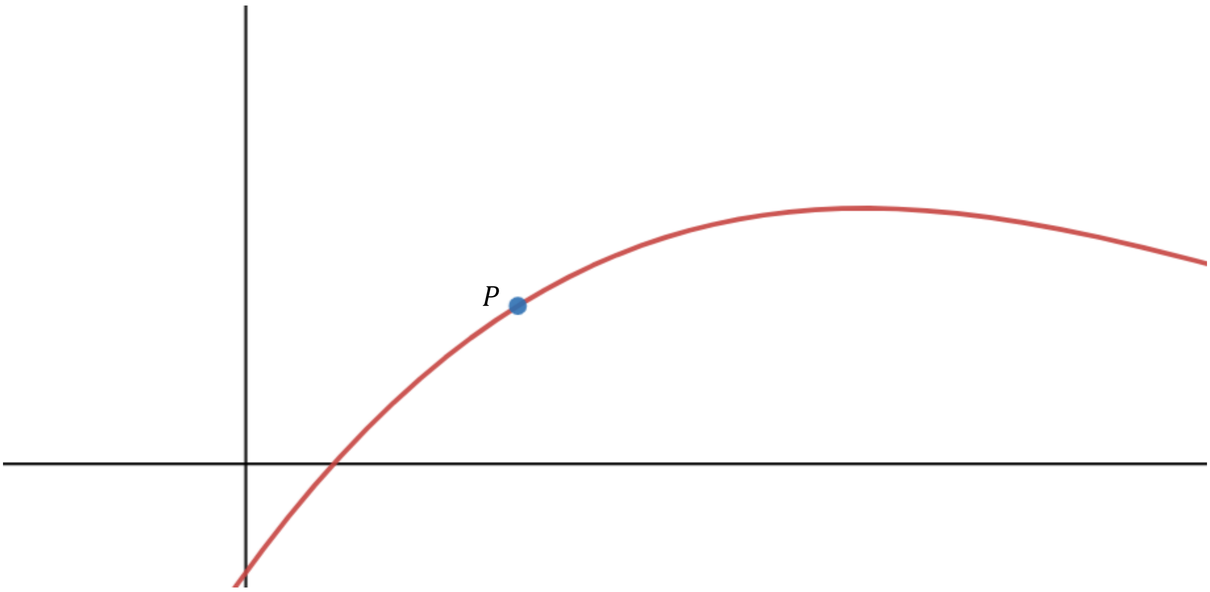
which is how it is often written.

In case you were wondering, the reason why the capital Greek letter sigma (Σ) is used is because it is the Greek version of the letter s for “sum”. π is the Greek version of the letter p, and so capital Pi (Π) notation is used for products.

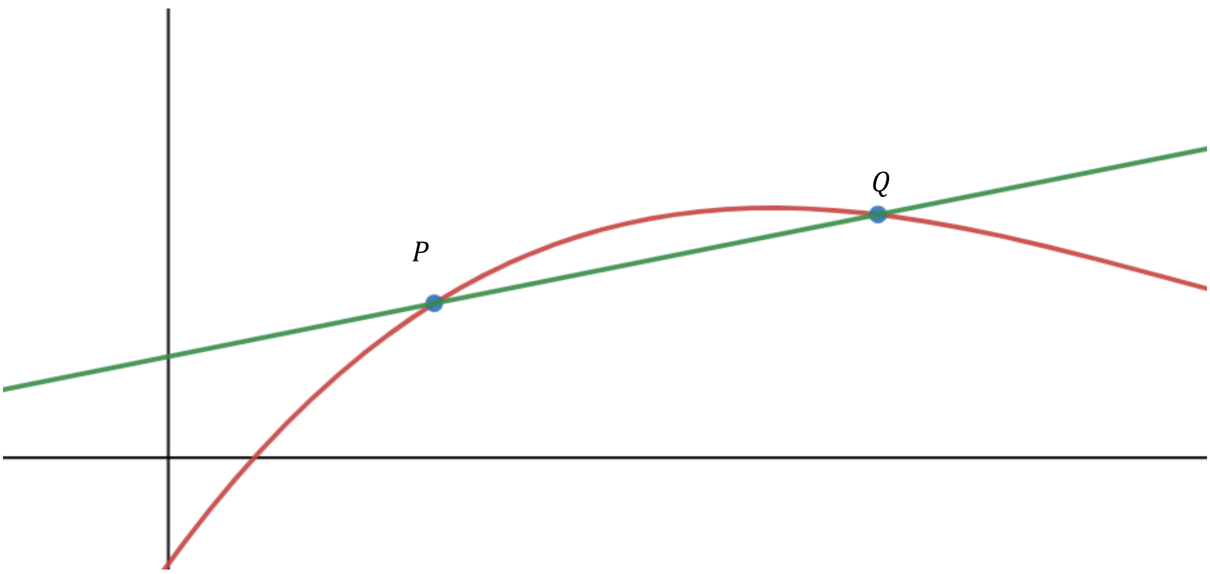
$$\prod_{n=a}^{b} f(n)=f(a)\times f(a+1)\times f(a+2)\times\cdots\times f(b-2)\times f(b-1)\times f(b)$$

Calculus Part 2: The Derivative

This chapter is about derivatives, the way in which the gradient of a curved graph can be calculated. First though, the gradient of a straight line is simply defined as its $\frac{change\ in\ y}{change\ in\ x}$. It is the rate at which x increases with respect to y. Finding the gradient of a curve is less simple. Here is some arbitrary continuous curve with the function $y = f(x)$:

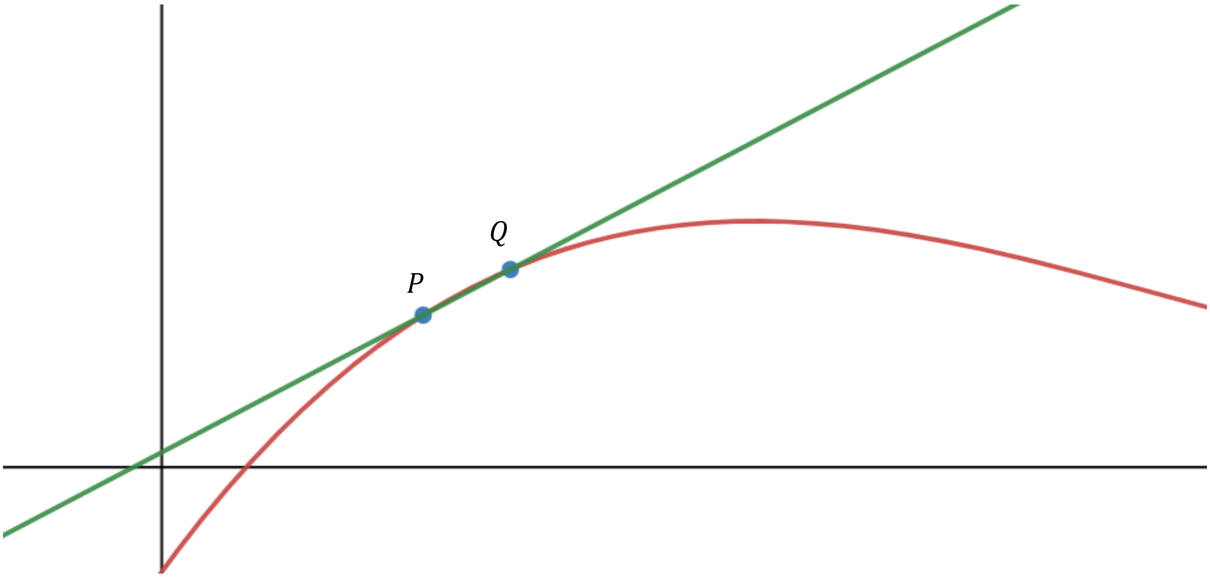


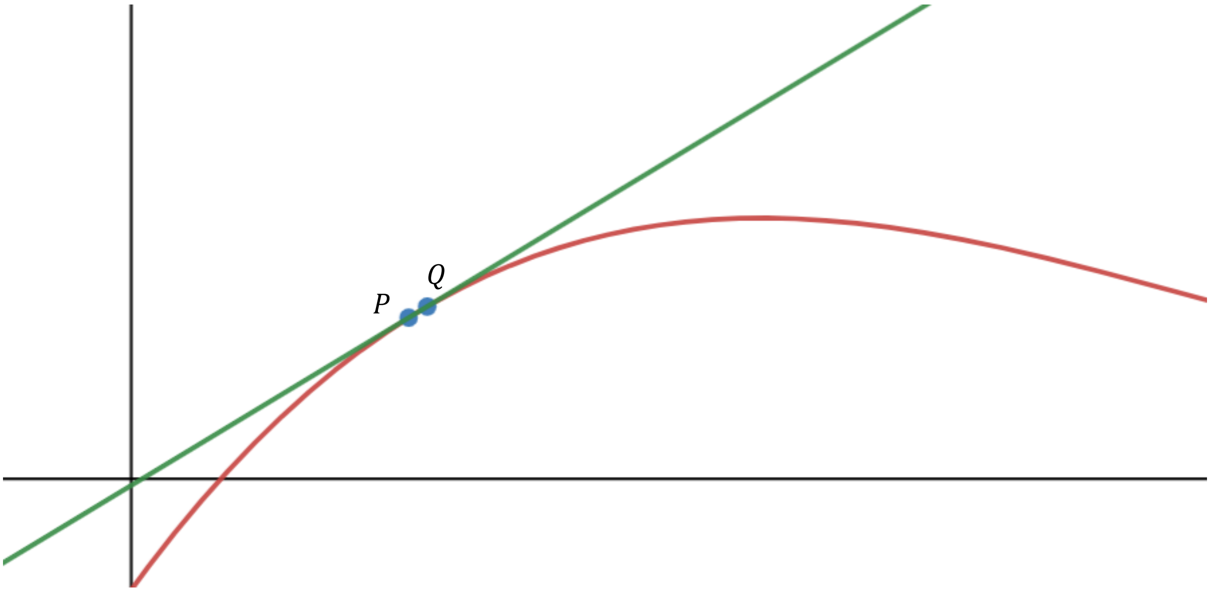
How could I find the gradient of the curve at the labelled point P? It depends how we define the "gradient of a curve at a point". By drawing a few points on the curve and joining them together with straight lines, you could create an approximation of this graph using straight lines and would therefore be able to find the gradient at a certain point. By adding more points and splitting drawing more straight lines, the curve approximation gets closer and closer to the actual curve. The gradient of a point on the curve should therefore be defined as the gradient of the infinitely small line segment which makes it. That is to say, the gradient of the tangent to the curve at that point, (the tangent at a point is the straight line which touches the curve but does not "pierce" it). To estimate this tangent, we can draw a line between our point and some other point on the curve which I will call Q. I will call the x-coordinate of Q c and I will call the x-coordinate of P x.



This means that point P has coordinates $\left(x, f(x)\right)$ and Q has coordinates $\left(c, f(x)\right)$. The gradient of this line is defined as the $\frac{change\ in\ y}{change\ in\ x} = \frac{f(x)-f(c)}{x-c}$.

As point Q moves close to point P the line connecting them gets closer to that of the actual tangent line at P.





As point Q gets closer to point P, or as c gets closer to x the gradient approximation gets better and better. We can express this idea using limits.

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

This is known as the difference quotient.

This limit equals the gradient of the curve $y = f(x)$ at some coordinate x. Because the limit is in terms of x it is also a function of x. I will call this function $f'(x)$. x and c are two arbitrarily chosen letters, meaning they can be switched. To find the gradient at some point $\left(c, f(c)\right)$ use the limit:

$$f'(c) = \lim_{x \to c} \frac{f(c) - f(x)}{c - x}$$

$$f'(c) = \lim_{x \to c} \frac{f(x) - f(c)}{x - c}$$

$f(x)$ is continuous, meaning that the gradient of $f(x)$ also changes continuously which means that $f'(x)$ is also continuous which means that $f'(c) = \lim_{x \to c} f'(x)$ so

$$\lim_{x \to c} f'(x) = \lim_{x \to c} \frac{f(x) - f(c)}{x - c}$$

This difference quotient could be used to find gradients of curves but there is an easier way to write it.

Instead of saying the x-coordinate of P is x and the x-coordinate of Q is c and that $c \to x$, we could have just as easily have said that the x-coordinate of P is x and the difference between the x-coordinates of P and Q is h and $h \to 0$.

The coordinates of P would still be $(x, f(x))$, and the coordinates of Q would be $(x + h, f(x + h))$. The gradient of the line joining these point $= \frac{f(x+h)-f(x)}{h}$. Remember that $h \to 0$ meaning the gradient of the curve $y = f(x)$ at some point $(x, f(x))$ is given by:

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

This expression can be written using differentials. $\lim_{h \to 0} f(x + h) - f(x) = 0$. This limit approaches zero and represents a change in y. It is therefore an infinitesimal change in y and can therefore be written as dy. $\lim_{h \to 0} h = 0$. This limit also approaches zero, though not necessarily at the same rate, it represents an infinitesimal change in x and can therefore be written as dx. The gradient of some curve at some point can written then as $\frac{dy}{dx}$. The value of $\frac{dy}{dx}$ will be different at different points on the curve because it is key to remember that $\frac{dy}{dx}$ is not about the values of dy and dx themselves, but rather the relationship between the two, how one changes with respect to the other. $\frac{dy}{dx}$ should therefore be a function of x which shows how the relationship between the two differentials changes for different x values. This function is known as the "gradient function" of $f(x)$ or as its "derivative". For a given x value though, $\frac{dy}{dx}$ is just some number, that being the value gradient at that point.

The derivative of some function multiplied by a constant, $\frac{d}{dx}(k \times f(x))$ can be found by using the difference quotient. ($\frac{d}{dx}$ is just a notation which means "the derivative of … with respect to x".)

$$\frac{d}{dx}(k \times f(x)) = \lim_{h \to 0} \frac{k \times f(x + h) - k \times f(x)}{h}$$

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

$$k \times \lim_{h \to 0} \frac{f(x + h) - f(x)}{h}$$

$$= k \times f'(x)$$

$$So\frac{d}{dx}\left(k \times f(x)\right) = k \times \frac{d}{dx}f(x)$$

I will now show how we can use the difference quotient to find the derivatives of certain functions, starting with $\sin(x)$.

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

Using the addition formulae:

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

By factoring out the $\sin(x)$:

$$sin'(x) = \lim_{h\to 0}\frac{\sin(x)\,(\cos(h) - 1) + \sin(h)\cos(x)}{h}$$

Splitting this into two fractions:

$$sin'(x) = \lim_{h\to 0}\left(\frac{\sin(x)\,(\cos(h) - 1)}{h} + \frac{\sin(h)\cos(x)}{h}\right)$$

Instead of writing $\lim_{h\to 0}$ just at the beginning, I could write it separately at all of the places where there is an expression with h in it. For example:

$$sin'(x) = \frac{\sin(x)\left(\cos\left(\lim_{h\to 0}(h)\right) - 1\right)}{\lim_{h\to 0}(h)} + \frac{\sin\left(\lim_{h\to 0}(h)\right)\cos(x)}{\lim_{h\to 0}(h)} = \frac{\sin(x)\,(1-1)}{0} + \frac{0\cos(x)}{0} = \frac{0}{0} + \frac{0}{0}$$

which is not very helpful.

This $\frac{0}{0}$ appears because I put the limit separately on the top and on the bottom of the fraction. It would therefore be smarter not to do this.

$$\sin'(x) = \lim_{h\to 0}\frac{\sin(x)(\cos(h) - 1)}{h} + \lim_{h\to 0}\frac{\sin(h)\cos(x)}{h}$$

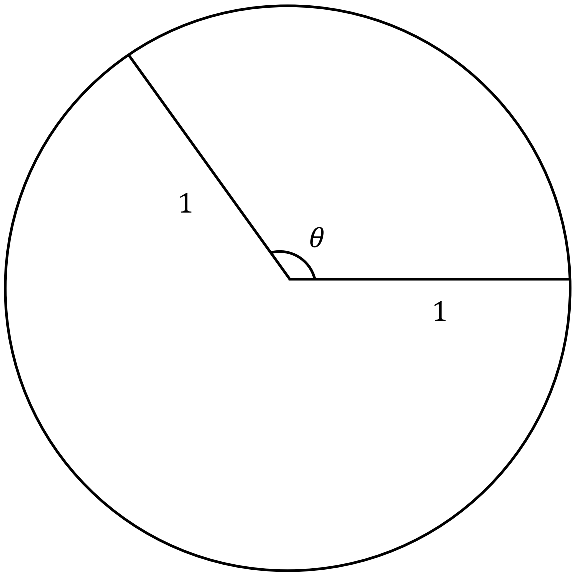
I can now bring the $\sin(x)$ and $\cos(x)$ outside of the limits, as they do not have hs in them and so do not need to be inside of the limit.

$$\sin'(x) = \sin(x)\lim_{h\to 0}\frac{\cos(h) - 1}{h} + \cos(x)\lim_{h\to 0}\frac{\sin(h)}{h}$$

Now we seem to have hit a bit of a roadblock. How do I work out the values of these limits? This can be done geometrically. It would, however, be much easier if we change the units which we currently use for angles because degrees were chosen more or less arbitrarily. I will define a new measurement of angle.

Looking at a unit circle (circle with a radius of 1), its $circumference = 2\pi r = 2\pi$.

Some sector of the unit circle, angle θ



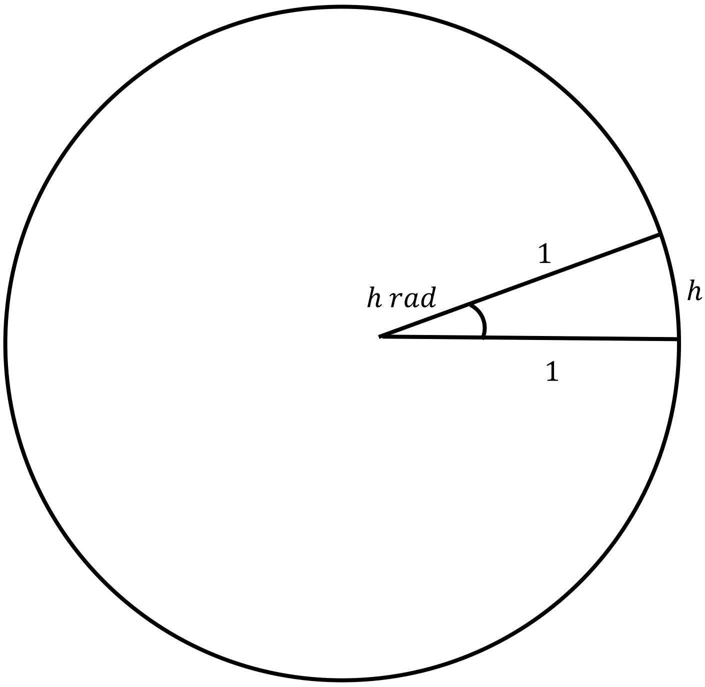
To find the arc length (The curved length of the sector), you know that the ratio between this length and the total circumference is the same as the ratio between the θ and the angle of a full rotation.

So $\frac{Arc\ Length}{Circumference} = \frac{\theta}{angle\ of\ full\ rotation} \rightarrow \frac{Arc\ Length}{2\pi} = \frac{\theta}{angle\ of\ full\ rotation}$

Note that, because I am defining a new unit for measuring angles, I get to decide what the value of an angle of a full rotation is. If I let this value equal 2π, then:

$$\frac{Arc\ Length}{2\pi} = \frac{\theta}{2\pi} \rightarrow Arc\ Length = \theta$$

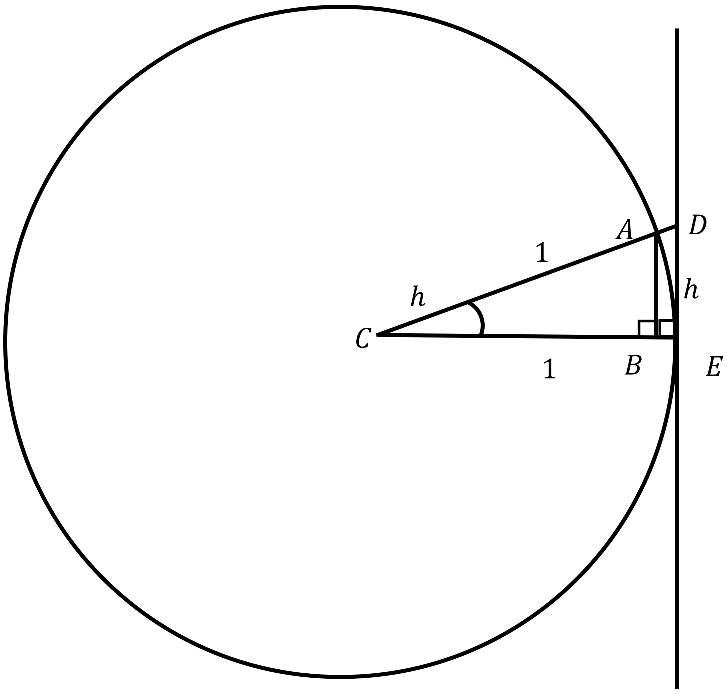
This could prove to be a very convenient property so I will define my new angle of measurement as being such that the angle of full rotation $= 2\pi\ units$, meaning that a right-angle $= \frac{\pi}{2}$ units. I will call this unit, the radian. I will talk more about radians later, but for now they can be used to solve our limits problem.



In this diagram, the angle is labelled as $h\ radians$ and since (when using radians) $Arc\ Length = \theta$, the arc length is also labelled as h. The radii of the circle are labelled as 1 because this is a unit circle.

I will do some general geometry with this diagram, and will then see what happens as $h \to 0$.

Constructing the following right triangles and labelling points:



From triangle ABC: $\sin(h) = \frac{AB}{AC} = \frac{AB}{1} = AB$ so $AB = \sin(h)$

From triangle CDE: $\tan(h) = \frac{DE}{CE} = \frac{DE}{1} = DE$ so $DE = \tan(h)$

AE is the arc length and so $AE = h$.

It is clear that $AB < AE$ because the distance between A and E is greater than the distance between A and B, also AB is a straight line and AE is curved. This means that $\sin(h) < h$. Remember we are interested in $\frac{\sin(h)}{h}$.

$$\sin(h) < h$$

$$\frac{\sin(h)}{h} < 1$$

As for comparing AB and DE, it is unclear which is larger. To resolve this, I can compare the areas of triangle CDE and arc ACE.

$$Area\ of\ CDE = \frac{1}{2} \times CE \times DE = \frac{1}{2} \times 1 \times \tan(h) = \frac{\tan(h)}{2}$$

For the area of the sector: $\frac{h}{2\pi}$ is the proportion of the whole circle taken up by the sector. This means that $\frac{h}{2\pi} \times circle\ area = sector\ area$

$$Area\ of\ ACE = \frac{h}{2\pi} \times \pi r^2 = \frac{h\pi r^2}{2\pi} = \frac{h}{2}$$

Clearly $Area\ of\ ACE < Area\ of\ CDE$ so:

$$\frac{h}{2} < \frac{\tan(h)}{2}$$

$$h < \tan(h)$$

Again, remember we are interested in $\frac{\sin(h)}{h}$.

$$h < \tan(h)$$

$$h < \frac{\sin(h)}{\cos(h)}$$

$$h\cos(h) < \sin(h)$$

I can multiply both sides by $\cos(h)$ without changing the sign because h is close to 0, meaning $\cos(h)$ is close to 1 which is positive.

$$\cos(h) < \frac{\sin(h)}{h}$$

I now have two inequalities: $\frac{\sin(h)}{h} < 1$, $\cos(h) < \frac{\sin(h)}{h}$

Combining these inequalities together:

$$\cos(h) < \frac{\sin(h)}{h} < 1$$

As I let $h \to 0$ now:

$$\cos(0) < \lim_{h\to 0}\frac{\sin(h)}{h} < 1$$

$$1 < \lim_{h\to 0}\frac{\sin(h)}{h} < 1$$

Whilst it may seem obsurd that 1 can be less than this limit and 1 can be greater than it, that is just because we are taking a limit. If h was some very small number close to 0, we may have gotten, for example: $0.99999 < \lim_{h\to 0}\frac{\sin(h)}{h} < 1$. The 0.99999 could have had an arbitrary number of nines. All of this means that it only makes sense to say that $\lim_{h\to 0}\frac{\sin(h)}{h} = 1$.

I still need to deal with the other limit: $\lim_{h\to 0}\frac{\cos(h)-1}{h}$.

$$\lim_{h\to 0}\frac{\cos(h)-1}{h}$$

$$= \lim_{h\to 0}\frac{(\cos(h)-1)(\cos(h)+1)}{h(\cos(h)+1)}$$

$$= \lim_{h \to 0} \frac{\cos^2(h) - \cos(h) + \cos(h) - 1}{h(\cos(h) + 1)}$$

$$= \lim_{h \to 0} \frac{\cos^2(h) - 1}{h(\cos(h) + 1)}$$

Recall that $\sin^2(\theta) + \cos^2(\theta) = 1$ so $\cos^2(\theta) = 1 - \sin^2(\theta)$

$$\lim_{h \to 0} \frac{\cos^2(h) - 1}{h(\cos(h) + 1)}$$

$$= \lim_{h \to 0} \frac{1 - \sin^2(h) - 1}{h(\cos(h) + 1)}$$

$$= \lim_{h \to 0} \frac{-\sin^2(h)}{h(\cos(h) + 1)}$$

This can be split into two fractions one of which is the same as the limit we just solved.

$$= \lim_{h \to 0} \left(\frac{\sin(h)}{h} \times \frac{-\sin(h)}{\cos(h) + 1} \right)$$

$$= \lim_{h \to 0} \frac{\sin(h)}{h} \times \lim_{h \to 0} \frac{-\sin(h)}{\cos(h) + 1}$$

$$= 1 \times \lim_{h \to 0} \frac{-\sin(h)}{\cos(h) + 1}$$

$$= \lim_{h \to 0} \frac{-\sin(h)}{\cos(h) + 1}$$

I can now let $h \to 0$:

$$\lim_{h \to 0} \frac{-\sin(h)}{\cos(h) + 1}$$

$$= \frac{-\sin(0)}{\cos(0) + 1}$$

$$= \frac{-0}{1 + 1} = \frac{0}{2} = 0$$

$$So \lim_{h \to 0} \frac{\cos(h) - 1}{h} = 0$$

We now know the values of both of these limits, which are known as the fundamental trigonometric limits.

I can now differentiate $\sin(x)$

$$\sin'(x) = \sin(x) \lim_{h \to 0} \frac{\cos(h) - 1}{h} + \cos(x) \lim_{h \to 0} \frac{\sin(h)}{h}$$

$$sin'(x) = \sin(x) \times 0 + \cos(x) \times 1$$

$$\sin'(x) = \cos(x)$$

Differentiating $cosine$ is a similar story:

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

$$\cos'(x) = \lim_{h\to 0} \frac{\cos(x)\cos(h) - \sin(x)\sin(h) - \cos(x)}{h}$$

$$\cos'(x) = \lim_{h\to 0} \frac{\cos(x)\,(\cos(h) - 1) - \sin(x)\sin(h)}{h}$$

$$\cos'(x) = \lim_{h\to 0} \left(\frac{\cos(x)\,(\cos(h) - 1)}{h} - \frac{\sin(x)\sin(h)}{h}\right)$$

$$\cos'(x) = \cos(x) \lim_{h\to 0} \frac{\cos(h) - 1}{h} - \sin(x) \lim_{h\to 0} \frac{\sin(h)}{h}$$

$$\cos'(x) = \cos(x) \times 0 - \sin(x) \times 1$$

$$\cos'(x) = -\sin(x)$$

Next I wish to find a function wich differentiates to itself. This means that

$$f(x) = f'(x)\ for\ all\ x$$

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

I will rearrange for $f(x)$ and will take the limit at the end.

$$f(x) = \frac{f(x+h) - f(x)}{h}$$

$$h\,f(x) = f(x+h) - f(x)$$

$$h\,f(x) + f(x) = f(x+h)$$

$$f(x)(h+1) = f(x+h)$$

It seems that here, I have hit a roadblock. This is because there is actually more than one solution. To find one of them I will $Let\ f(0) = 0$. Because the above is true for all x, I can let $x = 0$ and the resulting equation should hold true.

$$f(0)(h+1) = f(0+h)$$

$$0 = f(h)$$

I will now see what happens if I let $x = h$.

$$f(h)(h+1) = f(h+h)$$

We know that $f(h) = 0$ so:

$$0(h+1) = f(2h)$$

$$0 = f(2h)$$

Let $x = 2h$:

$$f(2h)(h+1) = f(2h+h)$$

$$0(h+1) = f(3h)$$

$$0 = f(3h)$$

In general, $f(nh) = 0$ for all positive integers n. Since h is infinitsesimally small (when the limit is applied) this covers all positive numbers. Also h could be approaching 0 from the negatives, meaning that the above also applies for negative inputs to f. This all means that $f(x) = 0$ for all x.

$$If\ y = 0$$

$$Then \frac{dy}{dx} = 0$$

There may however be more interesting solutions. To find them all, instead of letting $f(0) =$ some specific number I will $let\ f(0) = m$ where m is just some number meaning we can get our answer in terms of m.

$$f(x+h) = f(x)(h+1)$$

$$Let\ x = 0$$

$$f(0+h) = f(0)(h+1)$$

$$f(h) = m(h+1)$$

$$Let\ x = h$$

$$f(h+h) = f(h)(h+1)$$

$$f(2h) = m(h+1)(h+1)$$

$$f(2h) = m(h+1)^2$$

$$Let\ x = 2h$$

$$f(2h+h) = f(2h)(h+1)$$

$$f(3h) = m(h+1)^2(h+1)$$

$$f(3h) = m(h+1)^3$$

$$Let\ x = 3h$$

$$f(3h+h) = f(3h)(h+1)$$

$$f(4h) = m(h+1)^3(h+1)$$

$$f(4h) = m(h+1)^4$$

In general:

$$f(kh) = m(h+1)^k$$

Where k is a positive integer.

$$Let\ x = kh$$

$$k = \frac{x}{h}$$

$$f(x) = m(h+1)^{\frac{x}{h}}$$

Now take the limit $h \to 0$

$$f(x) = m \lim_{h \to 0} (h+1)^{\frac{x}{h}}$$

This is already the answer, but I will change one thing. I could replace all instances of h with $\frac{1}{m}$ and change the limit from $h \to 0$ to $n \to \infty$ because (if we assume that $h > 0$) as n gets bigger (closer to ∞, $\frac{1}{n}$ gets closer to 0. $\lim_{h \to 0} f(h)$ could be replaced with $\lim_{n \to \infty} f\left(\frac{1}{n}\right)$ like this:

$$f(x) = m \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^{nx}$$

$$f(x) = m \lim_{n \to \infty} \left(\left(1 + \frac{1}{n}\right)^{n}\right)^{x}$$

Where m is any constant. I will talk more about this limit later.

Next, I will differentiate another function which we have met, the logarithm. I will not specify the base, which means that what I am doing applies to any arbitrary base.

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

Which of our laws of logarithms might we be able to apply here? I have one logarithm being subtracted from another so I could use that one.

$$\log'(x) = \lim_{h \to 0} \frac{\log\left(\frac{x+h}{x}\right)}{h}$$

$$\log'(x) = \lim_{h \to 0} \frac{\log\left(\frac{x}{x} + \frac{h}{x}\right)}{h}$$

$$\log'(x) = \lim_{h \to 0} \frac{\log\left(1 + \frac{h}{x}\right)}{h}$$

$$\log'(x) = \lim_{h \to 0} \frac{1}{h} \log\left(1 + \frac{h}{x}\right)$$

Now I have a logarithm being multiplied by something.

$$\log'(x) = \lim_{h\to 0} \log\left(\left(1+\frac{h}{x}\right)^{\frac{1}{h}}\right)$$

This limit looks very similar to one we have seen before. To make this limit more similar to that one I will $let \frac{1}{n} = \frac{h}{x}$ which means that $n = \frac{x}{h}$ and that $\frac{1}{h} = \frac{n}{x}$. This means that, as long as $x > 0$ (which it always is because the logarithm is only defined for positive values of x), as $h \to 0$, $n \to \infty$.

$$\log'(x) = \lim_{n\to\infty} \log\left(\left(1+\frac{1}{n}\right)^{\frac{n}{x}}\right)$$

$$\log'(x) = \lim_{n\to\infty} \log\left(\left(1+\frac{1}{n}\right)^{n\times\frac{1}{x}}\right)$$

$$\log'(x) = \lim_{n\to\infty} \log\left(\left(\left(1+\frac{1}{n}\right)^{n}\right)^{\frac{1}{x}}\right)$$

I can now bring the exponent to the front of the log (using $\log(x^n) = n\log(x)$)

$$\log'(x) = \lim_{n\to\infty} \frac{1}{x}\log\left(\left(1+\frac{1}{n}\right)^{n}\right)$$

$$\log'(x) = \lim_{n\to\infty} \frac{\log\left(\left(1+\frac{1}{n}\right)^{n}\right)}{x}$$

$$\log'(x) = \frac{\log\left(\lim_{\mathrm{n}\to\infty}\left(\left(1+\frac{1}{n}\right)^{n}\right)\right)}{x}$$

$\lim_{\mathrm{n}\to\infty}\left(\left(1+\frac{1}{n}\right)^{n}\right)$ has now shown up in two different places. Whatever this limit approaches (assuming it converges), I will assign the letter e.

Using this definition: $e = \lim_{\mathrm{n}\to\infty}\left(\left(1+\frac{1}{n}\right)^{n}\right)$

$$\log'(x) = \frac{\log(e)}{x}$$

I said before that this is the logarithm of some arbitrary base. If I let this base be e, then:

$$\log_e{}'(x) = \frac{\log_{\mathrm{e}}(e)}{x}$$

$$\log_e{}'(x) = \frac{1}{x}$$

I will give $\log_e(x)$ a special name: the "natural logarithm" or $\ln(x)$

$$\ln'(x) = \frac{1}{x}$$

The derivative of some logarithm with arbitrary base b rewritten in terms of the natural logarithm by using $\log_a(b) = \frac{\log_c(b)}{\log_c(a)}$ where c is any number.

$$\log_b{}'(x) = \frac{\log_b(e)}{x}$$

$$\log_b{}'(x) = \frac{\log_c(e)}{\log_c(b)\,x}$$

c can be any number so I will $let\ c = e$

$$\log_b{}'(x) = \frac{\log_e(e)}{\log_e(b)\,x}$$

$$\log_b{}'(x) = \frac{1}{x\ln(b)}$$

What if I wanted to find the derivative of the sum of two functions, e.g., $\frac{d}{dx}\big(f(x) \pm g(x)\big)$?

$$Let\ j(x) = f(x) \pm g(x)$$

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

$$j'(x) = \lim_{h\to 0}\frac{f(x+h) \pm g(x+h) - f(x) \mp g(x)}{h}$$

$$j'(x) = \lim_{h\to 0}\frac{\big(f(x+h) - f(x)\big) \pm \big(g(x+h) - g(x)\big)}{h}$$

$$j'(x) = \lim_{h\to 0}\left(\frac{f(x+h) - f(x)}{h} \pm \frac{g(x+h) - g(x)}{h}\right)$$

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

$$j'(x) = f'(x) \pm g'(x)$$

$$\frac{d}{dx}\big(f(x) \pm g(x)\big) = f'(x) \pm g'(x)$$

I will next move on to the "chain rule" which will allow us to differentiate functions which contain other functions, for example $\ln(\sin(x))$. In general:

$$y = f\big(g(x)\big)$$

$$Find\ \frac{dy}{dx}$$

To do this, I can introduce a new variable, which is referred to as a "dummy variable" as it is not present in the original question nor in the final answer, but it does help to get there.

$$Let\ u = g(x)$$

I now have two equations:

$$y = f(u)$$

$$u = g(x)$$

The first could be plotted on a graph with a vertical y-axis and horizontal u-axis. This graph would have a gradient function $f'(u)$ or $\frac{dy}{du}$. The second could be plotted on a graph with a vertical u-axis and a horizontal x-axis. The gradient function of this graph would be $g'(x)$ or $\frac{du}{dx}$.

$$\frac{dy}{du} = f'(u)$$

$$\frac{du}{dx} = g'(x)$$

Remember, I wanted to find $\frac{dy}{dx}$. How could I get from $\frac{dy}{du}$ and $\frac{du}{dx}$ to $\frac{dy}{dx}$? I can multiply them together of course.

$$\frac{dy}{du} \times \frac{du}{dx} = f'(u) \times g'(x)$$

$$\frac{dy}{du} \times \frac{du}{dx} = \frac{dy\,du}{du\,dx} = \frac{dy}{dx}$$

$$So\,\frac{dy}{dx} = f'(u) \times g'(x)$$

Remember that u was just a dummy variable and now needs to be removed from the final answer. Substitute $u = g(x)$ into the above:

$$\frac{dy}{dx} = f'\big(g(x)\big) \times g'(x)$$

$$\frac{d}{dx}\Big(f\big(g(x)\big)\Big) = f'\big(g(x)\big) \times g'(x)$$

This is called the chain rule.

For example:

$$\frac{d}{dx}(\sin(\ln(x)))$$

$$= \sin'(\ln(x)) \times \ln'(x)$$

$$= \cos(\ln(x)) \times \frac{1}{x}$$

$$= \frac{\cos(\ln(x))}{x}$$

So far, I have said things like, “differentiate both sides with respect to x”, but using the chain rule, I don’t need to be so specific.

$$If\ f(y) = g(x)$$

By differentiating both sides with respect to x (the chain rule can be used on the right side because y is a function of x and its derivative with respect to x is $\frac{dy}{dx}$):

$$f'(y)\frac{dy}{dx} = g'(x)$$

$$f'(y)dy = g'(x)dx$$

Alternatively, by differentiating both sides with respect to y:

$$f'(y) = g'(x)\frac{dx}{dy}$$

$$f'(y)dy = g'(x)dx$$

Or we could differentiate both sides with respect to some dummy variable u:

$$f'(y)\frac{dy}{du} = g'(x)\frac{dx}{du}$$

$$f'(y)dy = g'(x)dx$$

However you chose to do it, you get this result. Instead of differentiating both sides with respect to some specific variable, you can instead differentiate both sides with respect to any variable, as long as you multiply the result by the differential of that variable. This isn't too useful but does occasionally allow us to cut out the middleman.

What if I wanted to differentiate a product (one function multiplied by another)?

$$y = f(x)g(x)$$

How can I find $\frac{dy}{dx}$? I don't know how to differentiate something multiplied by something add something. How can I turn a multiplication into an addition? By using logarithms.

$$\ln(y) = \ln\big(f(x)g(x)\big)$$

$$\ln(y) = \ln\big(f(x)\big) + \ln\big(g(x)\big)$$

Differentiate both sides with respect to x:

$$\frac{1}{y} \times \frac{dy}{dx} = \frac{1}{f(x)} \times f'(x) + \frac{1}{g(x)} \times g'(x)$$

$$\frac{1}{y} \times \frac{dy}{dx} = \frac{f'(x)}{f(x)} + \frac{g'(x)}{g(x)}$$

$$\frac{dy}{dx} = \frac{y\,f'(x)}{f(x)} + \frac{y\,g'(x)}{g(x)}$$

Substitute $y = f(x)g(x)$

$$\frac{dy}{dx} = \frac{f(x)\,g(x)\,f'(x)}{f(x)} + \frac{f(x)\,g(x)\,g'(x)}{g(x)}$$

$$\frac{dy}{dx} = f'(x)g(x) + f(x)g'(x)$$

This is called the product rule.

The same can be done for $y = \frac{f(x)}{g(x)}$

$$\ln(y) = \ln\left(\frac{f(x)}{g(x)}\right)$$

$$\ln(y) = \ln\big(f(x)\big) - \ln\big(g(x)\big)$$

$$\frac{1}{y} \times \frac{dy}{dx} = \frac{1}{f(x)} \times f'(x) - \frac{1}{g(x)} \times g'(x)$$

$$\frac{1}{y} \times \frac{dy}{dx} = \frac{f'(x)}{f(x)} - \frac{g'(x)}{g(x)}$$

$$\frac{dy}{dx} = y \times \frac{f'(x)}{f(x)} - y \times \frac{g'(x)}{g(x)}$$

$$\frac{dy}{dx} = \frac{f(x)}{g(x)} \times \frac{f'(x)}{f(x)} - \frac{f(x)}{g(x)} \times \frac{g'(x)}{g(x)}$$

$$\frac{dy}{dx} = \frac{f'(x)}{g(x)} - \frac{f(x)\,g'(x)}{g^2(x)}$$

$$\frac{dy}{dx} = \frac{f'(x)\,g(x)}{g^2(x)} - \frac{f(x)\,g'(x)}{g^2(x)}$$

$$\frac{dy}{dx} = \frac{f'(x)\,g(x) - f(x)\,g'(x)}{g^2(x)}$$

This is called the quotient rule.

We can find the derivative of x^n in a very similar way:

$$y = x^n$$

$$\ln(y) = \ln(x^n)$$

$$\ln(y) = n\ln(x)$$

$$\frac{1}{y} \times \frac{dy}{dx} = n \times \frac{1}{x}$$

$$\frac{dy}{dx} = n \times \frac{y}{x}$$

$$\frac{dy}{dx} = n \times \frac{x^n}{x}$$

$$\frac{dy}{dx} = n \times x^{n-1}$$

$$So\,\frac{d}{dx}(x^n) = n\,x^{n-1}$$

Using this, $1 = x^0$ and so $\frac{d}{dx}1 = \frac{d}{dx}x^0 = 0 \times x^{-1} = \frac{0}{x} = 0$. Because $\frac{d}{dx}\left(k \times f(x)\right) = k \times \frac{d}{dx}f(x)$, the derivative of any constant is zero.

This method can also be used to prove that $\frac{d}{dx}e^x = e^x$. Even though I have already shown that another way, it could also be done like this:

$$y = e^x$$

$$\ln(y) = \ln(e^x)$$

$$\ln(y) = x$$

$$\frac{1}{y} \times \frac{dy}{dx} = 1$$

$$\frac{dy}{dx} = y$$

$$\frac{dy}{dx} = e^x$$

Moving on from these rules for differentiation, I will now use them to find the derivatives of the remaining trigonometric functions as well as their inverses.

$$\tan'(x)$$

$$= \frac{d}{dx}\left(\frac{\sin(x)}{\cos(x)}\right)$$

$$= \frac{\sin'(x)\cos(x) - \sin(x)\cos'(x)}{\cos^2(x)}$$

$$= \frac{\cos(x)\cos(x) + \sin(x)\sin(x)}{\cos^2(x)}$$

$$= \frac{\cos^2(x) + \sin^2(x)}{\cos^2(x)}$$

$$= \frac{1}{\cos^2(x)}$$

$$= \sec^2(x)$$

$$So\ \tan'(x) = \sec^2(x)$$

$$\cot'(x)$$

$$= \frac{d}{dx}\left(\frac{\cos(x)}{\sin(x)}\right)$$

$$= \frac{\cos'(x)\sin(x) - \cos(x)\sin'(x)}{\sin^2(x)}$$

$$= \frac{-\sin(x)\sin(x) - \cos(x)\cos(x)}{\sin^2(x)}$$

$$= \frac{-(\sin^2(x) + \cos^2(x))}{\sin^2(x)}$$

$$= \frac{-1}{\sin^2(x)}$$

$$= -\csc^2(x)$$

$$\sec'(x)$$

$$= \frac{d}{dx}\left(\frac{1}{\cos(x)}\right)$$

$$= \frac{1'\cos(x) - 1\cos'(x)}{\cos^2(x)}$$

$$= \frac{0\cos(x) + 1\sin(x)}{\cos^2(x)}$$

$$= \frac{\sin(x)}{\cos(x) \times \cos(x)}$$

$$= \frac{\sin(x)}{\cos(x)} \times \frac{1}{\cos(x)}$$

$$= \tan(x)\sec(x)$$

$$So\ \sec'(x) = \tan(x)\sec(x)$$

$$\csc'(x)$$

$$= \frac{d}{dx}\left(\frac{1}{\sin(x)}\right)$$

$$= \frac{1'\sin(x) - 1\,sin'(x)}{\sin^2(x)}$$

$$= \frac{0\sin(x) - 1\cos(x)}{\sin^2(x)}$$

$$= \frac{-\cos(x)}{\sin(x) \times \sin(x)}$$

$$= -\frac{\cos(x)}{\sin(x)} \times \frac{1}{\sin(x)}$$

$$= -\cot(x)\csc(x)$$

Finally, the derivatives of the inverse trigonometric functions. The inverse functions are defined such that if $y = \sin(x)$ then $x = \arcsin(y)$ etc. These functions are only defined within a domain which prevents them from being considered one-to-many.

$$y = \arcsin(x)$$

$$\sin(y) = x$$

$$\cos(y)\frac{dy}{dx} = 1$$

$$\frac{dy}{dx} = \frac{1}{\cos(y)}$$

$\sin^2(y) + \cos^2(y) = 1$ so $\cos(y) = \sqrt{1 - \sin^2(y)}$

$$\frac{dy}{dx} = \frac{1}{\sqrt{1 - \sin^2(y)}}$$

Since $\sin(y) = x$, this means that:

$$\frac{dy}{dx} = \frac{1}{\sqrt{1 - x^2}}$$

$$\arcsin'(x) = \frac{1}{\sqrt{1 - x^2}}$$

$$y = \arccos(x)$$

$$\cos(y) = x$$

$$-\sin(y)\frac{dy}{dx} = 1$$

$$\frac{dy}{dx} = -\frac{1}{\sin(y)}$$

$\sin^2(y) + \cos^2(y) = 1$ so $\cos(y) = \sqrt{1 - \cos^2(x)}$

$$\frac{dy}{dx} = -\frac{1}{\sqrt{1 - \cos^2(y)}}$$

and $\cos(y) = x$ so

$$\frac{dy}{dx} = -\frac{1}{\sqrt{1 - x^2}}$$

$$\arccos'(x) = -\frac{1}{\sqrt{1 - x^2}}$$

$$y = \arctan(x)$$

$$\tan(y) = x$$

$$\sec^2(y)\frac{dy}{dx} = 1$$

$$\frac{dy}{dx} = \frac{1}{\sec^2(y)}$$

$\tan^2(y) + 1 = \sec^2(y)$ so

$$\frac{dy}{dx} = \frac{1}{1+\tan^2(y)} = \frac{1}{1+x^2}$$

$$\arctan'(x) = \frac{1}{1+x^2}$$

$$y = \operatorname{arccot}(x)$$

$$\cot(y) = x$$

$$-\csc^2(y)\frac{dy}{dx} = 1$$

$$\frac{dy}{dx} = \frac{-1}{\csc^2(y)}$$

$1 + \cot^2(y) = \csc^2(y)$ so

$$\frac{dy}{dx} = \frac{-1}{1+\cot^2(y)} = \frac{-1}{1+x^2}$$

$$\operatorname{arccot}'(x) = \frac{-1}{1+x^2}$$

$$y = \operatorname{arcsec}(x)$$

$$\sec(y) = x$$

$$\tan(y)\sec(y)\frac{dy}{dx} = 1$$

$$\frac{dy}{dx} = \frac{1}{\tan(y)\sec(y)} = \frac{1}{x\tan(y)}$$

$\tan^2(y) + 1 = \sec^2(y)$ so $\tan(y) = \sqrt{\sec^2(y) - 1}$

$$\frac{dy}{dx} = \frac{1}{x\sqrt{\sec^2(y)-1}} = \frac{1}{x\sqrt{x^2-1}}$$

$$\operatorname{arcsec}'(x) = \frac{1}{x\sqrt{x^2-1}}$$

$$y = \operatorname{arccsc}(x)$$

$$\csc(y) = x$$

$$-\cot(y)\csc(y)\frac{dy}{dx} = 1$$

$$\frac{dy}{dx} = \frac{-1}{\cot(y)\csc(y)} = \frac{-1}{x\cot(y)}$$

$1 + \cot^2(y) = \csc^2(y)$ so $\cot(y) = \sqrt{\csc^2(y) - 1}$

$$\frac{dy}{dx} = \frac{-1}{x\sqrt{\csc^2(y) - 1}} = \frac{-1}{x\sqrt{x^2 - 1}}$$

$$\text{arccsc}'(x) = \frac{-1}{x\sqrt{x^2 - 1}}$$

Derivatives can be used to find the gradient of a curve at some point. They can also be used to find the stationary points of a curve (the points where the gradient is zero). The second derivative can be used to find the rate of change of the gradient. If this is positive, then the gradient is increasing and the curve is convex at that point and if it is negative, then the gradient is decreasing and so the curve is concave at that point. One other, very useful, application of derivatives is Maclaurin and, more generally, Taylor series.

If some function $f(x)$ could be written as an infinite polynomial (a polynomial with an infinite number of terms in it), you can sometimes use a Maclaurin series to find what this polynomial would be. To do this:

$$Let\ f(x) = c_0 + c_1x + c_2x^2 + c_3x^3 + c_4x^4 + c_5x^5 + \cdots$$

where $c_0, c_1, c_2, \ldots$ are constants to be found.

We want the polynomial on the right to be identical to the function on the left. In other words, we want them to have the same value for a given value of x. This means that I can let x equal any number since we want the above to be true for any value of x. c_0 can then be easily found by letting $x = 0$ as this makes all other terms disappear.

$$f(0) = c_0 + c_1(0) + c_2(0)^2 + c_3(0)^3 + c_4(0)^4 + c_5(0)^5 + \cdots$$

$$c_0 = f(0)$$

If the curve $y = f(x)$ is the same curve as the endless polynomial curve, then the gradients of either curve at a given point would be the same. This means that both sides can be differentiated with respect to x.

$$f'(x) = c_1 + 2c_2x + 3c_3x^2 + 4c_4x^3 + 5c_5x^4 + \cdots$$

$$Let\ x = 0$$

$$f'(0) = c_1 + 2c_2(0) + 3c_3(0)^2 + 4c_4(0)^3 + 5c_5(0)^4 + \cdots$$

$$c_1 = f'(0)$$

Differentiating both sides again ($f''(x)$ means the second derivative of $f(x)$ or the derivative of the derivative of $f(x)$)

$$f''(x) = 2c_2 + 2 \times 3c_3x + 3 \times 4c_4x^2 + 4 \times 5c_5x^3 + \cdots$$

$$Let\ x = 0$$

$$f''(0) = 2c_2 + 2 \times 3c_3(0) + 3 \times 4c_4(0)^2 + 4 \times 5c_5(0)^3 + \cdots$$

$$2c_2 = f''(x)$$

$$c_2 = \frac{f''(x)}{2}$$

Writing all of the $'$s will quickly become annoying to do and even more annoying to read, so I will introduce a new notation: $f^{(n)}(x)$ meaning the nth derivative of $f(x)$

$$f^{(3)}(x) = 2 \times 3c_3 + 2 \times 3 \times 4c_4x + 3 \times 4 \times 5c_5x^2 + \cdots$$

$$Let\ x = 0$$

$$f^{(3)}(0) = 2 \times 3c_3 + 2 \times 3 \times 4c_4(0) + 3 \times 4 \times 5c_5(0)^2 + \cdots$$

$$2 \times 3c_3 = f^{(3)}(0)$$

$$c_3 = \frac{f^{(3)}(0)}{2 \times 3}$$

$$f^{(4)}(x) = 2 \times 3 \times 4c_4 + 2 \times 3 \times 4 \times 5c_5x + \cdots$$

$$Let\ x = 0$$

$$f^{(4)}(0) = 2 \times 3 \times 4c_4 + 2 \times 3 \times 4 \times 5c_5(0) + \cdots$$

$$2 \times 3 \times 4c_4 = f^{(4)}(0)$$

$$c_4 = \frac{f^{(4)}(0)}{2 \times 3 \times 4}$$

$$f^{(5)}(x) = 2 \times 3 \times 4 \times 5c_5 + \cdots$$

$$Let\ x = 0$$

$$f^{(5)}(0) = 2 \times 3 \times 4 \times 5c_5 + \cdots$$

$$2 \times 3 \times 4 \times 5c_5 = f^{(5)}(0)$$

$$c_5 = \frac{f^{(5)}(0)}{2 \times 3 \times 4 \times 5}$$

Substituting the values of c_1, c_2, c_3 etc., back into the original equation:

$$f(x) = c_0 + c_1x + c_2x^2 + c_3x^3 + c_4x^4 + c_5x^5 + \cdots$$

$$f(x) = f(0) + f'(0)x + \frac{f''(x)}{2}x^2 + \frac{f^{(3)}(0)}{2 \times 3}x^3 + \frac{f^{(4)}(0)}{2 \times 3 \times 4}x^4 + \frac{f^{(5)}(0)}{2 \times 3 \times 4 \times 5}x^5 + \cdots$$

To clean up this expression I will create a new notation, the factorial notation. $x!$ (x factorial) $= 1 \times 2 \times 3 \times 4 \times \ldots \times (x-1) \times x$

Using this new notation, as well as using the $f^{(n)}(x)$ notation, the equation becomes:

$$f(x) = f(0) + f^{(1)}(0)x + \frac{f^{(2)}(x)}{2}x^2 + \frac{f^{(3)}(0)}{3!}x^3 + \frac{f^{(4)}(0)}{4!}x^4 + \frac{f^{(5)}(0)}{5!}x^5 + \cdots$$

There is a clear pattern here: the nth term in the series (if n starts at 0) is $\frac{f^{(n)}(x)}{n!}x^n$. This is certainly true it seems for the x^3 terms and onwards and the second and third terms can be rewritten to fit the pattern.

$$f(x) = \frac{f^{(0)}(0)}{1}x^0 + \frac{f'(0)}{1!}x^1 + \frac{f''(x)}{2!}x^2 + \frac{f^{(3)}(0)}{2 \times 3}x^3 + \frac{f^{(4)}(0)}{2 \times 3 \times 4}x^4 + \frac{f^{(5)}(0)}{2 \times 3 \times 4 \times 5}x^5 + \cdots$$

$f(0) = f^{(0)}(0)$ because the $0th$ derivative of $f(x)$ just means that $f(x)$ has been differentiated 0 times, and so it is just $f(x)$.

The only problem is the first term, which is close, but would require $0!$ in the denominator to be exactly what we wanted, but 1 is in the denominator instead. $0!$ is undefined by the current factorial definition, so to make this series fit the pattern I will extend the definition of the factorial function to include $0! = 1$ which means that:

$$f(x) = \frac{f^{(0)}(0)}{0!}x^0 + \frac{f'(0)}{1!}x^1 + \frac{f''(x)}{2!}x^2 + \frac{f^{(3)}(0)}{2 \times 3}x^3 + \frac{f^{(4)}(0)}{2 \times 3 \times 4}x^4 + \frac{f^{(5)}(0)}{2 \times 3 \times 4 \times 5}x^5 + \cdots$$

To write this using sigma notation:

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

As mentioned earlier, this only works if the function can be written as polynomial. A non-continuous function for example cannot be written as a polynomial because polynomials are continuous, therefore non-continuous functions do not have a Maclaurin series. Polynomials are also defined for all values of x, meaning functions such as $\ln(x)$ also have no Maclaurin series because a polynomial would have a value for negative values of x but $\ln(x)$ does not. Also $\ln(x)$ is not defined where $x = 0$, nor are any of its derivatives meaning a Maclaurin series could not be formed anyway. e^x has a value for all values of x and is a continuous function, so I can make a Maclaurin series with it.

$$Let\ f(x) = e^x$$

$$e^x = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!}x^n$$

$$e^x = \frac{f^{(0)}(0)}{0!}x^0 + \frac{f^{(1)}(0)}{1!}x^1 + \frac{f^{(2)}(0)}{2!}x^2 + \frac{f^{(3)}(0)}{3!}x^3 + \frac{f^{(4)}(0)}{4!}x^4 + \cdots$$

Since e^x is its own derivative, all derivatives of e^x are e^x.

$$e^x = \frac{e^0}{0!}x^0 + \frac{e^0}{1!}x^1 + \frac{e^0}{2!}x^2 + \frac{e^0}{3!}x^3 + \frac{e^0}{4!}x^4 + \cdots$$

$$e^x = \frac{1}{0!}x^0 + \frac{1}{1!}x^1 + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \frac{1}{4!}x^4 + \cdots$$

$$e^x = \frac{x^0}{0!} + \frac{x^1}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots$$

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

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \frac{x^5}{120} + \cdots$$

This is the Maclaurin series for e^x. By letting $x = 1$:

$$e = \frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \frac{1}{5!} + \cdots$$

$$e = 1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \frac{1}{120} + \cdots$$

$$e = \sum_{n=0}^{\infty} \frac{1}{n!}$$

Another function which is continuous and has a value for every value of x is the sine function meaning I can create a Maclaurin series for it.

$$Let\ f(x) = \sin(x)$$

$$f^{(1)}(x) = \cos(x)$$

$$f^{(2)}(x) = -\sin(x)$$

$$f^{(3)}(x) = -\cos(x)$$

$$f^{(4)}(x) = \sin(x)$$

$$f^{(5)}(x) = \cos(x)$$

$\sin(0) = 0$ meaning that all of the even numbered derivatives will be zero. $\cos(0) = 1$ so, the rest of them will alternate between 1 and -1.

$$\sin(x) = \frac{f^{(0)}(0)}{0!}x^0 + \frac{f^{(1)}(0)}{1!}x^1 + \frac{f^{(2)}(0)}{2!}x^2 + \frac{f^{(3)}(0)}{3!}x^3 + \frac{f^{(4)}(0)}{4!}x^4 + \frac{f^{(5)}(0)}{5!} + \cdots$$

$$\sin(x) = \frac{\sin(0)}{0!}x^0 + \frac{\cos(0)}{1!}x^1 + \frac{-\sin(0)}{2!}x^2 + \frac{-\cos(0)}{3!}x^3 + \frac{\sin(0)}{4!}x^4 + \frac{\cos(0)}{5!} + \cdots$$

$$\sin(x) = \frac{0}{0!}x^0 + \frac{1}{1!}x^1 + \frac{0}{2!}x^2 + \frac{-1}{3!}x^3 + \frac{0}{4!}x^4 + \frac{1}{5!} + \cdots$$

$$\sin(x) = \frac{x^1}{1!} - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots$$

To write this in sigma notation, each term is $\frac{x^{2n+1}}{(2n+1)!}$ (because the exponents and fractions are the odd numbers) but every odd term is negative, and every even term is positive. Multiplying each term by $(-1)^n$ will make every odd term negative and every even term positive as required. (The term $\left(\frac{x^1}{1!}\right)$ is positive and is considered the 0^{th} term so is an even term.)

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

This Maclaurin series can be used to find very good decimal approximations for $\sin(x)$ for any value of x. It works best if x is smaller though, which is fine because the graph flips every π radians and so only values between 0 and π need to be found using this series in order to obtain all the information about the whole graph.

The same can all also be said about this cosine function.

$$Let\ f(x) = \cos(x)$$

$$f^{(1)}(x) = -\sin(x)$$

$$f^{(2)}(x) = -\cos(x)$$

$$f^{(3)}(x) = \sin(x)$$

$$f^{(4)}(x) = \cos(x)$$

$$f^{(5)}(x) = -\sin(x)$$

$\sin(0) = 0$ so the odd derivatives will become 0 and $\cos(0) = 1$ so the even derivative will alternate between 1 and -1.

$$\cos(x) = \frac{f^{(0)}(0)}{0!}x^0 + \frac{f^{(1)}(0)}{1!}x^1 + \frac{f^{(2)}(0)}{2!}x^2 + \frac{f^{(3)}(0)}{3!}x^3 + \frac{f^{(4)}(0)}{4!}x^4 + \frac{f^{(5)}(0)}{5!} + \cdots$$

$$\cos(x) = \frac{\cos(0)}{0!}x^0 + \frac{-\sin(0)}{1!}x^1 + \frac{-\cos(0)}{2!}x^2 + \frac{\sin(0)}{3!}x^3 + \frac{\cos(0)}{4!}x^4 + \frac{-\sin(0)}{5!} + \cdots$$

$$\cos(x) = \frac{1}{0!}x^0 + \frac{0}{1!}x^1 + \frac{-1}{2!}x^2 + \frac{0}{3!}x^3 + \frac{1}{4!}x^4 + \frac{0}{5!} + \cdots$$

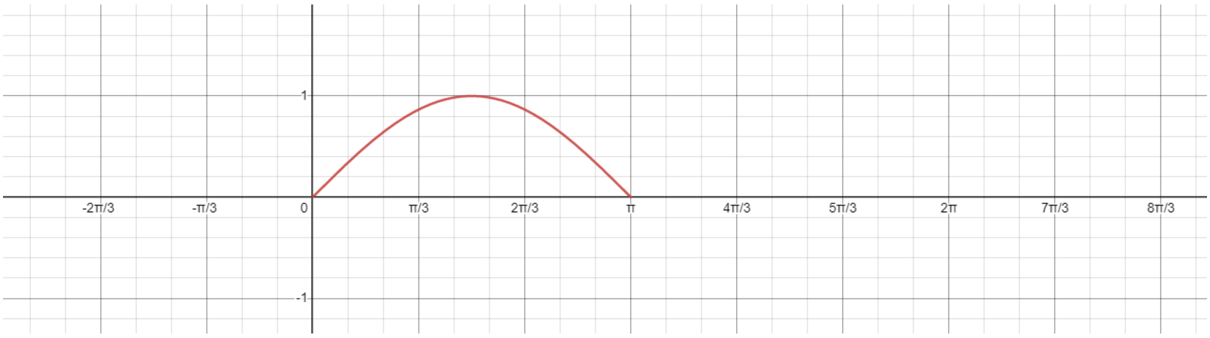
$$\cos(x) = \frac{x^0}{0!} - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots$$

Writing this in sigma notation, the exponents and factorials are the even numbers so each term will be $\frac{x^{2n}}{(2n)!}$ and the even terms are positive, and the odd terms are negative, so each term needs to be multiplied by $(-1)^n$.

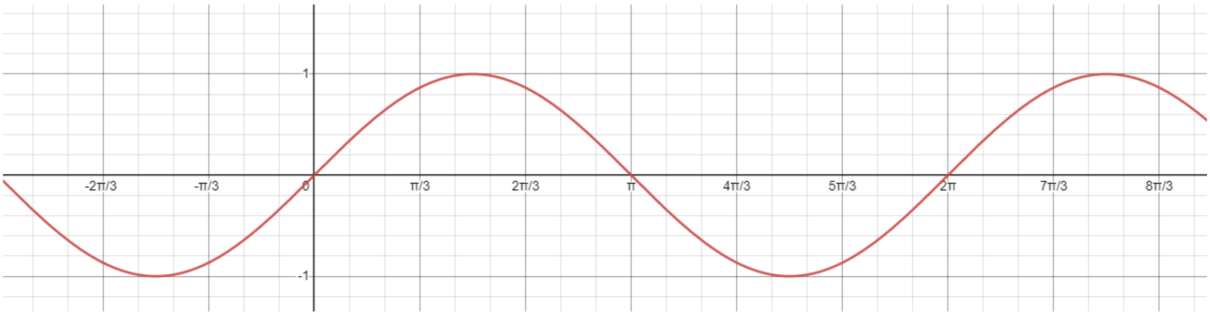
$$\cos(x) = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!}$$

The Maclaurin series for sine and cosine can be used to fill the gaps on their graphs from 0 to π radians and hence the rest of the graphs can be completed.

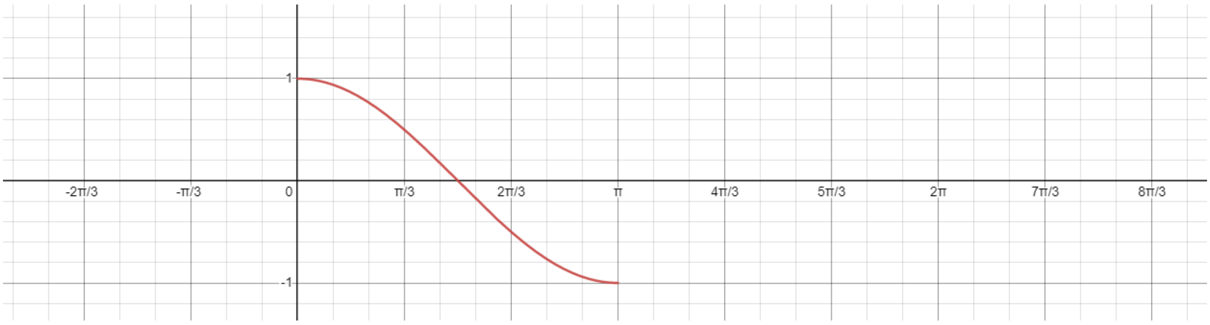
Sine graph from 0 to π using Maclaurin series:



Using this to complete the entire graph:

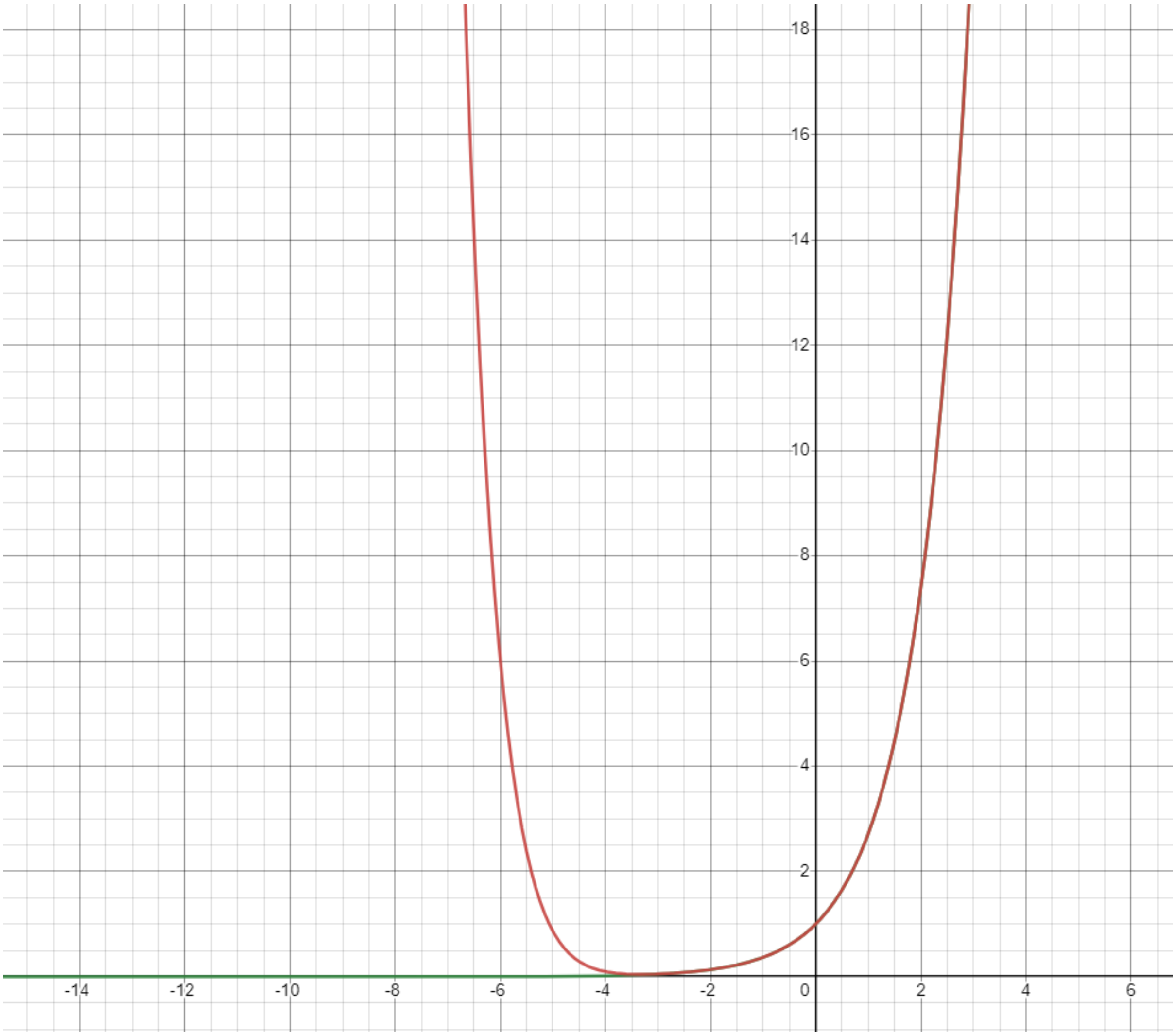


Cosine graph from 0 to π using Maclaurin series:



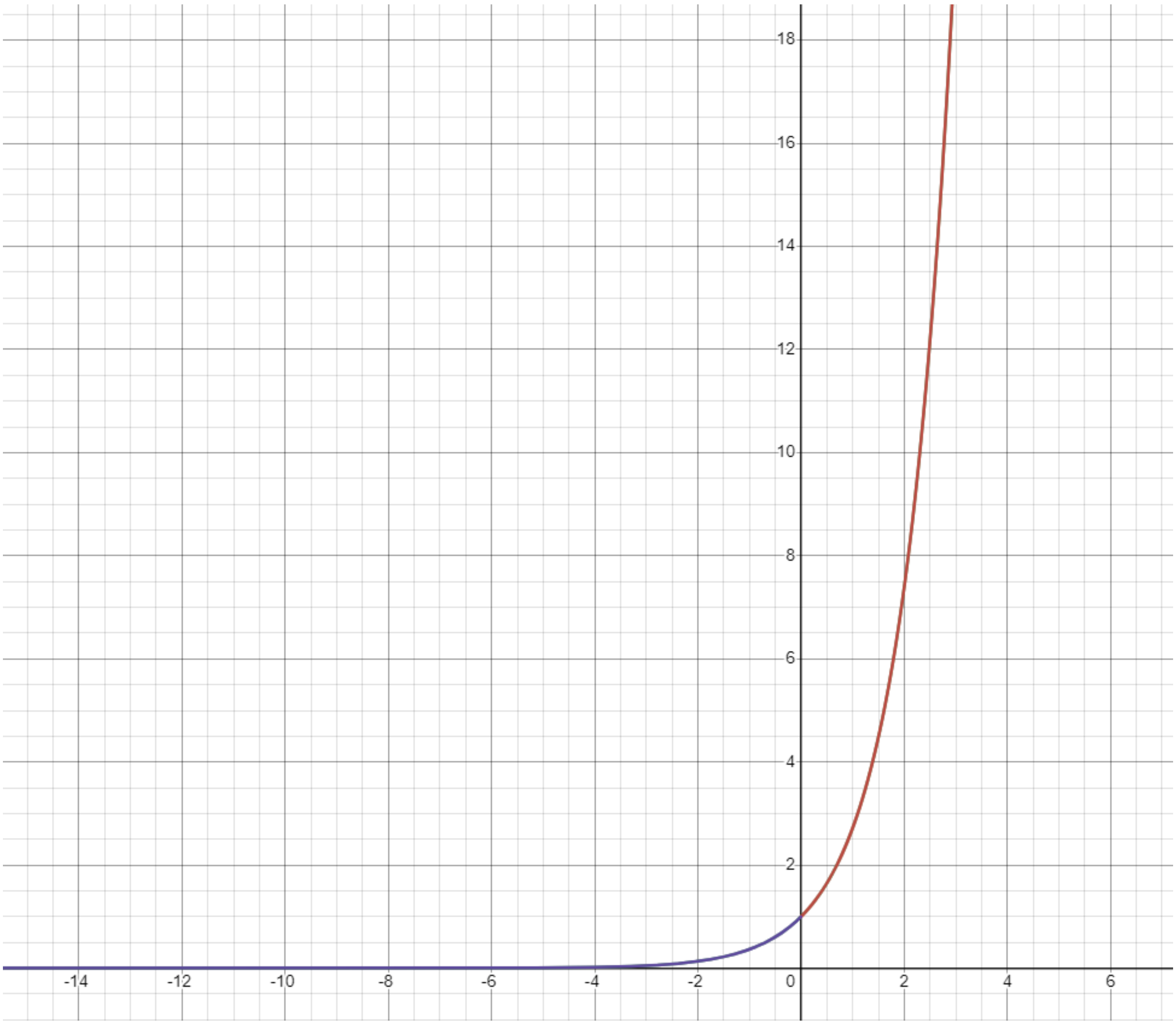
Using this to complete the graph:

The graph of $y = e^x$ can also be plotted using its Maclaurin series. The green shows the true graph, and the red shows the graph produced by the Maclaurin series with 10 terms.



As you can see, the approximation from the Maclaurin series is much better when $x > 0$ meaning when finding the value of $y = e^x$ for negative x values it would be more efficient to remember that $e^{-x} = \frac{1}{e^x}$ and use the Maclaurin series to more efficiently find a precise value for the function at a positive value of x and take its reciprocal to find the value of the function at a negative value of x.

After doing this, you may get a graph which looks more like this one:



The red was plotted using:

$$y = \sum_{n=0}^{10} \frac{x^n}{n!} \; for\; x > 0$$

The purple was plotted using:

$$y = \frac{1}{\sum_{n=0}^{10} \frac{(-x)^n}{n!}} \; for\; x < 0$$

which, for some negative x value, takes the reciprocal of the positive value of x.

The green line was still there when plotting these curves, but it is no longer visible because these approximations are so precise with just 10 terms.

The more general form of a Maclaurin series is called a Taylor series. Instead of being centred at $x = 0$ a Taylor series is centred at $x = a$, that is to say, it can be centred anywhere. Assuming $f(x)$ can be written as a polynomial:

$$f(x) = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + c_4 x^4 + \cdots$$

$$f(a) = c_0 + c_1 a + c_2 a^2 + c_3 a^3 + c_4 a^4 + \cdots$$

The problem here is that, unlike the Maclaurin series, the as do not allow the terms to disappear. The way to solve this is by writing the original polynomial in a different form.

$$f(x) = c_0 + c_1(x-a) + c_2(x-a)^2 + c_3(x-a)^3 + c_4(x-a)^4 + \cdots$$

$$f(a) = c_0 + c_1(a-a) + c_2(a-a)^2 + c_3(a-a)^3 + c_4(a-a)^4 + \cdots$$

$$f(a) = c_0 + c_1(0) + c_2(0)^2 + c_3(0)^3 + c_4(0)^4 + \cdots$$

$$f(a) = c_0$$

Differentiating both sides of $f(x) = c_0 + c_1(x-a) + c_2(x-a)^2 + c_3(x-a)^3 + c_4(x-a)^4 + \cdots$ (The left can be differentiated using the chain rule. The derivative of $x - a$ with respect to x is 1.):

$$f'(x) = c_1 + 2!\,c_2(x-a) + 3c_3(x-a)^2 + 4c_4(x-a)^3 + \cdots$$

$$f'(a) = c_1$$

$$f^{(2)}(x) = 2!\,c_2 + 3!\,c_3(x-a) + 3 \times 4c_4(x-a)^2 + \cdots$$

$$f^{(2)}(a) = 2!\,c_2$$

$$c_2 = \frac{f^{(2)}(a)}{2!}$$

$$f^{(3)}(x) = 3!\,c_3 + 4!\,c_4(x-a) + \cdots$$

$$f^{(3)}(a) = 3!\,c_3$$

$$c_3 = \frac{f^{(3)}(a)}{3!}$$

$$f^{(4)}(x) = 4!\,c_4 + \cdots$$

$$f^{(4)}(a) = 4!\,c_4$$

$$c_4 = \frac{f^{(4)}(a)}{4!}$$

Substituting the values of $c_1, c_2, c_3\ etc.$, into the original:

$$f(x) = f(a) + f'(a)(x-a) + \frac{f^{(2)}(a)}{2!}(x-a)^2 + \frac{f^{(3)}(a)}{3!}(x-a)^3 + \frac{f^{(4)}(a)}{4!}(x-a)^4 + \cdots$$

$$f(x) = \frac{f^{(0)}(a)(x-a)^0}{0!} + \frac{f^{(1)}(a)(x-a)^1}{1!} + \frac{f^{(2)}(a)}{2!}(x-a)^2 + \frac{f^{(3)}(a)}{3!}(x-a)^3 + \cdots$$

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

This is the general Taylor series, centred at $x = a$. Letting $a = 0$ gives you the general Maclaurin series because a Maclaurin series is just a Taylor series centred at $x = 0$.

This can be used to find series for functions which have no value where $x = 0$ such at the natural logarithm. There are multiple different Taylor series for a given function, centred at different x values which are useful in different situations. I will not be going into these Taylor series in detail here.

I will finish this chapter by talking about a problem involving limits. Often when dealing with limits, when you simply let the limiting variable equal what it is approaching, you get $\frac{0}{0}$ which is an indeterminant form (there may be an answer, but this does not help you get there). In other words, if trying to find:

$$\lim_{x \to c} \frac{f(x)}{g(x)}$$

By letting $x = c$ you get $\frac{0}{0}$. Or to be more mathematical:

$$f(c) = 0 \,\&\, g(c) = 0$$

So, this is our problem:

$$find \lim_{x \to c} \frac{f(x)}{g(x)}$$

$$when\, f(c) = 0$$

$$and\, g(c) = 0$$

Where have we seen a limit as $x \to c$ before? The second difference quotient which I have not yet used!

$$\lim_{x \to c} f'(x) = \lim_{x \to c} \frac{f(x) - f(c)}{x - c}$$

$$and \lim_{x \to c} g'(x) = \lim_{x \to c} \frac{g(x) - g(c)}{x - c}$$

Where $f(x)$ and $g(x)$ are continuous.

Trying to get the original expression in terms of these:

$$\lim_{x \to c} \frac{f(x)}{g(x)} = \lim_{x \to c} \frac{f(x) - f(c)}{g(x) - g(c)}$$

I am allowed to do this because $f(c) = 0$ and $g(c) = 0$.

Dividing the top and bottom by $(x - c)$:

$$\lim_{x \to c} \frac{f(x)}{g(x)} = \lim_{x \to c} \frac{\left(\frac{f(x) - f(c)}{x - c}\right)}{\left(\frac{g(x) - g(c)}{x - c}\right)}$$

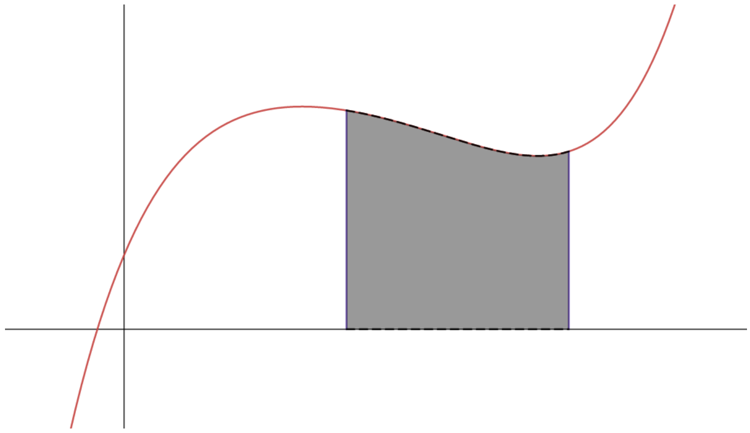
$$\lim_{x \to c} \frac{f(x)}{g(x)} = \frac{f'(x)}{g'(x)}$$

This is called L'Hospital's Rule (or L'Hôpital's Rule).

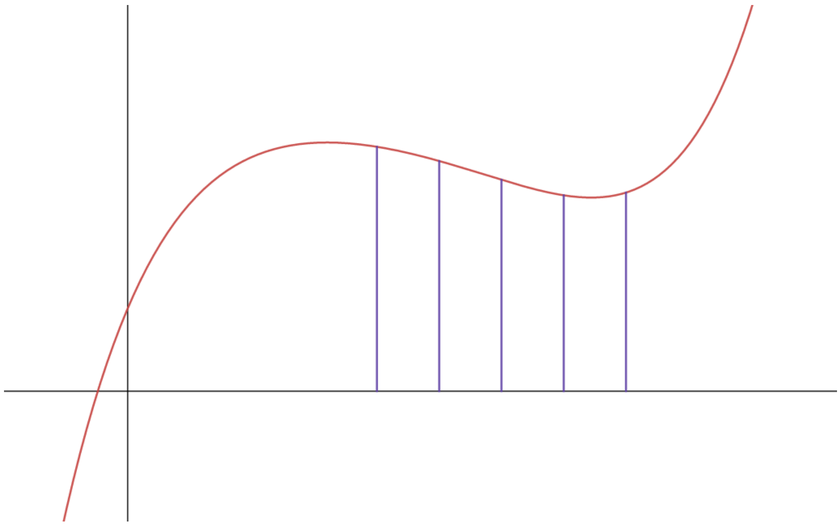
Remember, this is only true given that our assumptions are true: $f(x)$ and $g(x)$ are continuous and $f(c) = g(c) = 0$.

Calculus Part 3: The Integral

Integration is all about find the area under a given curve. For example, here is the curve of an arbitrary continuous function $y = f(x)$, with a shaded region bound by the x-axis, the lines $x = a$ and $x = b$ and the curve $y = f(x)$.

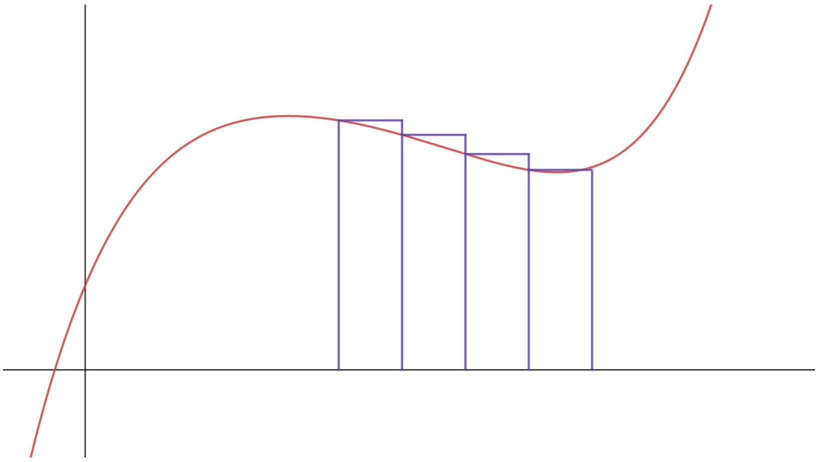


To find the area of this shaded region, it could be approximated by a series of trapeziums or rectangles. Trapeziums would leave less between themselves and the curve as well as less overlap. Rectangles though would be considerably easier to work with as the formula for the area of a rectangle is much simpler. Both would approach the exact area as the number of strips increases (and necessarily the length of each strip decreases).

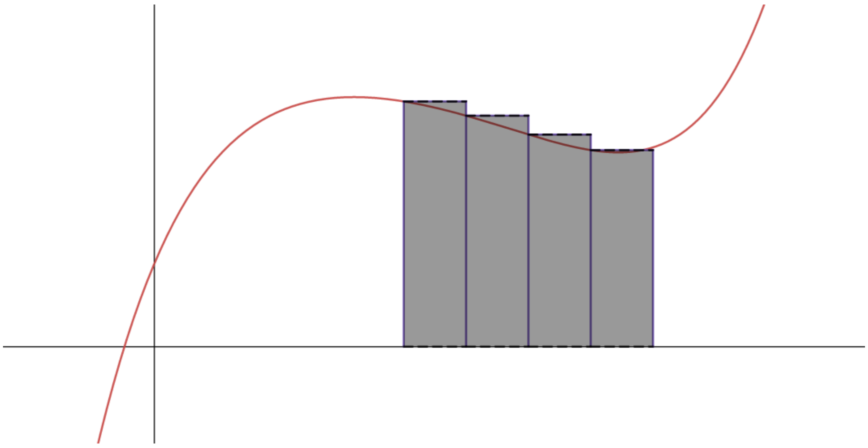


Here I have split the region into strips.

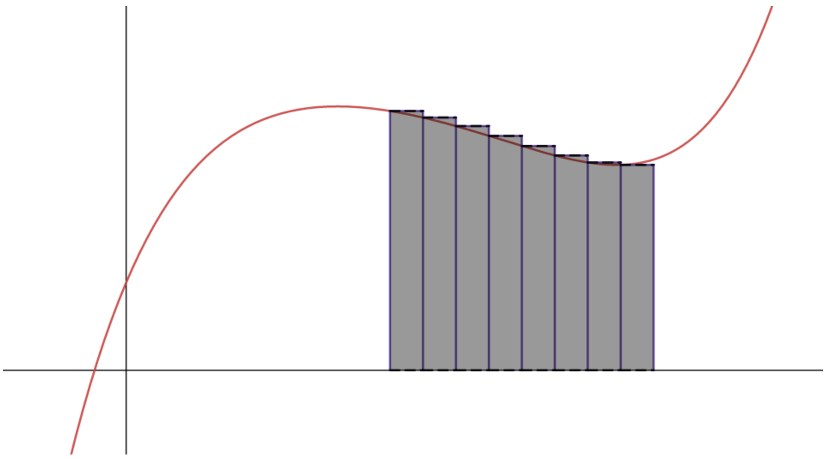
By turning each of these strips into rectangles:



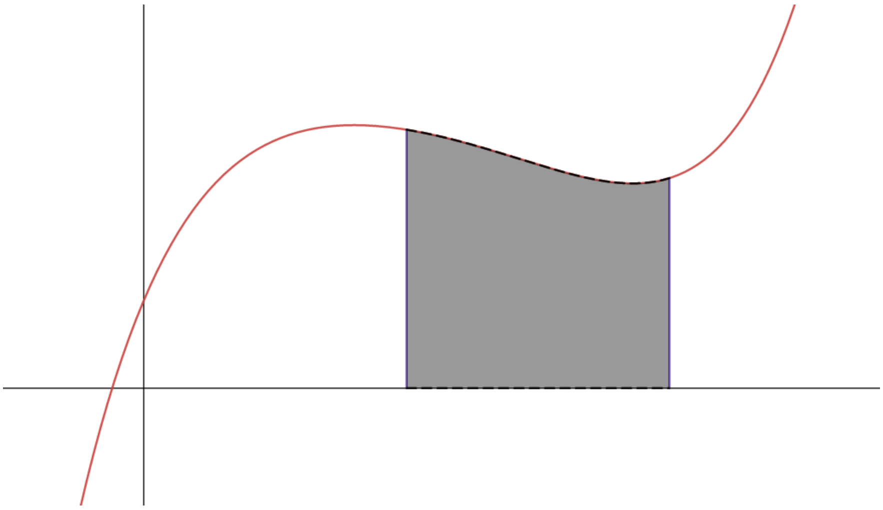
The area under the curve on the interval can be approximated.



As the number of strips increases the approximation becomes more precise.



As the number of strips approaches infinity and the width of each strip approaches zero, the sum of their areas approaches the exact area under the curve.



If I split area under the curve $y = f(x)$ from $x = a$ to $x = b$ into rectangles width h, the first rectangle would have a height $f(a)$ and width h and hence an area $h \times f(a)$. The next rectangle would be h further to the right than the first one was and would hence have a height $f(a + h)$ and a width h. It would have an area of $h \times f(a + h)$. The next rectangle would have an area of $h \times f(a + 2h)$ and so on. The last rectangle would have an area of $h \times f(b - h)$. How many rectangles would there be in total? This would be the width of the region divided by the width of each rectangle, that is $\frac{b-a}{h}$. The sum of areas of the rectangles would then be:

$$h \times f(a) + h \times f(a + h) + h \times f(a + 2h) + h \times f(a + 3h) + \cdots + h \times f(b - h)$$

$$= h\big(f(a) + f(a + h) + f(a + 2h) + f(a + 3h) + \cdots + f(b - h)\big)$$

To write this in sigma notation the general form of each term (starting at $n = 0$) is $h \times f(a + nh)$. There are $\frac{b-a}{h}$ terms in total, but because we are starting at $n = 0$ the sigma notation should take n from 0 to $\frac{b-a}{h} - 1$. Another way of doing this would be to find the value of n at the last term:

$$h \times f(b - h) = h \times f(a + nh)$$

$$f(b - h) = f(a + nh)$$

$$b - h = a + nh$$

$$nh = b - a - h$$

$$n = \frac{b - a - h}{h}$$

$$n = \frac{b - a}{n} - 1$$

$$Area \approx \sum_{n=0}^{\frac{b-a}{h}-1} h \times f(a + nh) = h \sum_{n=0}^{\frac{b-a}{h}-1} f(a + nh)$$

As the width of each rectangle approaches zero:

$$Area = \lim_{h \to 0} h \sum_{n=0}^{\frac{b-a}{h}-1} f(a+nh)$$

h represents a change in x and is approaching 0 meaning it could have been written as the differential dx.

$$Area = dx \sum_{n=0}^{\frac{b-a}{dx}-1} f(a+n\,dx)$$

This is more commonly written as the Riemann Sum:

$$\lim_{n \to \infty} \sum_{i=1}^{n} f(x_i^*)\Delta x$$

Where Δx is the width of each rectangle (which I called h or dx) x_i^* represents the values of x for different rectangles. Both mean the same thing.

The way it is written is not important, nor are the specific details of the summation. The important thing is the concept. The area under a curve can be found by summing an infinite number of infinitesimally small areas together. This is called continuous summation (as opposed to discrete summation which is what we are thus far used to) and can be represented by a tall letter s for sum $\int$. The area under a curve $y = f(x)$ from a to b is written using this notation as:

$$\int_a^b f(x)\,dx$$

This is called a definite integral.

With this notation the differential is typically written on the inside of the integral instead of on the outside.

This concept is called integration and is at the heart of this chapter.

One final note about integration for now:

$$\int_a^c f(x)dx - \int_a^b f(x)dx = \int_b^c f(x)dx$$

because the area from a to c minus the area from a to b must of course be the area from b to c.

Before continuing with integration, I will take a detour to talk about something seemingly irrelevant, averages.

An average is a number which proves, in a general sense, an idea of the typical value of each item in a set. There are many different types of averages, the mode (the most common item in the set), the median (the item in the middle of the set when the set is order in ascending or descending order) and the mean. There are different types of mean, including the arithmetic mean (adding all items in a set together and then dividing by the total number of items in the set) and the geometric mean (multiplying all items in a set together and taking the nth root of the result where n is the number of items in the set).

For some set S with n items in it defined as: $S = \{S_1, S_2, S_3, S_4, \dots S_n\}$ the average of S is given by

$$\mu = \frac{S_1 + S_2 + S_3 + \cdots + S_{n-1} + S_n}{n}$$

where μ (mu) denotes the mean value of the set.

One fact worth noting about the mean is that the mean of a set will always be between the least and greatest values in the set.

$$Let\ S = \{S_1, S_2, S_3, S_4, \dots S_n\}$$

Where S_1 is the least value in S and S_n is the greatest value in S_n.

If all items in the set $= S_1$, then $\sum S = n\, S_1$ (where $\sum S$ means the sum of all items in S). In this case $\mu = \frac{n\, S_1}{n} = S_1$. Otherwise $\sum S > n\, S_1$ meaning $\mu > \frac{n\, S_1}{n} = S_1$. All of this means that $\mu \geq S_1$.

If all values in the set $= S_n$ then $\sum S = n\, S_n$ so $\mu = \frac{n\, S_n}{n} = S_n$. Otherwise $\sum S < n\, S_n$ so $\mu < \frac{n\, S_n}{n} = S_n$. This means that if S_n is the greatest value in some set $\mu \leq S_n$.

This all means that if S_1 is the least value in some set S and S_n is the greatest value in S and μ is the mean value of the set, then $S_1 \leq \mu \leq S_n$.

The concept of the arithmetic mean can be applied to a continuous curve as well. Finding the mean value of a continuous curve on a given interval can be done using calculus. Specifically, by using the rectangles from before. Instead of adding their areas ($f(x)dx$) together, add the heights ($f(x)$) and divide the result by the number of rectangles. If each rectangle has width h, then

$$number\ of\ rectangles = \frac{b-a}{h}$$

$$sum\ of\ heights = \sum_{n=0}^{\frac{b-a}{h}-1} f(a+nh)$$

$$\mu \approx \sum_{n=0}^{\frac{b-a}{h}-1} f(a+nh) \div \frac{b-a}{h}$$

$$\mu \approx \sum_{n=0}^{\frac{b-a}{h}-1} f(a+nh) \times \frac{h}{b-a}$$

$$\mu \approx h \sum_{n=0}^{\frac{b-a}{h}-1} f(a+nh) \div (b-a)$$

To find the exact mean, let $h \to 0$ or let h become dx:

$$\mu = dx \sum_{n=0}^{\frac{b-a}{dx}-1} f(a+n\,dx) \div (b-a)$$

$$Since\ dx \sum_{n=0}^{\frac{b-a}{dx}-1} f(a + n\ dx) = \int_a^b f(x)dx$$

$$\mu = \frac{\int_a^b f(x)dx}{b-a}$$

The mean value of the curve $y = f(x)$ on some interval $[a, b]$ is given by

$$\mu = \frac{\int_a^b f(x)dx}{b-a}$$

Since $S_{min} \leq \mu \leq S_{max}$, if the curve of $y = f(x)$ is continuous, every value between the maximum and minimum values of $f(x)$ (S_{min} and s_{max}) on the interval $[a, b]$ must have a point on the curve somewhere on the interval. Which means there must be some x-value which I will call c which lies on the interval and for which $f(c) = \mu$.

$$so\ f(c) = \frac{\int_a^b f(x)dx}{b-a}$$

$$\int_a^b f(x)dx = (b-a) \times f(c)$$

For some function $f(x)$ which is continuous on the interval $[a, b]$, there exists some number c such that:

$$a \leq c \leq b$$

$$\& \int_a^b f(x)dx = (b-a) \times f(c)$$

This is called the Mean Value Theorem.

I have so far discussed differentiation and integration separately, but is it possible that there could be a link between them? I may seem at first that the answer would be no, how is gradient linked to area? To answer this question, I will ask two more. What really *is* finding the gradient of a curve? It is dividing on variable by another. On the other hand, finding the area under curve, integration is summing an infinite number of terms together. Each of those terms is one variable multiplied by another. If differentiation is simply glorified division and integration is nothing but over the top multiplication, and division is the inverse of differentiation, it reasons that, perhaps, differentiation is the inverse of integration. This paragraph exists to provide the motivation for what I will soon do.

Differentiating turns a function into another function (the function which gives the gradient of the original at a given x value). If integration is the inverse of this, it must also turn an input function into some output function. It does not do this. It takes an input function and two numbers and gives the area between these numbers under the curve. To resolve this I will say that, if my input function is

$f(x)$ then the output should be a function which gives the area between some arbitrary point a and some value of x. This means that the output will be a function of x, the fact the a is some arbitrary point, the resulting function will produce an arbitrary result, this doesn't matter and will make more sense later. I will call the output $F(x)$.

$$F(x) = \int_a^x f(t)dt$$

I am using t here as not to confuse the variable of the horizontal axis with the variable which the function F should be in terms of.

If integration is indeed the inverse of differentiation, then this function $F(x)$ should differentiate to $f(x)$. I will test this using the difference quotient.

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

$$F'(x) = \lim_{h \to 0} \frac{\int_a^{x+h} f(t)dt - \int_a^x f(t)dt}{h}$$

$$F'(x) = \lim_{h \to 0} \frac{\int_x^{x+h} f(t)dt}{h}$$

Via the mean value theorem, there exists some number c such that:

$$x \leq c \leq x + h$$

$$\int_x^{x+h} f(t)dt = (x + h - x) \times f(c)$$

$$\int_x^{x+h} f(t)dt = h \times f(c)$$

$$F'(x) = \lim_{h \to 0} \frac{h \times f(c)}{h}$$

$$F'(x) = \lim_{h \to 0} f(c)$$

As $h \to 0$ the interval $[x, x+h]$ converges to just a single value x. Which means as $h \to 0$ the interval on which c can exists becomes just a single point x or as $h \to 0$, $c \to x$.

$$so\ F'(x) = f(x)\ as\ required$$

This means that doing differentiation in reverse (taking the antiderivative) of $f(x)$ yields:

$$\int_a^x f(x)dx$$

There are still two problems though. Firstly, this function is arbitrary because a is some arbitrary number. The other problem is that the antiderivative of $f(x)$ could be any function. E.g., the antiderivative of $\cos x$ is $\sin x$ since $\frac{d}{dx}(\sin x) = \cos x$. But $\frac{d}{dx}(\sin(x) + 3) = \cos(x)$ as well. $\frac{d}{dx}(\sin(x) - 17) = \cos(x)$. In general, $\frac{d}{dx}(\sin(x) + C) = \cos(x)$ where C could be any constant.

Both of these problems cancel each other out, because if the antiderivative gives us an arbitrary function, we would also expect that integration (which we now know is the same thing) would as well.

Now for the part which makes all of this actually useful:

$$Let\ F(x) = \int_a^x f(t)dt$$

$$F(q) - F(p) = \int_a^q f(t)dt - \int_a^p f(t)dt$$

$$F(q) - F(p) = \int_p^q f(t)dt$$

Another note on notation:

$$\int_a^x f(t)dt\ \ is\ usually\ written\ as\ \int f(x)dx$$

This is called an indefinite integral, also known as an antiderivative or a primitive.

$$If\ F(x) = \int f(x)dx$$

$$Then\ \int_a^b f(x)dx = F(b) - F(a)$$

This is called the ***Fundamental Theorem of Calculus***.

This means that:

$$\int_a^b f(x)\,dx = F(b) - F(a) = -\big(F(a) - F(b)\big) = -\int_b^a f(x)\,dx$$

Examples:

$$\int \sin x\,dx = -\cos x + C$$

because $\frac{d}{dx}(-\cos x + C) = \sin x$

The area under a $sine$ wave from 0 to π is given by:

$$Area = \int_0^\pi \sin x \, dx$$

$$= [-\cos(\pi) + C] - [-\cos 0 + C]$$

$$= -\cos\pi + \cos 0 + C - C$$

$$= \cos 0 - \cos\pi$$

$$= 1 - (-1)$$

$$= 1 + 1$$

$$= 2$$

Notice how the C cancels out when finding definite integrals, meaning that when dealing with definite integrals, the $+C$ can be ignored.

Another example,

$$\int e^t dt = e^t + C$$

because $\frac{d}{dt}(e^t + C) = e^t$

The area bound by the t-axis, the curve $y = e^t$ and to the left of the line $t = x$ is given by:

$$Area = \int_{-\infty}^{x} e^t dt$$

The area has no lower bound for the t axis meaning we want the area from $-\infty$ to x.

$$Area = [e^t]_{-\infty}^{x}$$

This notation allows the $+C$ to be ignored.

$$Area = \lim_{n\to\infty}\left((e^x) - (e^{-n})\right)$$

$$Area = e^x - \lim_{n\to\infty}\frac{1}{e^n}$$

As $n \to \infty$, $e^n \to \infty$ and so $\frac{1}{e^n} \to 0$.

$$Area = e^x - 0$$

$$Area = e^x$$

The area under the curve $y = e^x$ from $-\infty$ to some x coordinate equals the height of the function at that x-coordinate as well as the gradient of the curve at that x-coordinate.

One final example.

When differentiating a polynomial, you multiply by the exponent and then subtract 1 from the power. Integrating is therefore this, but in reverse, add 1 to the exponent and then divide by the new exponent. This is done for each term.

$$\int (x^2 - 2)\, dx = \int (x^2 - 2x^0)\, dx = \frac{1}{3}x^3 - 2x + C$$

The area under the curve from 0 to 1 is given by:

$$Area = \int_0^1 (x^2 - 2x)\, dx$$

$$Area = \left[\frac{1}{3}x^3 - 2x\right]_0^1$$

$$Area = \left(\frac{1}{3}(1)^3 - 2(1)\right) - \left(\frac{1}{3}(0)^3 - 2(0)\right)$$

$$Area = \left(\frac{1}{3} - 2\right) - (0)$$

$$Area = -\frac{5}{3}$$

Here the area is negative. This is because the curve here goes below the x-axis. Negative and positive areas in the same definite integral will cancel each other out, for example $\int_0^{2\pi} \sin x\, dx = [-\cos x]_0^{2\pi} = (-\cos 2\pi) - (-\cos 0) = (-1) - (-1) = 0$

The rest of this chapter will cover different techniques for finding antiderivatives/indefinite integrals/primitives.

The first method is called integration by inspection, and simply means to ask the question, do I already know of a function which differentiates to this? For example, integrate $-\frac{1}{\sqrt{1-x^2}}$.

It may not at first seem obvious, but you need only look back to find that:

$$\arccos' x = \frac{-1}{\sqrt{1-x^2}}$$

$$so \int \frac{-1}{\sqrt{1-x^2}} dx = \arccos x + C$$

$$Also \frac{d}{dx}(\arcsin x) = \frac{1}{\sqrt{1-x^2}}$$

$$-\frac{d}{dx}(\arcsin x) = \frac{-1}{\sqrt{1-x^2}}$$

$$\frac{d}{dx}(-\arcsin x) = \frac{-1}{\sqrt{1-x^2}}$$

$$\int \frac{-1}{\sqrt{1-x^2}} dx = -\arcsin x + D$$

I have used D for the second of these integrals to demonstrate that this result does not necessarily mean that $\arccos x = -\arcsin x$ because the values of C and D may or may not be different. This result does however demonstrate that each of these functions is a vertical translation of the other.

Something worth noting now:

$$\frac{d}{dx}\left(\int f(x)dx + \int g(x)dx\right) = f(x) + g(x)$$

$$so \int \big(f(x) + g(x)\big)dx = \int f(x)dx + \int g(x)dx$$

$$\& \ \frac{d}{dx}\left(k \int f(x)dx\right) = k\, f(x)$$

$$\int k\, f(x)\, dx = k \int f(x)\, dx$$

$$Also \frac{d}{dx}\left(\frac{1}{a} f(ax)\right) = \frac{1}{a}\frac{d}{dx}\big(f(ax)\big) = \frac{1}{a}\ a\, f'(ax) = f'(ax)$$

$$So \int f'(ax)\, dx = \frac{1}{a} f(ax)$$

$$e.g., \int e^{2x}\, dx = \frac{1}{2} e^{2x} + cC$$

The next method of integration by substitution. This method (sometimes called u-substitution) involves changing an integral from being in terms of one variable (such as x) to being in terms of another (such as u) making it easier to solve. The easiest way to demonstrate this method is via examples.

E.g.,

$$Find \int x\sqrt{2x + 5}$$

The first step is deciding what to substitute, it is usually a good idea to substitute the function which is inside another function, in this case $u = 2x + 5$ seems to be the most obvious substitution to make.

$$Let\ u = 2x + 5$$

Here us is a dummy variable as it wasn't in the original question, nor will it be in the final answer.

Substitute this back into the original.

$$\int x\sqrt{u}\, dx$$

$$= \int x\, u^{\frac{1}{2}}\, dx$$

I still need to deal with the x and the dx. For x, rearrange $u = 2x + 5$ into $x = \frac{u-5}{2}$ and substitute this back in.

$$\int \frac{u^{\frac{1}{2}}(u-5)}{2}\ dx$$

$$= \frac{1}{2}\int u^{\frac{1}{2}}(u-5)\ dx$$

$$= \frac{1}{2}\int u^{\frac{3}{2}} - 5u^{\frac{1}{2}}\ dx$$

To deal with the dx, differentiate both sides of $u = 2x + 5$.

$$du = 2\ dx$$

$$dx = \frac{du}{2}$$

Substitute this back in.

$$\frac{1}{2}\int u^{\frac{3}{2}} - 5u^{\frac{1}{2}}\frac{du}{2}$$

$$= \frac{1}{4}\int u^{\frac{3}{2}} - 5u^{\frac{1}{2}}\ du$$

Now I have a polynomial which I can integrate.

$$= \frac{1}{4}\left[\frac{2}{5}u^{\frac{5}{2}} - \frac{10}{3}u^{\frac{3}{2}}\right]$$

$$= \frac{1}{10}u^{\frac{5}{2}} - \frac{5}{6}u^{\frac{3}{2}}$$

The original integral was in terms of x and so the answer must also be in terms of x. Substitute back in $u = 2x + 5$:

$$\int x\sqrt{2x+5} = \frac{(2x+5)^{\frac{5}{2}}}{10} - \frac{5(2x+5)^{\frac{3}{2}}}{6} + C$$

It is only essential that you put the $+C$ on at the end, but it is very important that you do not forget it.

Another example:

$$Find\ \int \frac{e^{\arctan x}}{1+x^2}$$

$$Let\ u = \arctan x$$

$$x = \tan u$$

$$dx = \sec^2 u\ \ du$$

Substitute these back in.

$$\int \frac{e^u}{1+\tan^2 u} \sec^2 u \ du$$

$$1+\tan^2 u = \sec^2 u$$

$$so\ we\ have \int \frac{e^u}{\sec^2 u} \sec^2 u \ du$$

$$= \int e^u \ du$$

$$= e^u$$

$$= e^{\arctan x} + C$$

You can also check that you indefinite integral is correct by differentiating it and seeing if you get back to the original.

$$\frac{d}{dx}(e^{\arctan x})$$

$$= e^{\arctan x} \times \frac{d}{dx}(\arctan x)$$

$$= e^{\arctan x} \times \frac{1}{1+x^2}$$

$$= \frac{e^{\arctan x}}{1+x^2}$$

Which is what we started with, so the answer was correct.

The next method of integration I will discuss is called integration by parts. A method which allows the product of functions (e.g., $f(x) \times g(x)$) to be integrated.

To differentiate a product of functions we can use:

$$\frac{d}{dx}(uv) = u'v + uv'$$

where u and v are functions of x and u' and v' are the derivatives of u and v respectively.

Integrating both sides:

$$uv = \int u'v \, dx + \int uv' \ dx$$

$$\int uv' \ dx = uv - \int u'v \, dx$$

For example,

$$find \int x \cos x \ dx$$

You first need to decide which to differentiate and which to integrate (out of x and $\cos x$.) In this case, differentiating x gives 1 which will simplifying things significantly.

$$Let\ u = x$$

$$Let\ v' = \cos x$$

Differentiating u:

$$u' = 1$$

Integrating v':

$$v = \sin x$$

$$\int x \cos x \ dx$$

$$= uv - \int u'v \, dx$$

$$= x \sin x - \int 1 \sin x \ dx$$

$$= x \sin x - (-\cos x)$$

$$= x \sin x + \cos x + C$$

This method can also be used to integrate $\ln x$ by writing it as $1 \times \ln x$

$$Let\ v' = 1$$

$$Let\ u = \ln x$$

$$v = x$$

$$u' = \frac{1}{x}$$

$$\int uv' \ dx = uv - \int u'v \, dx$$

$$\int 1 \ln x \, dx = x \ln x - \int \frac{1}{x} x \, dx$$

$$\int \ln x \, dx = x \ln x - \int 1 \, dx$$

$$\int \ln x \ dx = x \ln x - x + C$$

Another example of integration by parts:

$$\int \sin(x) \cos(x) \, dx$$

$$Let\ u = \sin x$$

$$Let\ v' = \cos x$$

$$u' = \cos x$$

$$v = \sin x$$

$$\int \sin(x)\cos(x)\,dx = \sin(x)\cos(x) - \int \cos(x)\sin(x)\,dx$$

In this case we have ended up with what we started with. By adding $\int \cos(x)\sin(x)\,dx$ to both sides:

$$2\int \cos(x)\sin(x)\,dx = \sin(x)\cos(x)$$

$$\int \cos(x)\sin(x)\,dx = \frac{\sin(x)\cos(x)}{2} + C$$

Don't forget the $+C$!

I will next show how integration by parts can be written a different way. (Notation note: $f^{(-1)}(x)$ means the (-1)th derivative of $f(x)$ or its indefinite integral. $f^{(-n)}(x)$ means the nth indefinite integral of $f(x)$).

Setting out the variables in this table:

$$u = f(x) \quad v' = g(x)$$

$$u' = f^{(1)}(x) \quad v = g^{(-1)}(x)$$

Using the integration by parts formula:

$$\int f(x)g(x)\,dx = f(x)g^{(-1)}(x) - \int f^{(1)}(x)g^{(-1)}(x)(x)\,dx$$

The top left value in the table has been differentiated to get the bottom left value. The top right value has been integrated to get the bottom left value. The top left value has been multiplied by the bottom left value, subtracted from this is the integral of the product of the bottom row of the table.

If $\int f^{(1)}(x)g^{(-1)}(x)(x)\,dx$ is known, or is equal to the original integral, then you can stop here. If not, you will need to use integration by parts again.

$$u' = f^{(1)}(x) \quad v = g^{(-1)}(x)$$

$$u'' = f^{(2)}(x) \quad V = g^{(-2)}(x)$$

Where V is the integral of v.

Again, to create this table I differentiated the top left to get the bottom left and then integrated the top right to get the bottom right. Using the table as described above, the integral of the product of the top row equals the top left item multiplied by the bottom right item minus the integral of the product of the bottom row, or:

$$\int f^{(1)}(x)g^{(-1)}(x)dx = f^{(1)}(x)g^{(-2)}(x) - \int f^{(2)}(x)g^{(-2)}(x)dx$$

Substituting this into what we had before:

$$\int f(x)g(x)\,dx = f(x)g^{(-1)}(x) - \left(f^{(1)}(x)g^{(-2)}(x) - \int f^{(2)}(x)g^{(-2)}(x)dx\right)$$

$$\int f(x)g(x)\,dx = f(x)g^{(-1)}(x) - f^{(1)}(x)g^{(-2)}(x) + \int f^{(2)}(x)g^{(-2)}(x)dx$$

This process can be continued until either you reach a point at which the integral on the right is known or is the same as the original (ignoring constants as they can be taken out of the integral) and so an equation can be formed to find the integral. Writing all of us and vs becomes unnecessary when we use the table so I will write it instead like this:

$$\begin{array}{cc} D & I \\ f(x) & g(x) \\ f^{(1)}(x) & g^{(-1)}(x) \\ f^{(2)}(x) & g^{(-2)}(x) \end{array}$$

and so on.

Notice that the terms alternate between being added and subtracted. To implement this into the table I will add a new column for the sign.

$$\begin{array}{ccc} S & D & I \\ + & f(x) & g(x) \\ - & f^{(1)}(x) & g^{(-1)}(x) \\ + & f^{(2)}(x) & g^{(-2)}(x) \end{array}$$

and so on.

Though it doesn't line up correctly here you get the idea. Also, the S column is usually not labelled. To use this table once created, consider the left and middle columns as being the same column, i.e., the D column consists of $f(x), -f^{(1)}(x), f^{(2)}(x), \ldots$

Multiply the top left item by the one down and to the right of it. Write this down. Check if the product of the second row is either integrable (able to be integrated) or is the same ass the original. If either of these are the case then stop and add the integral of this row to what you have already written down. If not, then add the product of the second item in the D column the third item in the I column to what you have written down already. Check the integral of the third row. Repeat this process.

Example:

$$\int x^2 e^{3x} dx$$

I will differentiate x^2 as it is polynomial and so will become a constant if I differentiate it enough times.

$$\begin{array}{ccc} & D & I \\ + & x^2 & e^{3x} \\ - & 2x & \frac{1}{3}e^{3x} \end{array}$$

$$+ \quad 2 \quad \frac{1}{9}e^{3x}$$

Here I can stop because I can integrate the bottom row.

The first item of D column multiplied by the second item of the I column is $\frac{1}{3}x^2e^{3x}$. The product of the second item of the D column multiplied by the third item of the I column is $-\frac{2}{9}xe^{3x}$. The product of the bottom row is $\frac{2}{9}e^{3x}$.

$$\int x^2e^{3x}\,dx = \frac{1}{3}x^2e^{3x} - \frac{2}{9}xe^{3x} + \int \frac{2}{9}e^{3x}\,dx$$

$$\int \frac{2}{9}e^{3x}\,dx = \frac{2}{9}\times\frac{1}{3}e^{3x} = \frac{2}{27}e^{3x}$$

$$so \int x^2e^{3x}\,dx = \frac{1}{3}x^2e^{3x} - \frac{2}{9}xe^{3x} + \frac{2}{27}e^{3x}$$

Sometimes integration by parts can give seemingly nonsensical results, for example:

$$\int \tan x\,dx$$

$$= \int \frac{\sin x}{\cos x}dx$$

$$= \int \sin x \sec x\,dx$$

$$\begin{array}{ccc} & D & I \\ + & \sec x & \sin x \\ - & \tan x \sec x & -\cos x \end{array}$$

$$\int \tan x\,dx = \sec x \times -\cos x - \int \tan x \sec x \times -\cos x\,dx$$

$$\int \tan x\,dx = -\frac{\cos x}{\cos x} + \int \frac{\tan x}{\cos x} \times \cos x\,dx$$

$$\int \tan x\,dx = -1 + \int \tan x\,dx$$

Subtracting $\int \tan x\,dx$ from both sides:

$$0 = -1$$

This is clearly a nonsensical result. So, what has happened here? What happened is I forgot the $+C$. If I added a $+C$ on the right and a $+D$ on the right, I get:

$$\int \tan x\,dx + D = -1 + \int \tan x\,dx + C$$

$$D - C = -1$$

Which is a perfectly reasonable result because D and C are just arbitrary constants. If you were wondering, this is not an issue when adding the integral to both sides, for example:

$$\int \sin(x)\cos(x)\,dx = \sin(x)\cos(x) - \int \cos(x)\sin(x)\,dx$$

If I write the $+C$ and $+K$ here:

$$\int \sin(x)\cos(x)\,dx + D = \sin(x)\cos(x) - \left(\int \cos(x)\sin(x)\,dx + K\right)$$

$$2\int \sin(x)\cos(x)\,dx + D + K = \sin(x)\cos(x)$$

$D + K$ is just the sum of two arbitrary constants and so $D + K$ is also some arbitrary constant so we can write it as $+B$ for example.

$$2\int \sin(x)\cos(x)\,dx + B = \sin(x)\cos(x)$$

$$\int \sin(x)\cos(x)\,dx + \frac{B}{2} = \frac{\sin(x)\cos(x)}{2}$$

$\frac{B}{2}$ is also just some arbitrary constant and so could be written as $+C$.

$$\int \sin(x)\cos(x)\,dx + C = \frac{\sin(x)\cos(x)}{2}$$

The constant is already assumed to be there by the indefinite integral notation anyway so:

$$\int \sin(x)\cos(x)\,dx = \frac{\sin(x)\cos(x)}{2}$$

So having the product of the row when using the $D\ I$ method (the name of the method using the table) being the same as the original integral does not always work, but sometimes it does. By the way, $\tan x$ can be integrated properly by using a u-substitution.

$$I = \int \tan x\,dx$$

$$I = \int \frac{\sin x}{\cos x}dx$$

$$Let\ u = \cos x$$

$$du = -\sin x\,dx$$

$$dx = \frac{du}{-\sin x}$$

$$I = \int \frac{\sin x}{u}\frac{du}{-\sin x}$$

$$I = \int -\frac{1}{u}\,du$$

$$= -\ln(u)$$

Because $\ln'(x) = \frac{1}{x}$

$$= -\ln(\cos(x))$$

$$= \ln((\cos(x))^{-1})$$

$$= \ln\left(\frac{1}{\cos(x)}\right)$$

$$= \ln(\sec(x)) + C$$

This integral brings up another key point, the integral of the reciprocal function.

$$\int \frac{1}{x}\, dx = \ln(x) + C$$

$$because \frac{d}{dx}(\ln(x) + C) = \frac{1}{x}$$

But the reciprocal function is defined for all x (except 0) whereas the natural logarithm is defined only for positive inputs. How then, could I find the area under the curve $y = \frac{1}{x}$ where $x < 0$? To answer this question, it is essential to realise that $\frac{1}{-x} = -\frac{1}{x}$ which means that if you plot the graph $y = \frac{1}{x}$ for only positive x, the negative side of the graph will be the same but reflected in the x-axis (and obviously also flipped in the y-axis). The area under the curve between two negative values (which will be negative because at this interval the curve is below the x-axis) will be the same as the area between the corresponding positive values.

$$e.g., \int_{-5}^{-3} \frac{1}{x}\, dx = -\int_{3}^{5} \frac{1}{x}\, dx = [\ln x]_3^5$$

If a and b are both positive:

$$\int_{-a}^{-b} \frac{1}{x}\, dx = -\int_{b}^{a} \frac{1}{x}\, dx = \int_{a}^{b} \frac{1}{x}\, dx = [\ln x]_a^b$$

$$\& \int_{a}^{b} \frac{1}{x}\, dx = [\ln x]_a^b$$

This means that regardless of whether a and b are both positive or are both negative, we can pretend that they are both positive and use that instead. The modulus (or absolute value) function can be used here. $|x|$ (modulus of x or absolute value of x) in this context means: if x is positive, keep it the same, if x is negative, make it positive.

Since this is what we are doing when taking the natural logarithm of a and b, we can say that:

$$\int \frac{1}{x}\, dx = \ln|x| + C$$

Whilst this isn't technically correct, for all intents and purposes it is.

If one a is negative and b is positive we can find

$$\int_a^b \frac{1}{x}\,dx$$

By finding the area from a to 0 and then from 0 to b because the fundamental theorem of calculus should only be used when $f(x)$ is continuous on the interval $[a,b]$.

$$\int_a^b \frac{1}{x}\,dx$$

$$= \int_a^0 \frac{1}{x}\,dx + \int_0^b \frac{1}{x}\,dx$$

$$= [\ln|x|]_a^0 + [\ln|x|]_0^b$$

$$= \ln|0| - \ln|a| + \ln|b| - \ln|0|$$

$$= \ln|b| - \ln|a|$$

$$= [\ln|x|]_a^b$$

$$= \int_a^b \frac{1}{x}\,dx$$

which means that, it so happens that defining $\int \frac{1}{x}\,dx$ as $\ln|x| + C$ allows you to integrate over a non-continuous interval with no problem. It is very important to note that this should not be done in general, it just so happens to work in this case.

For example, the function $f(x) = \frac{1}{x^2}$ is not defined for $x = 0$ and so is not continuous over any interval containing 0. The curve is always above the x-axis since x^2 and hence $\frac{1}{x^2}$ is always > 0. This means any definite integral of this function from a to b should be positive as long as $b > a$.

$$\int_{-2}^{1} \frac{1}{x^2}\,dx$$

Using the fundamental theorem of calculus:

$$\int_{-2}^{1} x^{-2}\,dx$$

$$= [-x^{-1}]_{-2}^{1}$$

$$= \left[-\frac{1}{x}\right]_{-2}^{1}$$

$$= \left(-\frac{1}{1}\right) - \left(-\frac{1}{-2}\right)$$

$$= (-1) - \left(\frac{1}{2}\right)$$

$$= -\frac{3}{2}$$

Which is clearly not the correct answer because the answer should have been positive. For this reason, it is essential that you only use the fundamental theorem of calculus over a continuous interval of $f(x)$ (except with the special case of $\int \frac{1}{x}\,dx$). To find the integral properly:

$$I = \int_{-2}^{1} x^{-2}\,dx$$

$$I = \int_{-2}^{0^-} x^{-2}\,dx + \int_{0^+}^{1} x^{-2}\,dx$$

Since $-\frac{1}{x}$ is not defined for $x = 0$ we must use limits. 0^- means approaching 0 from the left. This can be thought of as -0.000001 for example. Likewise, 0^+ means approaching 0 from the right and can be though of as 0.000001 for example.

$$I = \left[-\frac{1}{x}\right]_{-2}^{0^-} + \left[-\frac{1}{x}\right]_{0^+}^{1}$$

$$I = \left[\left(-\frac{1}{0^-}\right) - \left(-\frac{1}{-2}\right)\right] + \left[\left(-\frac{1}{1}\right) - \left(-\frac{1}{0^+}\right)\right]$$

$$I = \left[\frac{-1}{0^-} - \frac{1}{2}\right] + \left[-1 + \frac{1}{0^+}\right]$$

Since 0^- is essentially a negative number which is very close to 0, as $h \to 0^-$, $\frac{-1}{h} \to +\infty$. Likewise, as $h \to 0^+$, $\frac{1}{h} \to \infty$.

$$I = \left(\infty - \frac{1}{2}\right) + (\infty - 1)$$

$$I = 2\infty - \frac{3}{2}$$

$$I = \infty$$

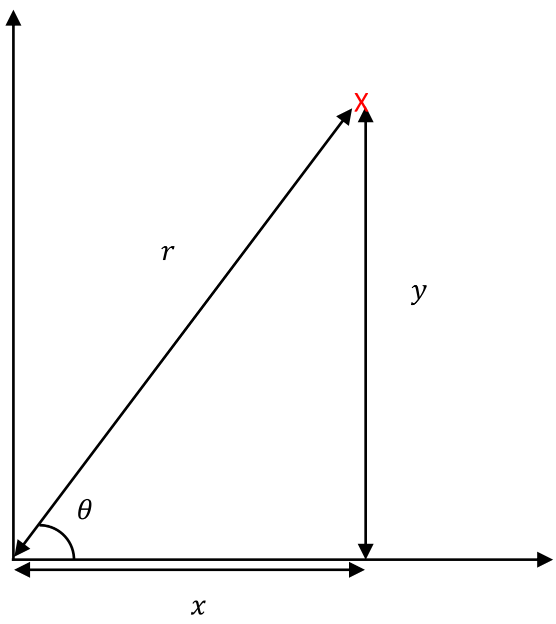
In other words, the integral diverges, because it does not have a finite value. The area under the curve on this interval is infinite. Also, when doing the integration this way, it is positive which is what was expected.

Polar Coordinates

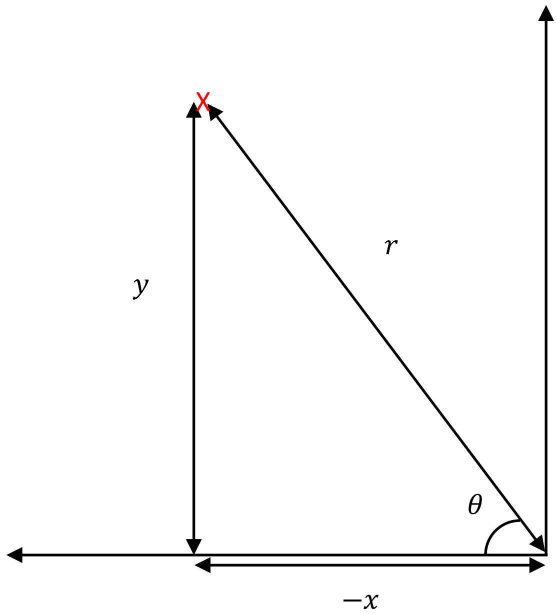
We are so far familiar with representing points on a grid using the “cartesian” or “rectangular” coordinate system, in which two numbers (x, y) can represent a point. The x coordinate describes how far to the right do move from the origin and the y coordinate describes how far up to move from that point to reach the specified point. There are, however, other ways to represent points on a grid, such as the polar coordinate system. A point can be described using two numbers (r, θ) where the θ coordinate describes which angles (anti-clockwise about the origin starting facing the positive directly to the right) to turn and the r coordinate then describes how far to move in this direction to reach the specified point.

Plotting the graph $y = 3$ would mean drawing a line such that every point on that line has a y coordinate of 3. That is to say, a horizontal line at $y = 3$. Plotting $r = 3$ means to draw a curve such that the distance from any given point to the origin is equal to three. In other words, a circle, radius 3 and centre $(0,0)$. Plotting $\theta = 3$ would be to draw a line going through the origin ($(0,0)$) such that the angle made between this line and the positive x axis $= 3$ radians. Because the value of r can be negative, this line would extend infinitely in both directions.

Converting between cartesian and polar coordinates can be done using trigonometry. Here is a diagram showing a point in the top right quadrant of a coordinate grid.



The point marked X has cartesian coordinates (x, y) and polar coordinates (r, θ). Using trigonometry, $\sin\theta = \frac{y}{r}$ so $y = r\sin\theta$ and $\cos\theta = \frac{x}{r}$ so $x = r\cos\theta$. This means that for a point in the top right quadrant, if you know r and you know θ then you can write its coordinates in cartesian form as $(r\cos\theta\,, r\sin\theta)$.



Here I have a similar diagram for the top left quadrant. The point still has coordinates (x, y), but since x is negative, and the horizontal distance is positive, the distance is $-x$. The vertical distance is still y. Using trigonometry here: $\sin\theta_1 = \frac{y}{r}$ so $y = r\sin\theta_1$. $\cos\theta_1 = \frac{-x}{r}$ so $x = -r\cos\theta_1$. When using polar coordinates, θ should represent the anti-clockwise angle starting from facing the positive x-axis. The diagram instead has labelled θ_1, the clockwise angle from the negative x-axis. In radians, $\pi - \theta_1 = \theta$ so $\theta_1 = \pi - \theta$. Substituting this into what we had before:

$$y = r\sin(\pi - \theta)$$

using the addition formula for $sine$:

$$y = r(\sin\pi\cos\theta - \sin\theta\cos\pi)$$

$$y = r\big(0\cos\theta - \sin\theta \times (-1)\big)$$

$$y = r\sin\theta$$

$$x = -r\cos(\pi - \theta)$$

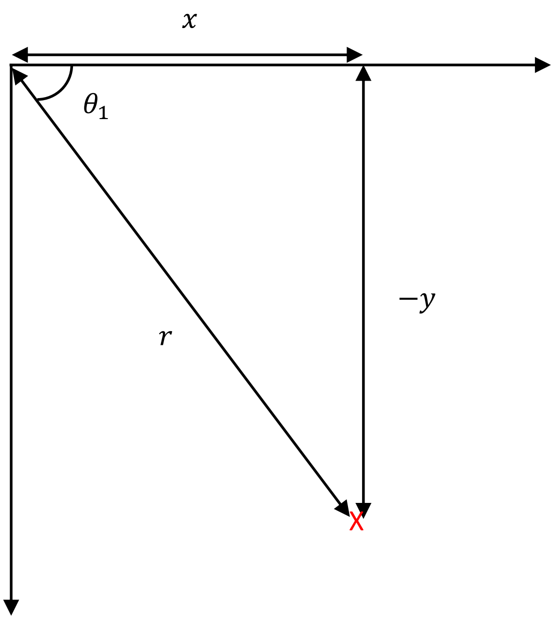
using the addition formula for $cosine$:

$$x = -r(\cos\pi\cos\theta + \sin\pi\sin\theta)$$

$$x = -r(-1\cos\theta + 0\sin\theta)$$

$$x = r\cos\theta$$

For the bottom right quadrant:



The point here has coordinates (x, y) where x is positive and y is negative. Since all lengths must be positive, the horizontal length is x and the vertical length is $-y$.

Using trigonometry: $\sin\theta_1 = \frac{-y}{r}$, $y = -r\sin\theta_1$. $\cos\theta_1 = \frac{x}{r}$, $x = r\cos\theta_1$. Here, θ_1 is the angle clockwise from the positive x-axis meaning that θ (the anti-clockwise angle from the positive x-axis) $= 2\pi - \theta_1$ so $\theta_1 = 2\pi - \theta$.

$$y = -r\sin\theta_1$$

$$y = -r\sin(2\pi - \theta)$$

$$y = -r(\sin 2\pi\cos\theta - \sin\theta\cos 2\pi)$$

$$y = -r\big(0\cos\theta - \sin\theta \times (1)\big)$$

$$y = r\sin\theta$$

$$x = r\cos\theta_1$$

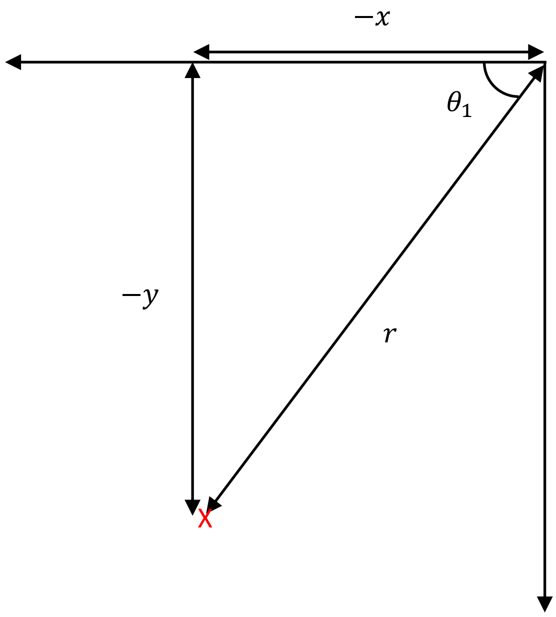
$$x = r\cos(2\pi - \theta)$$

$$x = r(\cos 2\pi\cos\theta + \sin 2\pi\ sin\theta)$$

$$x = r(1\cos\theta + 0\cos\theta)$$

$$x = r\cos\theta$$

For the bottom left quadrant:



Here the point has coordinates (x, y) where x and y are both negative, since lengths are positive, the horizontal length is $-x$ and the vertical length is $-y$.

$$\sin\theta_1 = \frac{-y}{r}$$

$$y = -r\sin\theta_1$$

$$\cos\theta_1 = \frac{-x}{r}$$

$$x = -r\cos\theta_1$$

$$\theta = \pi + \theta_1$$

$$\theta_1 = \theta - \pi$$

$$y = -r\sin(\theta - \pi)$$

$$y = -r(\sin\theta\cos\pi - \sin\pi\cos\theta)$$

$$y = -r(\sin\theta \times (-1) - 0\cos\theta)$$

$$y = r\sin\theta$$

$$x = -r\cos(\theta - \pi)$$

$$x = -r(\cos\theta\cos\pi + \sin\theta\sin\pi)$$

$$x = -r(\cos\theta \times (-1) + sin\theta \times 0)$$

$$x = r\cos\theta$$

So regardless of quadrant, $y = r\sin\theta$ and $x = r\cos\theta$.

For points which are in none of the four quadrants (points on the axis):

For points on the positive x-axis: $\theta = 0, y = 0, x = r$. $y = r\sin\theta$ because $y = 0$ and $r\sin 0 = 0$. $x = r\cos\theta$ because $x = r$ and $r\cos 0 = r$.

For points on the positive y-axis: $\theta = \frac{\pi}{2}, y = r, x = 0$. $y = r\sin\theta$ because $y = r$ and $r\sin\frac{\pi}{2} = r$. $x = r\cos\theta$ because $x = 0$ and $r\cos\frac{\pi}{2} = 0$

For points on the negative x-axis: $\theta = \pi, y = 0, x = -r$. $y = r\sin\theta$ because $y = 0$ and $r\sin\pi = 0$. $x = r\cos\theta$ because $x = -r$ and $r\cos\pi = -r$.

For points on the negative y-axis: $\theta = \frac{3\pi}{2}, y = -r, x = 0$. $y = r\sin\theta$ because $y = -r$ and $r\sin\frac{3\pi}{2} = -r$. $x = r\cos\theta$ because $x = 0$ and $r\cos\frac{3\pi}{2} = 0$.

This means that all coordinates can be converted form cartesian form to polar form by using $y = r\sin\theta$ and $x = r\cos\theta$.

The distance from the origin to a given point can be found by using the Pythagorean theorem so we also have the equation $x^2 + y^2 = r^2$.

Calculus has so far been done on equations of cartesian form but can also be done on equations of polar form. If the segment of a circle has an angle θ, because the angle of a full rotation is 2π, the sector is $\frac{\theta}{2\pi}$ of the full circle. This applies to both arc length and area. This gives us the formulae (which we have already met):

$$arc\ length = \frac{\theta}{2\pi} \times 2\pi r = r\theta$$

$$sector\ area = \frac{\theta}{2\pi} \times \pi r^2 = \frac{1}{2} r^2 \theta$$

To find the arc length or area of some polar curve between to values of θ which I will call α and β, split the curve into a series of thin sectors. As the angle of these sectors approaches 0, the arc length or area approximation gets closer to the true value. Calculus can be used here. If each sector has a radius r and an angle, $d\theta$ then the arc length of each will be $r\ d\theta$ and hence the arc length between α and β is given by:

$$arc\ length = \int_{\alpha}^{\beta} r\ d\theta$$

The area of each sector is $\frac{1}{2} r^2 d\theta$ and so the area of the region is given by:

$$Area = \int_{\alpha}^{\beta} \frac{1}{2} r^2 d\theta$$

$\frac{1}{2}$ is a constant and so can be taken out of the integral:

$$Area = \frac{1}{2}\int_{\alpha}^{\beta} r^2 d\theta$$

Where r is a function of θ.

Complex Numbers

There are certain equations with no solutions, such as $x^2 + 1 = 0$. This is because x^2 is always ≥ 0 and adding 1 to such a number will always result in a number greater than 0. Rearranging the equation, we get $x^2 = -1$ which is impossible because a positive number squared is positive, a negative number squared is positive and $0^2 = 0$. This means that there is no value for x such that $x^2 = -1$. There is a problem with this though. That is a very boring response. Remember, we make the rules in the universe of mathematics, so I will just create a number, call it i and say that $i^2 = -1$. That's it. Problem solved. But remember, we invent axioms and we discover their consequences. So, let's discover some consequences.

Starting with some terminology:

Real numbers are the numbers we are all familiar with, 1, 12, -32, π etc. A real number is any number with no imaginary part.

An imaginary number is any real number mulitplied by i, e.g., $i, 12i, -32i, \pi i$ etc.

A complex number is a number consisting of a real part and an imaginary part. A complex number z, can be written in the form $z = a + bi$ where a and b are real numbers.

i can be treated just like any other constant but with one special property, that being that $i^2 = -1$. To be more specific: $i = \sqrt{-1}$. This is important because $(-i)^2$ also equals -1 because $-i = -\sqrt{-1}$.

We can use the fact that $i^2 = -1$ to obtain:

$$i^0 = 1$$

$$i^1 = i$$

$$i^2 = -1$$

$$i^3 = -i$$

$$i^4 = 1$$

$$i^5 = i$$

and so on.

Complex numbers can be represented on the complex plane, also known as an argand diagram. An argand diagram has a real axis Re and an imaginary axis Im. On this diagram, the complex number $z = a + bi$ can be represented as the point with coordinates (a, b). If we represented this point using polar coordinates, we could represent it as (r, θ). Or we could represent it using cartesian coordinates in terms of r and θ as $(r\cos\theta, r\sin\theta)$. So $(a, b) = (r\cos\theta, r\ \sin\theta)$ so $a = r\cos\theta$ and $b = r\sin\theta$ and so $z = a + bi = r\cos\theta + ri\sin\theta = r(\cos\theta + i\sin\theta)$. This representation $z = r(\sin\theta + i\sin\theta)$ will be very useful later.

What happens if we use complex numbers as inputs to certain functions? One way to find out is by using the Maclaurin Series for these functions.

$$\sin(x) = \frac{x^1}{1!} - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \frac{x^9}{9!} - \cdots$$

$$\sin(ix) = \frac{(ix)^1}{1!} - \frac{(ix)^3}{3!} + \frac{(ix)^5}{5!} - \frac{(ix)^7}{7!} + \frac{(ix)^9}{9!} - \cdots$$

$$\sin(ix) = \frac{i}{1!}x^1 - \frac{-i}{3!}x^3 + \frac{i}{5!}x^5 - \frac{-i}{7!}x^7 + \frac{i}{9!}x^9 - \cdots$$

$$\sin(ix) = \frac{i}{1!}x^1 + \frac{i}{3!}x^3 + \frac{i}{5!}x^5 + \frac{i}{7!}x^7 + \frac{i}{9!}x^9 + \cdots$$

$$\sin(ix) = i\left(\frac{x^1}{1!} + \frac{x^3}{3!} + \frac{x^5}{5!} + \frac{x^7}{7!} + \frac{x^9}{9!} + \cdots\right)$$

$$\cos(x) = \frac{x^0}{0!} - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} - \cdots$$

$$\cos(ix) = \frac{(ix)^0}{0!} - \frac{(ix)^2}{2!} + \frac{(ix)^4}{4!} - \frac{(ix)^6}{6!} + \frac{(ix)^8}{8!} - \cdots$$

$$\cos(ix) = \frac{1}{0!}x^0 - \frac{-1}{2!}x^2 + \frac{1}{4!}x^4 - \frac{-1}{6!}x^6 + \frac{1}{8!}x^8 - \cdots$$

$$\cos(ix) = \frac{x^0}{0!} + \frac{x^2}{2!} + \frac{x^4}{4!} + \frac{x^6}{6!} + \frac{x^8}{8!} + \cdots$$

This one is very interesting:

$$e^x = \frac{x^0}{0!} + \frac{x^1}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \frac{x^6}{6!} + \frac{x^7}{7!} + \frac{x^8}{8!} + \cdots$$

$$e^{ix} = \frac{(ix)^0}{0!} + \frac{(ix)^1}{1!} + \frac{(ix)^2}{2!} + \frac{(ix)^3}{3!} + \frac{(ix)^4}{4!} + \frac{(ix)^5}{5!} + \frac{(ix)^6}{6!} + \frac{(ix)^7}{7!} + \frac{(ix)^8}{8!} + \cdots$$

$$e^{ix} = \frac{1x^0}{0!} + \frac{ix^1}{1!} + \frac{-1x^2}{2!} + \frac{-ix^3}{3!} + \frac{1x^4}{4!} + \frac{ix^5}{5!} + \frac{-1x^6}{6!} + \frac{-ix^7}{7!} + \frac{1x^8}{8!} + \cdots$$

$$e^{ix} = \frac{x^0}{0!} + i\frac{x^1}{1!} - \frac{x^2}{2!} - i\frac{x^3}{3!} + \frac{x^4}{4!} + i\frac{x^5}{5!} - \frac{x^6}{6!} - i\frac{x^7}{7!} + \frac{x^8}{8!} + \cdots$$

Finding the real and imaginary parts separately:

$$e^{ix} = \left(\frac{x^0}{0!} - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} - \cdots\right) + i\left(\frac{x^1}{1!} - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots\right)$$

The real and imaginary parts may seem familiar here. These are the Maclaurin series for $sine$ and $cosine$, so

$$e^{ix} = \cos x + i \sin x$$

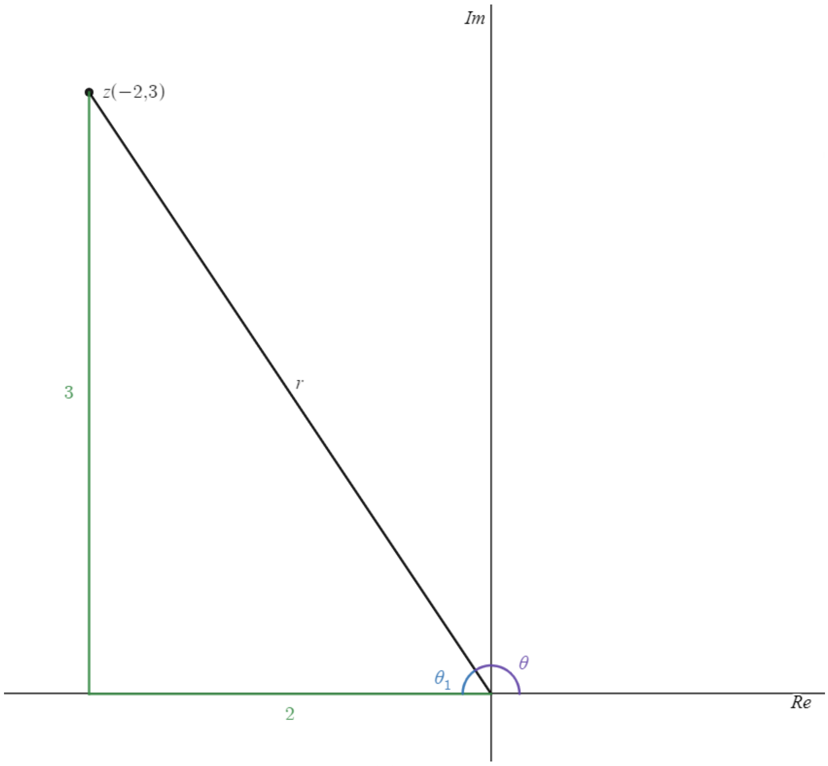
Usually, θ is written instead of x.

$$e^{i\theta} = \cos\theta + i\sin\theta$$

This is called Euler's formula.

This looks very similar to the polar coordinate representation of complex numbers. If a line is draw from some complex number z on an argand diagram, where the length of this line is r and the angle made by rotating counter-clockwise from the positive real axis to this line is θ, then $z = r(\cos\theta + i\sin\theta)$ which we now know equals $re^{i\theta}$. We can use this fact to find the natural logarithms of any complex number by writing them in this form.

For example, to find $\ln(-2+3i)$ draw $-2+3i$ on an argand diagram.



Using trigonometry, $\tan(\theta_1) = \frac{3}{2}$, $\theta_1 = \arctan\left(\frac{3}{2}\right)$. $\theta + \theta_1 = \pi$, $\theta = \pi - \theta_1$, $\theta = \pi - \arctan\left(\frac{3}{2}\right)$. $r^2 = 3^2 + 2^2, r = \sqrt{13}$

$$z = \sqrt{13}e^{i\left(\pi - \arctan\left(\frac{3}{2}\right)\right)}$$

$$\ln(z) = \ln\left(\sqrt{13}\right) + i\left(\pi - \arctan\left(\frac{3}{2}\right)\right)$$

Since $\sin(\theta \pm 2n\pi) = \sin\theta$ and $\cos(\theta \pm 2n\pi) = \cos\theta$ where n is some integer because $sine$ and $cosine$ repeat every 2π radians (360°), and since rotating 2π radians in either direction any number of times will end up in the same place, z could have been written as:

$$z = \sqrt{13}e^{i\left(\pi - \arctan\left(\frac{3}{2}\right) + 2n\pi\right)}$$

where n is some integer. This means that $\ln(z) = \ln(\sqrt{13}) + i\left(\pi(1+2n) - \arctan\left(\frac{3}{2}\right)\right)$ for all integers n. This means that the natural logarithm of complex numbers has an infinite number of values. Because functions should always have one output for each input, the natural logarithm for complex numbers will be defined as the above where $n = 0$.

Using the above, the natural logarithm of some positive real number x may have different possible values (if we ignore the part about $n = 0$). The angle $\theta = 2n\pi$ and $r = x$.

$$x = xe^{i(2n\pi)}$$

$$\ln(x) = \ln(x) + i(2n\pi)$$

This may seem to imply that $i(2n\pi) = 0$ by subtracting $\ln(x)$ from both sides. This is not the case as $\ln(x)$ on the left is not the same as $\ln(x)$ on the right, this is why it is important that we define functions to have only one single output or else confusion like this may occur.

For a complex number $z = -x$ where the imaginary part of z is 0 and the real part is negative, i.e., z is a negative real number, $r = x, \theta = 2\pi + 2n\pi$

$$z = re^{i\theta}$$

$$-x = xe^{i(\pi+2n\pi)}$$

$$\ln(-x) = \ln(x) + i(\pi + 2n\pi)$$

Considering only the base case (where $n = 0$):

$$\ln(-x) = \ln(x) + \pi i$$

If we use this definition of the natural logarithm when integrating, instead of using $\ln|x|$ we get the following:

$$where - a < 0 \; and - b < 0$$

$$\int_{-a}^{-b} \frac{1}{x} dx$$

$$= \ln(-b) - \ln(-a)$$

$$= (\ln(b) + \pi i) - (\ln(a) + \pi i)$$

$$= \ln(b) - \ln(a)$$

$$= \ln|-b| - \ln|-a|$$

We reach the same conclusion either way.

One more example, $\ln(i)$:

$z = 0 + 1i, \theta = \frac{\pi}{2} + 2n\pi, r = 1$

$$z = re^{i\theta}$$

$$i = 1e^{i\left(\frac{\pi}{2}+2n\pi\right)}$$

$$\ln(i) = i\left(\frac{\pi}{2} + 2n\pi\right)$$

Considering only where $n = 0$

$$\ln(i) = \frac{\pi i}{2}$$

I will now explore some other fascinating results:

$$i^i = e^{\ln(i^i)}$$

Because $e^{\ln(x)} = x$

$$i^i = e^{i\ln(i)}$$

$$i^i = e^{i\times i\left(\frac{\pi}{2}+2n\pi\right)}$$

Here I am using all values, not just the base. This is for the same reason that when finding all solutions to $x^2 = b$, you would take the positive and negative roots, $x = -\sqrt{b}$.

$$i^i = e^{-\left(\frac{\pi}{2}+2n\pi\right)}$$

Because $i \times i = -1$

$$i^i = \frac{1}{e^{\frac{\pi}{2}+2n\pi}}$$

So, there are an infinite number of solutions to $x = i^i$. One of these is where $n = 0$:

$$i^i = \frac{1}{\sqrt{e^\pi}}$$

A similar result:

$$\sqrt[i]{i} = i^{\frac{1}{i}}$$

$$\frac{1}{i} = \frac{1}{i} \times \frac{i}{i} = \frac{i}{i^2} = \frac{i}{-1} = -i$$

$$\sqrt[i]{i} = i^{-i}$$

$$\sqrt[i]{i} = \frac{1}{i^i}$$

$$\sqrt[i]{i} = e^{\frac{\pi}{2}+2n\pi}$$

Again, an infinite number of solutions, (where n is an integer). One solution is when $n = 0$:

$$\sqrt[i]{i} = \sqrt{e^\pi}$$

One more result: $\sqrt{i}$. For this, we know that (if the solution exists) it will be a complex number. The real part or imaginary part may be zero. Maybe both or neither will be. In any case, $\sqrt{i} = a + bi$

I now need to find the values of a and b.

$$\sqrt{i} = a + bi$$

$$\left(\sqrt{i}\right)^2 = (a + bi)^2$$

$$i = a^2 + 2abi + b^2 i^2$$

$$i = a^2 + 2abi - b^2$$

Equating real and imaginary parts gives us two equations:

$$0 + 1i = (a^2 - b^2) + (2ab)i$$

$$a^2 - b^2 = 0$$

$$2ab = 1$$

From the second equation:

$$b = \frac{1}{2a}$$

Substituting this into the first equation:

$$a^2 - b^2 = 0$$

$$a^2 - \left(\frac{1}{2a}\right)^2 = 0$$

$$a^2 - \frac{1}{4a^2} = 0$$

$$a^2 = \frac{1}{4a^2}$$

Multiplying both sides by $4a^2$:

$$4a^2(a^2) = 1$$

$$4(a^2)^2 = 1$$

$$\sqrt{4(a^2)^2} = \sqrt{1}$$

$$2a^2 = 1$$

$$a^2 = \frac{1}{2}$$

$$a = \pm\sqrt{\frac{1}{2}}$$

$$a = \pm\frac{1}{\sqrt{2}}$$

$$a = \pm\frac{\sqrt{2}}{2}$$

Substituting this into $b = \frac{1}{2a}$:

$$b = \pm\frac{1}{2 \times \frac{\sqrt{2}}{2}}$$

$$b = \pm\frac{1}{\sqrt{2}}$$

$$b = \pm\frac{\sqrt{2}}{2}$$

$$So\ \sqrt{i} = a + bi = \pm\frac{1}{\sqrt{2}} \pm \frac{\sqrt{2}}{2}i = \pm\frac{\sqrt{2}}{2} \pm \frac{\sqrt{2}}{2}i = \pm\left(\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i\right)$$

Since the square root function is defined to output the positive root:

$$\sqrt{i} = \frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i$$

You can check this result by squaring the right-hand side…

$$\left(\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i\right)^2$$

$$= \left(\frac{\sqrt{2}}{2}\right)^2 + 2\left(\frac{\sqrt{2}}{2}\right)\left(\frac{\sqrt{2}}{2}i\right) + \left(\frac{\sqrt{2}}{2}i\right)^2$$

$$= \frac{2}{4} + 2\frac{2}{4}i - \frac{2}{4}$$

$$= \frac{1}{2} + i - \frac{1}{2}$$

$$= i$$

…and getting the square of the left-hand side.

Revisiting Euler's Formula:

$$e^{i\theta} = \cos\theta + i\sin\theta$$

$$e^{-i\theta} = e^{i(-\theta)} = \cos(-\theta) + i\sin(-\theta) = \cos\theta - i\sin\theta$$

$$So\ e^{-i\theta} = \cos\theta - i\sin\theta$$

using these, the trigonometric functions can be redefined in terms of e:

$$e^{i\theta} + e^{-i\theta} = \cos\theta + i\sin\theta + \cos\theta - i\sin\theta$$

$$e^{i\theta} + e^{-i\theta} = 2\cos\theta$$

$$\cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2}$$

$$e^{i\theta} - e^{-i\theta} = (\cos\theta + i\sin\theta) - (\cos\theta - i\sin\theta)$$

$$e^{i\theta} - e^{-i\theta} = 2i\sin\theta$$

$$\sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i}$$

$$\tan\theta = \frac{\sin\theta}{\cos\theta}$$

$$\tan\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i} \div \frac{e^{i\theta} + e^{-i\theta}}{2}$$

$$\tan\theta = \frac{e^{i\theta} - e^{-i\theta}}{i(e^{i\theta} + e^{-i\theta})}$$

$$\tan\theta = \frac{e^{i\theta} - e^{-i\theta}}{i(e^{i\theta} + e^{-i\theta})} \times \frac{e^{i\theta}}{e^{i\theta}}$$

$$\tan\theta = \frac{e^{2i\theta} - 1}{i(e^{2i\theta} + 1)}$$

$$\cot\theta = \frac{1}{\tan\theta}$$

$$\cot\theta = \frac{i\left(e^{2i\theta} + 1\right)}{e^{2i\theta} - 1}$$

$$\sec\theta = \frac{1}{\cos\theta}$$

$$\sec\theta = \frac{2}{e^{i\theta} + e^{-i\theta}}$$

$$\csc\theta = \frac{1}{\sin\theta}$$

$$\csc\theta = \frac{2i}{e^{i\theta} - e^{-i\theta}}$$

From these functions, I will define a new set of functions, by simply removing the is from these.

$$\sinh x = \frac{e^x - e^{-x}}{2}$$

$$\cosh x = \frac{e^x + e^{-x}}{2}$$

$$\tanh x = \frac{e^{2x} - 1}{e^{2x} + 1}$$

$$\coth x = \frac{e^{2x} + 1}{e^{2x} - 1}$$

$$\operatorname{sech} x = \frac{2}{e^x + e^{-x}}$$

$$\operatorname{csch} x = \frac{2}{e^x - e^{-x}}$$

These are called the hyperbolic trigonometric functions (hence the h), I will discuss these in more detail in the next chapter.

Hyperbolic Trigonometric Functions

The hyperbolic functions were defined in the previous chapter. From these I will derive some identities:

$$\frac{1}{\sinh x} = 1 \div \frac{e^x - e^{-x}}{2} = \frac{2}{e^x - e^{-x}} = \operatorname{csch} x$$

$$\frac{1}{\cosh x} = 1 \div \frac{e^x + e^{-x}}{2} = \frac{2}{e^x + e^{-x}} = \operatorname{sech} x$$

$$\frac{1}{\tan x} = 1 \div \frac{e^{2x} - 1}{e^{2x} + 1} = \frac{e^{2x} + 1}{e^{2x} - 1} = \tanh x$$

$$\frac{\sinh x}{\cosh x} = \frac{e^x - e^{-x}}{2} \div \frac{e^x + e^{-x}}{2} = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{e^{2x} - 1}{e^{2x} + 1} = \tanh x$$

They seem to be the same as the identities as for the regular trigonometric functions.

$$\sinh^2 x + \cosh^2 x$$

$$= \left(\frac{e^x - e^{-x}}{2}\right)^2 + \left(\frac{e^x + e^{-x}}{2}\right)^2$$

$$= \frac{e^{2x} - 2 + e^{-2x}}{4} + \frac{e^{2x} + 2 + e^{-2x}}{4}$$

$$= \frac{2e^{2x} + 2e^{-2x}}{4}$$

$$= \frac{e^{2x} + e^{-2x}}{2}$$

The xs do not cancel out, so this identity is not the same as for the regular trigonometric functions. If I instead subtract though:

$$\sinh^2 x - \cosh^2 x$$

$$= \left(\frac{e^x - e^{-x}}{2}\right)^2 - \left(\frac{e^x + e^{-x}}{2}\right)^2$$

$$= \frac{e^{2x} - 2 + e^{-2x}}{4} - \frac{e^{2x} + 2 + e^{-2x}}{4}$$

$$= \frac{-4}{4}$$

$$= -1$$

$$\sinh^2 x - \cosh^2 x = -1$$

$$\cosh^2 x - \sinh^2 x = 1$$

Investigating the addition formulae assuming they are similar for these new functions:

$$\sinh A \cosh B \pm \sinh B \cosh A$$

$$= \frac{e^A - e^{-A}}{2} \times \frac{e^B + e^{-B}}{2} \pm \frac{e^B - e^{-B}}{2} \times \frac{e^A + e^{-A}}{2}$$

$$= \frac{(e^A - e^{-A})(e^B + e^{-B})}{4} \pm \frac{(e^B - e^{-B})(e^A + e^{-A})}{4}$$

$$= \frac{(e^A - e^{-A})(e^B + e^{-B}) \pm (e^B - e^{-B})(e^A + e^{-A})}{4}$$

$$= \frac{(e^{A+B} + e^{A-B} - e^{-A+B} - e^{-A-B}) \pm (e^{A+B} + e^{-A+B} - e^{A-B} - e^{-A-B})}{4}$$

For $+$:

$$\frac{(e^{A+B} + e^{A-B} - e^{-A+B} - e^{-A-B}) + (e^{A+B} + e^{-A+B} - e^{A-B} - e^{-A-B})}{4}$$

$$= \frac{2e^{A+B} - 2e^{-A-B}}{4}$$

$$= \frac{e^{(A+B)} - e^{-(A+B)}}{2}$$

$$= \sinh(A + B)$$

For $-$:

$$\frac{(e^{A+B} + e^{A-B} - e^{-A+B} - e^{-A-B}) - (e^{A+B} + e^{-A+B} - e^{A-B} - e^{-A-B})}{4}$$

$$= \frac{2e^{A-B} - 2e^{-A+B}}{4}$$

$$= \frac{e^{(A-B)} - e^{-(A-B)}}{2}$$

$$= \sinh(A - B)$$

$$So \sinh(A \pm B) = \sinh(A) \cosh(B) \pm \sinh(B) \cosh(A)$$

$$\cosh A \cosh B \pm \sinh A \sinh B$$

$$= \frac{e^A + e^{-A}}{2} \times \frac{e^B + e^{-B}}{2} \pm \frac{e^A - e^{-A}}{2} \frac{e^B - e^{-B}}{2}$$

$$= \frac{(e^A + e^{-A})(e^B + e^{-B})}{4} \pm \frac{(e^A - e^{-A})(e^B - e^{-B})}{4}$$

$$= \frac{(e^A + e^{-A})(e^B + e^{-B}) \pm (e^A - e^{-A})(e^B - e^{-B})}{4}$$

$$= \frac{(e^{A+B} + e^{A-B} + e^{-A+B} + e^{-A-B}) \pm (e^{A+B} - e^{-A+B} - e^{A-B} + e^{-A-B})}{4}$$

For $+$:

$$\frac{(e^{A+B} + e^{A-B} + e^{-A+B} + e^{-A-B}) + (e^{A+B} - e^{-A+B} - e^{A-B} + e^{-A-B})}{4}$$

$$= \frac{2e^{A+B} + 2e^{-A-B}}{4}$$

$$= \frac{e^{(A+B)} + e^{-(A+B)}}{2}$$

$$= \cosh(A + B)$$

For $-$:

$$\frac{(e^{A+B} + e^{A-B} + e^{-A+B} + e^{-A-B}) - (e^{A+B} - e^{-A+B} - e^{A-B} + e^{-A-B})}{4}$$

$$= \frac{2e^{A-B} + 2e^{-A+B}}{4}$$

$$= \frac{e^{(A-B)} + e^{-(A-B)}}{2}$$

$$= \cosh(A - B)$$

$$So \cosh(A \pm B) = \cosh(A)\cosh(B) \pm \sinh(A)\sinh(B)$$

$$\tanh(A \pm B) = \frac{\sinh(A \pm B)}{\cosh(A \pm B)}$$

$$\tanh(A \pm B) = \frac{\sinh A \cosh B \pm \sinh B \cosh A}{\cosh A \cosh B \pm \sinh A \sinh B}$$

Divide the numerator and denominator by $\cosh A \cosh B$:

$$\tanh(A \pm B) = \frac{\left(\frac{\sinh A}{\cosh A} \times \frac{\cosh B}{\cosh B} \pm \frac{\sinh B}{\cosh B} \times \frac{\cosh A}{\cosh A}\right)}{\left(\frac{\cosh A}{\cosh A} \times \frac{\cosh B}{\cosh B} \pm \frac{\sinh A}{\cosh A} \times \frac{\sinh B}{\cosh B}\right)}$$

$$\tanh(A \pm B) = \frac{\tanh A \times 1 \pm \tanh B \times 1}{1 \times 1 \pm \tanh A \times \tanh B}$$

$$\tanh(A \pm B) = \frac{\tanh A \pm \tanh B}{1 \pm \tanh A \tanh B}$$

I will next discuss where the name of these functions comes from. Why are they called the hyperbolic functions?

A unit circle can be represented by the equation $x^2 + y^2 = 1$. Each point on this circle has coordinates (x, y). Writing this in terms of r and θ, each point on the circle has coordinates $(\cos\theta, \sin\theta)$ because $x = r\cos\theta$ and $y = r\sin\theta$ and $r = 1$ because this is a unit circle. From an algebraic perspective:

$$x^2 + y^2 = 1$$

$$x = r\cos\theta$$

$$y = r\sin\theta$$

$$(r\cos\theta)^2 + (r\sin\theta)^2 = 1$$

$$r^2\cos^2\theta + r^2\sin^2\theta = 1$$

$$r^2(\cos^2\theta + \sin^2\theta) = 1$$

$$r^2(1) = 1$$

$$r^2 = 1$$

$$r = 1$$

This works because $\sin^2\theta + \cos^2\theta \equiv 1$. The same cannot be said for the new functions because $\cosh^2\theta - \sinh^2\theta \equiv 1$.

But what if I instead plotted the graph $x^2 - y^2 = 1$ and $let\ x = r\cosh\theta$ and $let\ y = r\sinh\theta$? This means:

$$(r\cosh\theta)^2 - (r\sinh\theta)^2 = 1$$

$$r^2\cosh^2\theta - r^2\sinh^2\theta = 1$$

$$r^2(\cosh^2\theta - \sinh^2\theta) = 1$$

$$r^2 = 1$$

$$r = 1$$

So, each point on the curve (x, y) could be written as $(\cosh\theta, \sinh\theta)$.

I will take a moment to talk about what this line "$let\ x = r\cosh\theta$ and $let\ y = r\sinh\theta$" means.

Equations are typically in the form $f(x, y) = g(x, y)$, in other words, equations are in terms of x and y. For a given value of x, a value of y can be found and vice-versa. Parametric equations instead involve two equations one of the form $f(x, t) = g(x, t)$ and the other of the form $f(y, t) = g(y, t)$ where t is called the "parameter". Instead of finding the value of x for a given value of y or vice-versa, a value of y and a value of x can be found for a given value of t.

In the above example, r is a constant which we later found to be 1 and θ is the parameter. As you change the value of θ and plot the corresponding values of x and y you get the same shape as you would get by plotting $x^2 - y^2 = 1$.

This is all to say that the cartesian equation $x^2 - y^2 = 1$ is equivalent to the parametric equations $x = \cosh\theta$, $y = \sinh\theta$ (for positive values of x because the former is valid for $x < 0$ but the latter is not because $\cosh\theta$ is always > 0). The shape of the curve formed by these equations is called a hyperbola, hence the name, hyperbolic functions.

To finish, I will differentiate these functions and their inverses.

$$\frac{d}{dx}\sinh x =$$

$$\frac{d}{dx}\frac{e^x - e^{-x}}{2}$$

$$= \frac{1}{2}\left(\frac{d}{dx}(e^x) - \frac{d}{dx}(e^{-x})\right)$$

$$= \frac{1}{2}\left(e^x - (-e^{-x})\right)$$

$$= \frac{e^x + e^{-x}}{2}$$

$$= \cosh x$$

$$\frac{d}{dx}\cosh x$$

$$= \frac{d}{dx}\left(\frac{e^x + e^{-x}}{2}\right)$$

$$= \frac{1}{2}\left(\frac{d}{dx}(e^x) + \frac{d}{dx}(e^{-x})\right)$$

$$= \frac{1}{2}\left(e^x + (-e^{-x})\right)$$

$$= \frac{e^x - e^{-x}}{2}$$

$$= \sinh x$$

$$\frac{d}{dx}\tanh x$$

$$= \frac{d}{dx}\frac{\sinh x}{\cosh x}$$

$$= \frac{\sinh' x \cosh x - \sinh x \cosh' x}{\cosh^2 x}$$

$$= \frac{\cosh x \cosh x - \sinh x \sinh x}{\cosh^2 x}$$

$$= \frac{\cosh^2 x - \sinh^2 x}{\cosh^2 x}$$

$$\frac{1}{\cosh^2 x}$$

$$= \operatorname{sech}^2 x$$

$$\frac{d}{dx} \coth x$$

$$= \frac{d}{dx} \frac{\cosh x}{\sinh x}$$

$$= \frac{\cosh' x \sinh x - \cosh x \sinh' x}{\sinh^2 x}$$

$$= \frac{\sinh x \sinh x - \cosh x \cosh x}{\sinh^2 x}$$

$$= \frac{\sinh^2 x - \cosh^2 x}{\sinh^2 x}$$

$$= \frac{-1}{\sinh^2 x}$$

$$= -\operatorname{csch}^2 x$$

$$\frac{d}{dx} \operatorname{sech} x$$

$$= \frac{d}{dx} \frac{1}{\cosh x}$$

$$= \frac{1' \cosh x - 1 \cosh' x}{\cosh^2 x}$$

$$= \frac{0 \cosh x - 1 \sinh x}{\cosh^2 x}$$

$$= -\frac{\sinh x}{\cosh^2 x}$$

$$= -\tanh x \operatorname{sech} x$$

$$\frac{d}{dx} \operatorname{csch} x$$

$$= \frac{d}{dx} \frac{1}{\sinh x}$$

$$= \frac{1' \sinh x - 1 \sinh' x}{\sinh^2 x}$$

$$= \frac{0 \sinh x - 1 \cosh x}{\sinh^2 x}$$

$$= -\frac{\cosh x}{\sinh^2 x}$$

$$= -\coth x \operatorname{csch} x$$

For the inverse hyperbolic functions:

$$find\ \frac{dy}{dx}\ when:$$

$$y = \operatorname{arcsinh} x$$

$$\sinh y = x$$

$$\cosh y\, dy = dx$$

$$\frac{dy}{dx} = \frac{1}{\cosh y}$$

$$\cosh^2 y - \sinh^2 y = 1$$

$$\cosh y = \sqrt{\sinh^2 y + 1}$$

$$\cosh y\ = \sqrt{x^2 + 1}$$

$$\frac{dy}{dx} = \frac{1}{\sqrt{x^2 + 1}}$$

$$y = \operatorname{arccosh} x$$

$$\cosh y = x$$

$$\sinh y\, dy = dx$$

$$\frac{dy}{dx} = \frac{1}{\sinh y}$$

$$\cosh^2 y - \sinh^2 y = 1$$

$$\sinh y = \sqrt{\cosh^2 y - 1}$$

$$\sinh y = \sqrt{x^2 - 1}$$

$$\frac{dy}{dx} = \frac{1}{\sqrt{x^2 - 1}}$$

$$y = \operatorname{arctanh} x$$

$$\tanh y = x$$

$$\operatorname{sech}^2 y\, dy = dx$$

$$\frac{dy}{dx} = \frac{1}{\operatorname{sech}^2 y}$$

$$\cosh^2 y - \sinh^2 y = 1$$

$$\frac{\cosh^2 y - \sinh^2 y}{\cosh^2 y} = \frac{1}{\cosh^2 y}$$

$$1 - \tanh^2 y = \operatorname{sech}^2 y$$

$$\operatorname{sech}^2 y = 1 - x^2$$

$$\frac{dy}{dx} = \frac{1}{1 - x^2}$$

$$y = \operatorname{arccoth} x$$

$$\coth y = x$$

$$-\operatorname{csch}^2 y\, dy = dx$$

$$\frac{dy}{dx} = \frac{-1}{\operatorname{csch}^2 y}$$

$$\cosh^2 y - \sinh^2 y = 1$$

$$\frac{\cosh^2 y - \sinh^2 y}{\sinh^2 y} = \frac{1}{\sinh^2 y}$$

$$\coth^2 y - 1 = \operatorname{csch}^2 y$$

$$\operatorname{csch}^2 y = x^2 - 1$$

$$\frac{dy}{dx} = \frac{-1}{x^2 - 1}$$

$$\frac{dy}{dx} = \frac{1}{1 - x^2}$$

$$y = \operatorname{arcsech} x$$

$$\operatorname{sech} y = x$$

$$-\tanh y \operatorname{sech} y\, dy = dx$$

$$\frac{dy}{dx} = \frac{-1}{\tanh y \operatorname{sech} y}$$

$$\frac{dy}{dx} = \frac{-1}{x \tanh y}$$

$$1 - \tanh^2 y = sech^2\, y$$

$$\tanh y = \sqrt{1 - \operatorname{sech}^2 y}$$

$$\tanh y = \sqrt{1 - x^2}$$

$$\frac{dy}{dx} = \frac{-1}{x\sqrt{1-x^2}}$$

$$y = \operatorname{arccsch} x$$

$$\operatorname{csch} y = x$$

$$-\coth y \operatorname{csch} y \, dy = dx$$

$$\frac{dy}{dx} = \frac{-1}{\coth y \operatorname{csch} y}$$

$$\frac{dy}{dx} = \frac{-1}{x \coth y}$$

$$\coth^2 y - 1 = \operatorname{csch}^2 y$$

$$\coth y = \sqrt{\operatorname{csch}^2 y + 1}$$

$$\coth y = \sqrt{x^2 + 1}$$

$$\frac{dy}{dx} = \frac{-1}{x\sqrt{x^2+1}}$$

Some of the derivative of inverse trig and hyperbolic functions need a modulus.

$$\operatorname{arcsec}' x = \frac{1}{|x|\sqrt{x^2-1}}$$

$$\operatorname{arccsc}' x = \frac{-1}{|x|\sqrt{x^2-1}}$$

$$\operatorname{arcsech}' x = \frac{-1}{|x|\sqrt{1-x^2}}$$

$$\operatorname{arccsch}' x = \frac{-1}{|x|\sqrt{x^2+1}}$$

This is because the gradients of these functions are always negative (always positive for $arcsec$). The modulus is necessary because the $\sqrt{}$ function only accounts for the positive root.

This fact can be found be plotting the graphs of $x = \sec y$, $x = \csc y$, $x = \operatorname{sech} y$ and $x = \operatorname{csch} y$ by using the Maclaurin series for $sine$, $cosine$ and the exponential function. You will also find that some of these functions are only valid within certain ranges.

One more thing worth noting:

Using $\sin(x) = \frac{e^{ix}-e^{-ix}}{2i}$ & $\cos(x) = \frac{e^{ix}+e^{-ix}}{2}$

$$\sin(ix) = \frac{e^{i(ix)} - e^{-i(ix)}}{2i}$$

$$\sin(ix) = \frac{e^{-x} - e^{x}}{2i}$$

$$\& \; since \frac{1}{i} = -i$$

$$so \sin(ix) = -i\left(\frac{e^{-x} - e^{x}}{2}\right)$$

$$\sin(ix) = i\left(\frac{e^{x} - e^{-x}}{2}\right)$$

$$\sin(ix) = i\sinh(x)$$

$$\cos(ix) = \frac{e^{i(ix)} + e^{-i(ix)}}{2}$$

$$\cos(ix) = \frac{e^{-x} + e^{x}}{2}$$

$$\cos(ix) = \cosh(x)$$

Also using $\sinh(x) = \frac{e^{x}-e^{-x}}{2}$ and $\cosh(x) = \frac{e^{x}+e^{-x}}{2}$

$$\sinh(ix) = \frac{e^{ix} - e^{-ix}}{2}$$

$$\sinh(ix) = i\frac{e^{ix} - e^{-ix}}{2i}$$

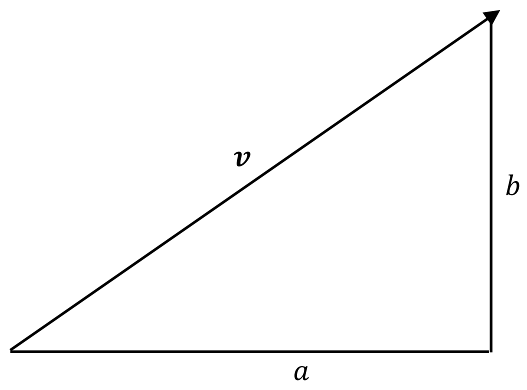
$$\sinh(ix) = i\sin(x)$$

$$\cosh(ix) = \frac{e^{ix} + e^{-ix}}{2}$$

$$\cosh(ix) = \cos(x)$$

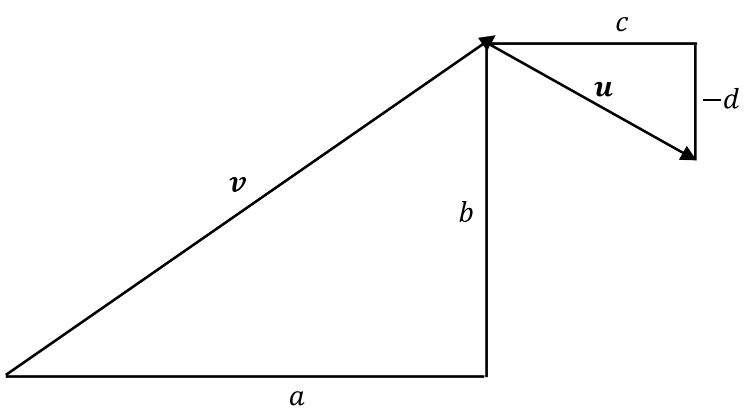
Vectors

Scalars are essentially just numbers; they can represent a magnitude. Vectors can be used to describe both a magnitude and a direction. This is typically done by using scalars to describe the different components of the direction separately. Vectors are typically written as **bold** letters in print or <u>underlined</u> letters when writing. For example, this vector $\boldsymbol{v}$:

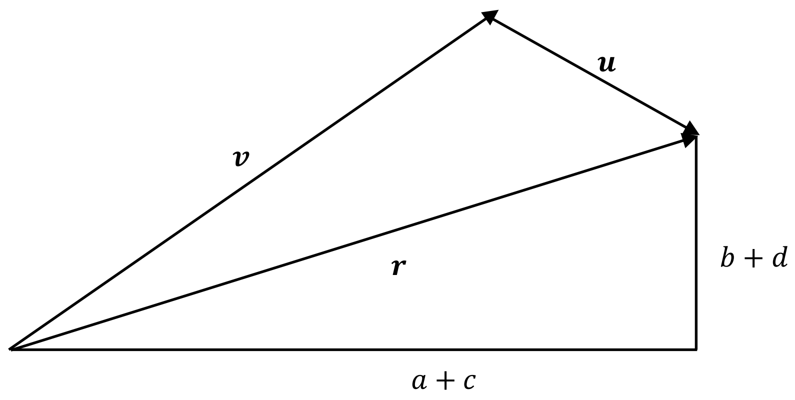


has a horizontal component of a and a vertical component of b so could be written using column vector notation as $\boldsymbol{v} = \binom{a}{b}$. The magnitude (length) of a vector can be found using the Pythagorean theorem. $\boldsymbol{v} = \binom{a}{b}$, $||\boldsymbol{v}|| = \sqrt{a^2 + b^2}$ where $||\boldsymbol{v}||$ means the magnitude of $\boldsymbol{v}$.

Two vectors $\boldsymbol{v} = \binom{a}{b}$ and $\boldsymbol{u} = \binom{c}{d}$ can be joined together like so:



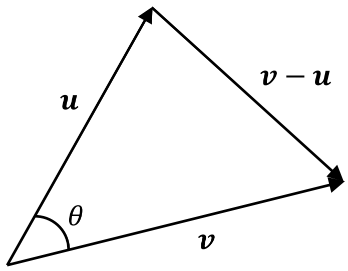
In this example d is negative and so the length has been labelled as $-d$ because lengths are always positive. If I were to create a vector going directly from the beginning of $\boldsymbol{v}$ to the end of $\boldsymbol{u}$ then the horizontal component of this vector would be $a + c$ and the vertical component would be $b - -d = b + d$. This is called the resultant vector and I will label it $\boldsymbol{r}$:



So $\boldsymbol{r} = \binom{a+c}{b+d}$. It would make sense to define vector addition in the following way: $\boldsymbol{u} + \boldsymbol{v}$ is the resultant vector from drawing one vector at the end of the other: $\binom{a}{b} + \binom{c}{d} = \binom{a+c}{b+d}$. From this, multiplying a vector by a scalar by a vector can be easily defined. $2\binom{a}{b} = \binom{a}{b} + \binom{a}{b} = \binom{a+a}{b+b} = \binom{2a}{2b}$, $3\binom{a}{b} = \binom{a}{b} + \binom{a}{b} + \binom{a}{b} = \binom{a+a+a}{b+b+b} = \binom{3a}{3b}$. In general, vector-scalar multiplication should be defined: $k\binom{a}{b} = \binom{ka}{kb}$. Visually, this would mean that the vectors direction would remain unchanged, but its direction would be multiplied by a factor of k. Multiplication of two vectors together is more complicated. I will speak on this later.

In the above diagram, $\boldsymbol{v} + \boldsymbol{u} = \boldsymbol{r}$. This means that $\boldsymbol{u} = \boldsymbol{r} - \boldsymbol{v}$. This is because another way of getting from the beginning to the end of vector $\boldsymbol{u}$ is by going backwards along $\boldsymbol{v}$ and forwards along $\boldsymbol{r}$ so $\boldsymbol{u} = -\boldsymbol{v} + \boldsymbol{r} = \boldsymbol{r} - \boldsymbol{v}$.

To find the angle θ between to vectors $\boldsymbol{u} = \binom{a}{b}$ and $\boldsymbol{v} = \binom{c}{d}$ I will use the following diagram:



This triangle can be formed, with side lengths equal to the magnitudes of the given vectors. Using this fact and the cosine rule: $c^2 = a^2 + b^2 - 2ab\cos(\theta)$ we can write the following:

$$\left|\left|\boldsymbol{v} - \boldsymbol{u}\right|\right|^2 = \left|\left|\boldsymbol{u}\right|\right|^2 + \left|\left|\boldsymbol{v}\right|\right|^2 - 2\left|\left|\boldsymbol{u}\right|\right|\left|\left|\boldsymbol{v}\right|\right|\cos(\theta)$$

Since $\left|\left|\boldsymbol{u}\right|\right| = \sqrt{a^2+b^2}$ and $\left|\left|\boldsymbol{v}\right|\right| = \sqrt{c^2+d^2}$, we can say $\left|\left|\boldsymbol{u}\right|\right|^2 = a^2 + b^2$ and $\left|\left|\boldsymbol{v}\right|\right|^2 = c^2 + d^2$. Also, since $\boldsymbol{v} - \boldsymbol{u} = \binom{c}{d} - \binom{a}{b} = \binom{c-a}{d-b}$, $\left|\left|\boldsymbol{v} - \boldsymbol{u}\right|\right| = \sqrt{(c-a)^2 + (d-b)^2}$ so $\left|\left|\boldsymbol{v} - \boldsymbol{u}\right|\right|^2 = (c-a)^2 + (d-b)^2$.

Using these in the above formula:

$$(c-a)^2 + (d-b^2) = a^2 + b^2 + c^2 + d^2 - 2\left|\left|\boldsymbol{u}\right|\right|\left|\left|\boldsymbol{v}\right|\right|\cos(\theta)$$

$$c^2 - 2ac + a^2 + d^2 - 2bd + b^2 = a^2 + b^2 + c^2 + d^2 - 2\left|\left|\boldsymbol{u}\right|\right|\left|\left|\boldsymbol{u}\right|\right|\cos\theta$$

$$2\left|\left|\boldsymbol{u}\right|\right|\left|\left|\boldsymbol{u}\right|\right|\cos\theta = a^2 + b^2 + c^2 + d^2 - c^2 + 2ac - a^2 - d^2 + 2bd - b^2$$

$$2\left|\left|\boldsymbol{u}\right|\right|\left|\left|\boldsymbol{u}\right|\right|\cos\theta = 2ac + 2bd$$

$$\left|\left|\boldsymbol{u}\right|\right|\left|\left|\boldsymbol{u}\right|\right|\cos\theta = ac + bd$$

This could be one useful way in which the multiplication of two vectors could be defined. For $\boldsymbol{u} = \binom{a}{b}$ and $\boldsymbol{v} = \binom{c}{d}$ $\boldsymbol{uv} = ac + bd = \left|\left|\boldsymbol{u}\right|\right|\left|\left|\boldsymbol{u}\right|\right|\cos\theta$ where θ is the angle between $\boldsymbol{u}$ and $\boldsymbol{v}$. From this fact comes the definition of the "dot product" of two vectors, written as

$$\boldsymbol{u} \cdot \boldsymbol{v} = ac + bd = \left|\left|\boldsymbol{u}\right|\right|\left|\left|\boldsymbol{u}\right|\right|\cos\theta$$

Calculus Part 4: Multivariable Functions

We have so far seen functions which take a single numerical input and produce a single numerical output. Multivariable functions can take multiple inputs to produce a single output, for example:

$$f(x, y) = 5x^2 - 13y$$

A single variable function can be represented by a curve on a graph where the horizontal coordinate of a point on the curve represents some x value (an input top the function) and the vertical coordinate at that point represents the corresponding y value (the output of the function for the given input). A multivariable function can be similarly represented in 3D space. A flat grid may be created and a set of coordinates (x, y) can be chosen. These values for x and y are then put through the function and the output determines the height above the plane where a point can be plotted. If this is done for all coordinates (all possible pairs of x and y) then the 3-dimensional curve will be completed. The purpose of this chapter is to investigate how calculus can be done on these 3D curves.

First, how should differentiation be defined for these 3D curves? One way could be to let the y-coordinate be considered a constant (for example if the point had coordinates $(2, 3, 5)$ (using (x, y, z))) then we would let $y = 3$ and treat it as a constant. This would turn z, a function of x and y into a function of just x. The gradient of this function can then be found at the point (in this case, the point with x-coordinate 2). This gives us what is called the *partial derivative* of z with respect to x $\frac{\partial z}{\partial x}$, where ∂ (del) is a symbol which distinguishes between regular derivatives which use d and partial derivatives which use ∂. The function x created by doing the above, when plotted will give a cross-section of the 3D curve going through the given point. Doing the same but letting x be constant turning z into a function of y allows us to do very much the same thing, the derivative of this curve is the partial derivative of z with respect to y $\frac{\partial z}{\partial y}$.

Using partial derivatives, the slope of the line can be found along the x or y axis. How could I find the derivative in some direction described for example by vector $\boldsymbol{u}$?

This is called the directional derivative and is written as $D_{\boldsymbol{u}}f(x, y)$ where $\boldsymbol{u}$ is the vector describing the direction in which the derivative will describe the slope. With single variable calculus the difference quotient could be used to find derivatives. Something similar will need to be done here.

For a curve $z = f(x, y)$ find the directional derivative with some vector unit $\boldsymbol{u} = \binom{u_1}{u_2}$ (I am saying that $\boldsymbol{u}$ is a unit vector because the length does not matter and 1 is an easy number to work with) at the fixed point (x_0, y_0) (I will explain why this is a fixed point later). To do this I will need to find the value of f for some point in the direction of this vector very close to (x_0, y_0) and subtract from this, the value of f at (x_0, y_0). This will give be a change in f, Δf. The distance between the point (x_0, y_0) and the point which is close to it in the direction of $\boldsymbol{u}$, I will call h. The directional derivative $\approx \frac{\Delta f}{h}$. As $h \to 0$, $\frac{\Delta f}{h} \to D_{\boldsymbol{u}}f(x_0, y_0)$. I will call $P(x_0, y_0)$ and Q The other point we have been referring to and will let its coordinates be (x, y). Since the coordinates of Q change as h changes this means that x and y are functions of h. Since $\boldsymbol{u}$ is a unit vector, $Q = P + h\boldsymbol{u}$. Technically Q and P are coordinates, but if I treat them as "position vectors" (the vector going from the origin to a given point) then the above can be done. $h\boldsymbol{u} = h\binom{u_1}{u_2} = \binom{hu_1}{hu_2}$. The "position vector" of $Q = \binom{x}{y}$ and the "position vector" of $P = \binom{x_0}{y_0}$. This means $\binom{x}{y} = \binom{x_0}{y_0} + \binom{hu_1}{hu_2}$. So $x = x_0 + hu_1$ and $y = y_0 + hu_2$. As I said before, x and y are both functions of h, these are those functions. This means that since $D_{\boldsymbol{u}}f(x,y) \approx \frac{\Delta f}{h}$, where $\Delta f = f(x,y) - f(x_0,y_0) = f(x_0 + hu_1, y_0 + hu_2) - f(x_0,y_0)$:

$$D_{\boldsymbol{u}}f(x_0,y_0) \approx \frac{f(x_0 + hu_1, y_0 + hu_2) - f(x_0,y_0)}{h}$$

$$D_{\boldsymbol{u}}f(x_0,y_0) = \lim_{h\to 0}\frac{f(x_0 + hu_1, y_0 + hu_2) - f(x_0,y_0)}{h}$$

The above limit finds the derivative of some function $f(x,y)$ at some point (x_0,y_0) in the direction $\boldsymbol{u}$. You may wonder why I have used (x_0,y_0) as opposed to (x,y). Why did I make P the fixed point as opposed to Q? The reason for this is because if I made Q the fixed point e.g., with coordinates (x_Q, y_Q) then as $h \to 0$, $P \to Q$ which would give me the gradient at Q as opposed to at P. Also because I am subtracting the value of the function at P from the value of the function at Q, I would have gotten the negative slope because that would be the wrong way around. It is only the right way round because I am trying the find the slope at P which means P must be fixed and Q variable so that as $h \to 0$, $Q \to P$.

I will look now at a new way to view the chain rule by using a "dependency diagram".

$$\begin{tikzcd}
\boxed{y} \\
\boxed{t} \arrow[u] \\
\boxed{x} \arrow[u]
\end{tikzcd}$$

This dependency diagram says that y depends on t and that t depends on x. In other words, y is a function of t and t is a function of x e.g., $y = f(t),\ t = g(x)$. This means that for a small change in x, Δx, there is some change in t, $\Delta t \approx \frac{dt}{dx}\Delta x$. (This approximation is better the smaller that Δx and Δt are.) For a small change in t, Δt there is a small change in y, $\Delta y \approx \frac{dy}{dt}\Delta t$. This comes from the fact that when Δx, Δt and Δy are small, the curve on that interval $\approx$ a straight line meaning $gradient \approx \frac{rise}{run}$ so $rise \approx gradient \times run$.

This all means that for some small Δx, $\Delta t \approx \frac{dt}{dx}\Delta x$ and $\Delta y \approx \frac{dy}{dt}\Delta t \approx \frac{dy}{dt}\frac{dt}{dx}\Delta x$

$$since\ y = f(t), \qquad \frac{dy}{dt} = f'(t)$$

$$since\ t = g(x), \qquad \frac{dt}{dx} = g'(x)$$

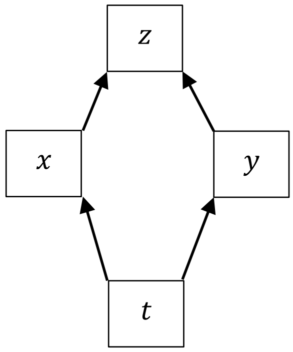
$$\Delta y \approx f'(t)g'(x)\Delta x$$

$$\Delta y = f'(g(x))g'(x)\Delta x$$

This means that, since Δx is small, the gradient $\approx$ constant $gradient \approx \frac{rise}{run} = \frac{\Delta y}{\Delta x} = f'(g(x))g'(x)$

As $\Delta x \to 0$, $gradient \to f'\big(g(x)\big)g'(x)$. This is another way to justify what we already know as the chain rule. A similar method can be used to create a chain rule for multivariable functions.

Let $z = f(x, y)$ where x and y are both functions of t. The dependency diagram for this would look like this:



This diagram says that z depends on x and y and that x depends on t and y depends on t. If I change t by some small amount Δt the change in x could be found by using $\Delta x \approx \frac{dx}{dt}\Delta\mathrm{t}$. When x changes by some small amount Δx, this causes z to change by some amount $\Delta z \approx \frac{\partial z}{\partial x}\Delta x \approx \frac{\partial z}{\partial x}\frac{dx}{dt}\Delta\mathrm{t}$. Here partial derivatives are used because I am differentiating z (a function of both x and y) with respect to just one variable. A small change in t, Δt would cause the small change in y, $\Delta y \approx \frac{dy}{dt}\Delta\mathrm{t}$. A small change in y, Δy causes a small change in z, $\Delta z \approx \frac{\partial z}{\partial y}\Delta y \approx \frac{\partial z}{\partial y}\frac{dy}{dt}\Delta\mathrm{t}$. Overall, changing t by some small amount Δt causes z to change by about $\frac{\partial z}{\partial x}\frac{dx}{dt}\Delta\mathrm{t}$ and then by about $\frac{\partial z}{\partial y}\frac{dy}{dt}\Delta\mathrm{t}$. Therefore, the overall change in z, $\Delta z \approx \frac{\partial z}{\partial x}\frac{dx}{dt}\Delta\mathrm{t} + \frac{\partial z}{\partial y}\frac{dy}{dt}\Delta\mathrm{t}$.

$$\frac{\Delta z}{\Delta t} \approx \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}$$

Since Δt is small, $slope \approx \frac{rise}{run} = \frac{\Delta z}{\Delta t}$

$$so\ slope \approx \frac{\Delta z}{\Delta t} \approx \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}$$

as $\Delta t \to 0$, $slope \to \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}$

$$hence\ \frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}$$

Returning to the directional derivative:

$$D_{\boldsymbol{u}}f(x_0, y_0) = \lim_{h\to 0}\frac{f(x_0 + hu_1, y_0 + hu_2) - f(x_0, y_0)}{h}$$

When using the difference quotient from single variable calculus, a change in y is divided by a change in x giving $\frac{dy}{dx}$. Here, a change in f is divided by a change in h giving $\frac{df}{dh}$. This means that $D_{\boldsymbol{u}}f(x_0, y_0) = \frac{df}{dh}$. $Q = (x_0, y_0)$ only when $h = 0$ so $D_{\boldsymbol{u}}f(x, y)|_{(x_0, y_0)} = \frac{df}{dh}\Big|_{h=0}$. This notation means the directional derivative of f at (x_0, y_0) equals $\frac{df}{dh}$ at $h = 0$ When $h = 0$, $x = x_0 + hu_1 = x_0$ and $y = y_0 + hu_2 = y_0$ so $\left.\frac{df}{dh}\right|_{h=0} = \left.\frac{df}{dh}\right|_{x=x_0, y=y_0}$ which is usually written as $\left.\frac{df}{dh}\right|_{(x_0,y_0)}$.

f is a function given by $f(x, y)$. Since $x = x_0 + hu_1$ and $y = y_0 + hu_2$, f could be written as:

$$f(x_0 + hu_1, y_0 + hu_2)$$

Using $\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}$:

$$\frac{df}{dh} = \frac{\partial f}{\partial x}\frac{dx}{dh} + \frac{\partial f}{\partial y}\frac{dy}{dh}$$

$$\frac{dx}{dh} = \frac{d}{dh}(x_0 + hu_1) = u_1$$

$$\frac{dy}{dh} = \frac{d}{dh}(y_0 + hu_2) = u_2$$

$$so\,\frac{df}{dh} = \frac{\partial f}{\partial x}u_1 + \frac{\partial f}{\partial y}u_2$$

Since $D_{\boldsymbol{u}}f(x, y)|_{(x_0,y_0)} = \left.\frac{df}{dh}\right|_{h=0}$,

$$D_{\boldsymbol{u}}f(x_0\,, y_0) = \left.\frac{\partial f}{\partial x}\right|_{(x_0,y_0)} u_1 + \left.\frac{\partial f}{\partial y}\right|_{(x_0,y_0)} u_2$$

This could be used to work out the point vector $\boldsymbol{u}$ which will yield the maximum slope.

Recalling the definition of the dot product: $\boldsymbol{u} \cdot \boldsymbol{v} = ac + bd = ||\boldsymbol{u}||\,||\boldsymbol{u}|| \cos\theta$

$$D_{\boldsymbol{u}}f(x_0, y_0) = \begin{pmatrix} \left.\frac{\partial f}{\partial x}\right|_{(x_0,y_0)} \\ \left.\frac{\partial f}{\partial y}\right|_{(x_0,y_0)} \end{pmatrix} \cdot \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

$$D_{\boldsymbol{u}}f(x_0, y_0) = \begin{pmatrix} \left.\frac{\partial f}{\partial x}\right|_{(x_0,y_0)} \\ \left.\frac{\partial f}{\partial y}\right|_{(x_0,y_0)} \end{pmatrix} \cdot \boldsymbol{u}$$

$$D_{\boldsymbol{u}}f(x_0, y_0) = \left.\begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix}\right|_{(x_0,y_0)} \cdot \boldsymbol{u}$$

$$D_{\boldsymbol{u}}f(x_0, y_0) = \left\| \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix} \right\|_{(x_0, y_0)} \left\| \boldsymbol{u} \right\| \cos\theta$$

Replacing specific values (x_0, y_0) with general (x, y):

$$D_{\boldsymbol{u}}f(x, y) = \left\| \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix} \right\| \left\| \boldsymbol{u} \right\| \cos\theta$$

I will let $\boldsymbol{p} = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix}$ as this is the vector of partial derivatives.

$\boldsymbol{u}$ was defined as a unit vector so $\left\|\boldsymbol{u}\right\| = 1$. The maximum value of the above is where $\cos\theta = 1$ (since $\theta \in \mathbb{R}$ (θ is real)). This means that the maximum slope of a multivariable function at some point (x, y) is the magnitude of the vector of partial derivatives $= \left\|\boldsymbol{p}\right\|$. This occurs when $\cos\theta = 1$ meaning $\theta = 0$. This means that the angle between the two vectors $\boldsymbol{u}$ and $\boldsymbol{p}$ is 0 meaning they both face the same direction. This means that to find the maximum slope of a curve at some point (x, y) you must find the directional derivative in the direction of $\boldsymbol{u}$ where $\boldsymbol{u}$ faces the same direction as $\boldsymbol{p}$. In other words, the direction of $\boldsymbol{p}$ is the direction in which the slope is at its greatest. This vector of partial derivatives is referred to as the "gradient" of f and is written as ∇f "del f". To summarise:

∇f is the vector of partial derivatives of f. It is the vector which describes the direction in which the instantaneous change in f is greatest. The magnitude of the change is $\left\|\nabla f\right\|$.

I will now move on to the integration of multivariable functions. Integration of single variable functions find the area under a curve, integration of multivariable functions find the volume under a curve. For some multi variable function $f(x, y)$:

I could integrate this with respect to x treating y as a function giving

$$\int_a^b f(x, y)dx$$

which would be a function of y. This function would give you the area under curve on the cross-section which is parallel to the x-axis. Setting y equal to some value would give you the area under this curve from a to b at the given y value (since this cross-sectional curve changes as you change the value of y). Integrating this area as y goes from c to d will give the volume under the curve on the intervals x: $[a, b]$ and y: $[c, d]$.

$$Volume = \int_{y=c}^{d} \int_{x=a}^{b} f(x, y)dx\, dy$$

This is called a double integral. Another way of thinking about this could be to integrate $f(x, y)$ as y goes from c to d and then as x goes from a to b giving the same volume but this time as:

$$\int_{x=a}^{b}\int_{y=c}^{d} f(x,y)dy\,dx$$

Another way to consider this would be to think of $f(x,y)dxdy$ as being the infinitesimal volume of some cuboid, height $f(x,y)$, width dx and depth dy. Integrating this along the x-axis from a to b gives the still infinitesimal volume:

$$\int_{x=a}^{b} f(x,y)dxdy$$

Integrating this volume along the y-axis from c to d gives a regular volume:

$$\int_{y=c}^{d}\int_{x=a}^{b} f(x,y)\,dxdy$$

However, you choose to look at it these all mean the same thing.

One of my favourite applications of the double integral is in solving the gaussian integral:

$$\int_{-\infty}^{\infty} e^{-x^2}dx$$

This integral may seem innocent, however no matter how hard you try you will never be able to find an antiderivative for e^{-x^2}. This is because this integral is non-elementary (cannot be expressed using the functions which we are used to). We will need to solve it a different way.

This problem may not at first seem to have anything to do with double integrals, but the way in which it was solved by Gauss was by turning it into one.

$$Let\ I = \int_{-\infty}^{\infty} e^{-x^2}dx$$

$$I^2 = \left(\int_{-\infty}^{\infty} e^{-x^2}dx\right)^2$$

$$I^2 = \int_{-\infty}^{\infty} e^{-x^2}dx \times \int_{-\infty}^{\infty} e^{-x^2}dx$$

Since x is a dummy variable (it does not exist in the final answer) it can be replaced by any other variable name, e.g., y meaning:

$$\int_{-\infty}^{\infty} e^{-x^2}dx = \int_{-\infty}^{\infty} e^{-y^2}dy$$

$$So\ I^2 = \int_{-\infty}^{\infty} e^{-x^2}dx \times \int_{-\infty}^{\infty} e^{-y^2}dy$$

Since $\int_{-\infty}^{\infty} e^{-y^2}dy$ is just some number, a constant (though we don't yet know what it is), we can treat it as a constant with respect to x. Hence, we can bring it into the first integral.

$$I^2 = \int_{-\infty}^{\infty}\left(e^{-x^2} \times \int_{-\infty}^{\infty} e^{-y^2}dy\right)dx$$

Since x is a constant with respect to y (x does not depend on y), e^{-x^2} can be brought inside the integral next to it.

$$I^2 = \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{\infty} e^{-x^2} e^{-y^2} dy\, dx$$

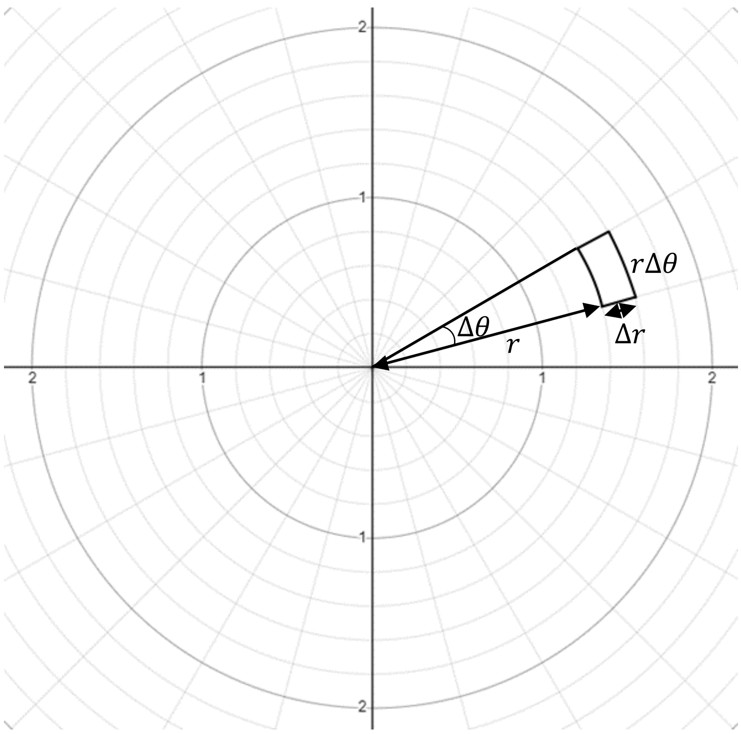
$$I^2 = \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{\infty} e^{-x^2-y^2} dy\, dx$$

$$I^2 = \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{\infty} e^{-(x^2+y^2)} dy\, dx$$

This integral certainly seems more difficult than the original but this is all necessary. The above represents the total volume under the curve $z = e^{-(x^2+y^2)}$. The way in which gauss found this was by writing the function in terms of polar coordinates. Since $x^2 + y^2 = r^2$:

$$z = e^{-r^2}$$

To find the total volume under this curve we will need to integrate some infinitesimal volume with respect to r and then with respect to θ. Looking at the polar grid (as opposed to the rectangular/cartesian grid):



The outlined region in the diagram has a length Δr and an arc length $r\Delta\theta$. The area of this shape $\approx r\Delta\theta\Delta r$ since the shape approximates a rectangle. As $\Delta\theta \to 0$ this shape approaches a rectangle and so its $area \to r\Delta r\Delta\theta$. So if I use the differential $d\theta$ which means the same thing as $\Delta\theta \to 0$ then:

$$Area = r\Delta r d\theta$$

If the volume under the curve of the graph $z = e^{-r^2}$ was split infinitesimal volumes, each of these could have a height of e^{-r^2} and a cross-sectional area of $rdrd\theta$. Each volume would hence be $e^{-r^2} rdrd\theta = re^{-r^2} drd\theta$. Integrating this as r goes from 0 to ∞ gives a function of θ representing the cross sectional area of the curve between $r = 0$ and $r = \infty$. This is a function of θ as this area depends on the angle at which this cross-section is drawn. In this particular case the angle θ does not affect this area and so does not appear in the function although, technically it is still a function of θ. This function of θ could then be integrated with respect to θ as θ goes from 0 to 2π (although this would technically work for any interval range 2π e.g., $[7\pi, 9\pi]$) This will give the volume under the entire curve. re^{-r^2} could be integrated as θ goes from 0 to 2π giving the cross-sectional area of the ring radius r around the centre. This area would be a function of r (actually this time because this area would change with r). This function could then be integrated as r goes from 0 to ∞ to obtain the full volume under the curve. This is all to say that:

$$I^2 = \int_{\theta=0}^{2\pi} \int_{r=0}^{\infty} re^{-r^2} dr\, d\theta$$

First, I must solve:

$$\int_0^{\infty} re^{-r^2} dr$$

This looks very similar to the original integral which could not be solved. I could try to solve this by using integration by parts. Since I cannot integrate e^{-r^2} I will chose to integrate r and differentiate e^{-r^2}:

$$\begin{array}{ccc} & D & I \\ + & e^{-r^2} & r \\ - & -2re^{-r^2} & \frac{1}{2}r^2 \end{array}$$

As I continue, the product of the bottom row becomes more and more complicated and the e^{-r^2} part never disappears so this isn't going to work. I will instead attempt a substitution.

$$Let\ u = -r^2$$

$$du = -2r\ dr$$

$$dr = \frac{1}{-2r} du$$

$$\int_0^{\infty} re^{-r^2} dr = \int_{r=0}^{\infty} \frac{re^u}{-2r} du$$

$$= -\frac{1}{2}\int_{r=0}^{\infty} e^u du$$

$$= -\frac{1}{2}[e^u]_{r=0}^{\infty}$$

$$= -\frac{1}{2}\left[e^{-r^2}\right]_0^{\infty}$$

$$= -\frac{1}{2}\left(\lim_{n\to\infty} e^{-n^2} - e^{-0^2}\right)$$

$$as\ n \to \infty$$

$$n^2 \to \infty,$$

$$-n^2 \to -\infty$$

$$e^{-n^2} = \frac{1}{e^{n^2}} \to \frac{1}{\infty} \to 0$$

$$-\frac{1}{2}\left(\lim_{n\to\infty} e^{-n^2} - e^{-0^2}\right)$$

$$= -\frac{1}{2}(0-1)$$

$$\frac{1}{2}$$

$$So \int_0^{\infty} re^{-r^2} dr = \frac{1}{2}$$

$$I^2 = \int_{\theta=0}^{2\pi} \int_{r=0}^{\infty} re^{-r^2} dr\, d\theta = I^2 = \int_{\theta=0}^{2\pi} \frac{1}{2} d\theta$$

$$I^2 = \left[\frac{1}{2}\theta\right]_0^{2\pi}$$

$$I^2 = \frac{2\pi}{2} - \frac{0}{2}$$

$$I^2 = \pi$$

$$I = \pm\sqrt{\pi}$$

Recalling the original integral: $I = \int_{-\infty}^{\infty} e^{-x^2} dx$, it can be observed that $e^{-x^2} > 0$ for all real x hence the area under the curve must be positive and so:

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$$

Mathematician Laplace (who we will hear more about later) solved the integral in a different, yet similar way.

$$Let\ f(x) = e^{-x^2}$$

$$then\ f(x) = f(-x)\ since\ x^2 = (-x)^2\ meaning\ the\ curve\ is\ symmetrical\ about\ y = 0$$

$$Let\ I = \int_{-\infty}^{\infty} e^{-x^2} dx$$

$$I = 2\int_0^{\infty} e^{-x^2} dx$$

This previous step will be very important later. The next step is similar to what Gauss did.

$$I^2 = 4\int_0^{\infty} e^{x^2} dx \int_0^{\infty} e^{-y^2} dy$$

$$I^2 = 4\int_{x=0}^{\infty} \int_{y=0}^{\infty} e^{-(x^2+y^2)} dy\, dx$$

Laplace then used the following substitution:

$$y = xt$$

Where y is a function of t (remember y is constant with respect to x). From the above we get the following:

$$y^2 = x^2t^2$$

$$t = \frac{y}{x}$$

$$dy = xdt$$

$$I^2 = 4\int_{x=0}^{\infty}\int_{y=0}^{\infty} e^{-(x^2+x^2t^2)}xdt\,dx$$

Changing the bounds from y to t: when $y = 0$, $t = \frac{y}{x} = \frac{0}{x} = 0$.

As $y \to \infty$, $t \to \frac{\infty}{x} = \infty$ since $x > 0$ as the bounds of x are from 0 to ∞. This is why the step mentioned earlier was so important.

(The bounds can be changed like this because the of the fundamental theorem of calculus. If $\int_{x=a}^{b} f(x)dx = F(a) - F(b)$. If the variable changes from x to something else, I need to find out what those variables are when $x = a$ and $x = b$ and change the old bounds to the new ones. In this case the new bounds are the same as the new ones.)

As mentioned before, the order in which these integrals are done does not matter so

$$I^2 = 4\int_{t=0}^{\infty}\int_{x=0}^{\infty} xe^{-x^2(1+t^2)}dx\,dt$$

$$\int_{0}^{\infty} xe^{-x^2(1+t^2)}dx$$

Solving this the same way as before, remembering that t is a constant with respect to x:

$$Let\ u = -x^2(1+t^2)$$

$$du = -2x(1+t^2)dx$$

$$dx = \frac{1}{-2x(1+t^2)}$$

$$\int_{x=0}^{\infty} \frac{xe^u}{-2x(1+t^2)}du$$

$$= \frac{1}{-2(1+t^2)}\int_{0}^{\infty} e^u du$$

$$= \frac{1}{-2(1+t^2)}[e^u]_{x=00}^{\infty}$$

$$= \frac{1}{-2(1+t^2)}\left[e^{-x^2(1+t^2)}\right]_0^{\infty}$$

$$t^2 > 0 \; so \; 1 + t^2 > 0 \; so - x^2(1 + t^2) < 0$$

$$so \; as \; x \to \infty, e^{-x^2(1+t^2)} \to 0$$

$$e^{-(0)^2(1+t^2)} = e^0 = 1$$

$$\frac{1}{-2(1 + t^2)}\left[e^{-x^2(1+t^2)}\right]_0^\infty$$

$$= \frac{1}{-2(1 + t^2)}(0 - 1)$$

$$= \frac{1}{2(1 + t^2)}$$

$$I^2 = 4\int_{t=0}^{\infty}\int_{x=0}^{\infty} xe^{-x^2(1+t^2)}dx\,dt$$

$$I^2 = 4\int_{t=0}^{\infty}\frac{1}{2(1 + t^2)}dt$$

$$I^2 = 2\int_{t=0}^{\infty}\frac{1}{1 + t^2}dt$$

Since we know that $\frac{d}{dx}\arctan x = \frac{1}{1+t^2}$

$$\int \frac{1}{1 + t^2}dt = \arctan t$$

$$I^2 = -2[\arctan t]_0^\infty$$

$$I^2 = 2\left[\lim_{n\to\infty}\arctan(n) - \arctan(0)\right]$$

$$\tan(0) = 0 \; so \; \arctan(0) = 0$$

$y = \tan x$ has a vertical asymptote at $x = \frac{\pi}{2}$. This means $y = \arctan x$ has a horizontal asymptote at $y = \frac{\pi}{2}$. This means that as $x \to \infty$, $\arctan x \to \frac{\pi}{2}$.

$$I^2 = 2\left(\frac{\pi}{2} - 0\right)$$

$$I^2 = \pi$$

$$I = \pm\sqrt{\pi}$$

As mentioned before, $I > 0$ so $I = \sqrt{\pi}$

$$\int_0^\infty e^{-x^2}dx = \sqrt{\pi}$$

Matrices and Linear Transformations

Transformation is essentially another word for function. Functions which are a used to take a number, sometimes multiple numbers, as an input and produce a single number as an output. This chapter will focus on transformations which take a vector as an input and produce a different vector as an output.

If, in this context a transformation takes a vector as an input and produces a vector as an output, a linear transformation is any transformation which fulfils two criteria:

- The origin must remain at the origin (the transformation of $\begin{pmatrix}0\\0\end{pmatrix}$ is $\begin{pmatrix}0\\0\end{pmatrix}$).
- When the transformation is applied all points on the grid (treating points as position vectors) al grid lines remain straight, parallel, and evenly spaced, in other words if the grid could be seen as a grid of parallelograms (e.g., squares). (This is assuming that the grid lines were straight, parallel, and evenly spaced before the transformation was applied.)

The above criteria may seem somewhat arbitrary, but their importance will become clear later.

First, I will discuss a new way in which vectors can be represented.

If I define the following two vectors:

$$\boldsymbol{i} = \begin{pmatrix}1\\0\end{pmatrix}$$

$$\boldsymbol{j} = \begin{pmatrix}0\\1\end{pmatrix}$$

(These are sometimes written as $\hat{\boldsymbol{\imath}}$ and $\hat{\boldsymbol{\jmath}}$).

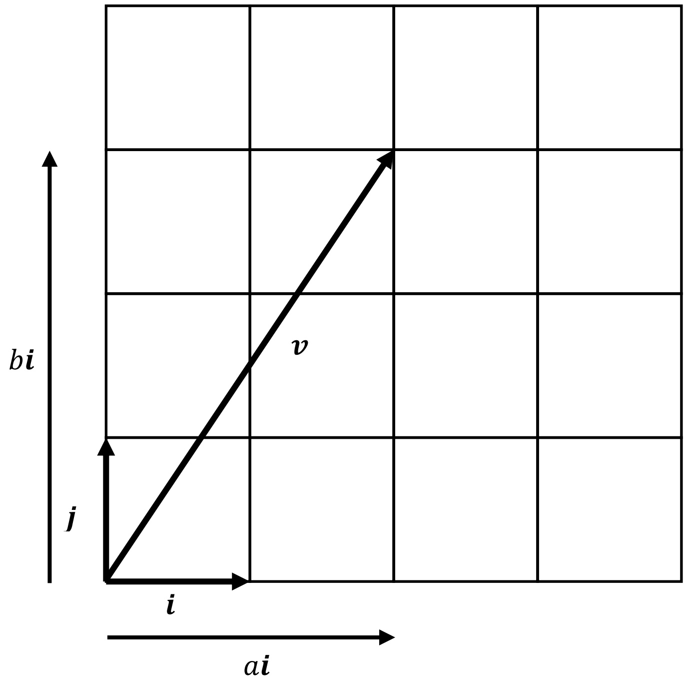
Then I can write any vector in terms of $\boldsymbol{i}$ and $\boldsymbol{j}$. For example, the vector $\boldsymbol{v} = \begin{pmatrix}a\\b\end{pmatrix}$ means a units to the right and b units up. This means that $\boldsymbol{v} = a\boldsymbol{i} + b\boldsymbol{j}$. I will call some transformation $\boldsymbol{M}$, and the vector obtained from applying transformation of $\boldsymbol{M}$ on some vector $\boldsymbol{v}$ I will call $\boldsymbol{Mv}$. This next part is where the criteria for a linear transformation become important.

$$Let\ \boldsymbol{v} = a\boldsymbol{i} + b\boldsymbol{j}$$

$$Let\ \boldsymbol{M}\ be\ some\ linear\ transformation$$

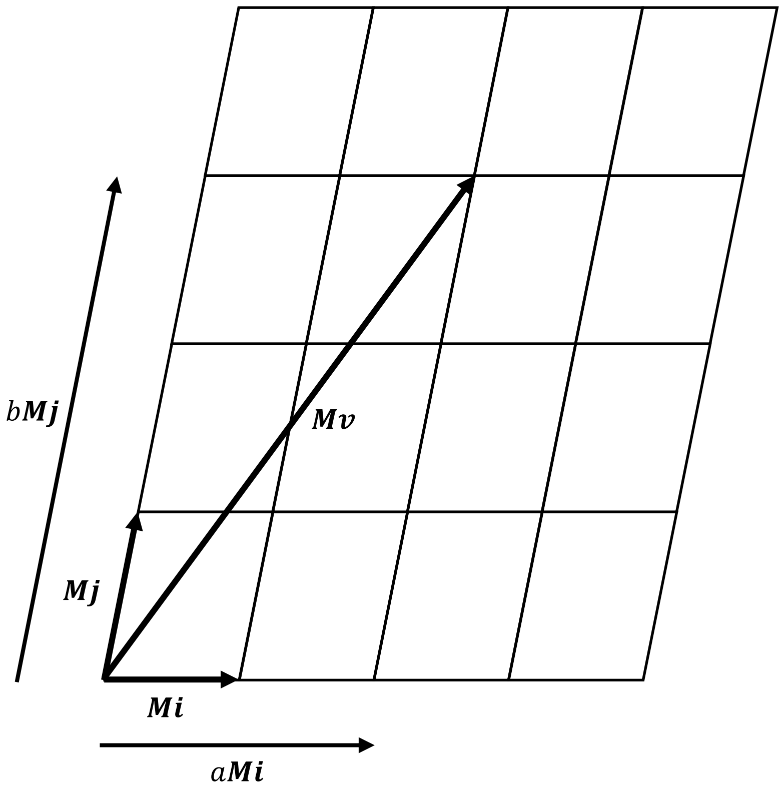
Since $\boldsymbol{M}$ is linear, and before the transformation is applied we are using a square grid, when $\boldsymbol{M}$ is applied to every point on the said grid, the grid remains as parallelograms. The reason why this is important is portrayed in the following diagrams.

Before transformation:



As shown in this diagram, $\boldsymbol{v} = a\boldsymbol{i} + b\boldsymbol{j}$.

After transformation $\boldsymbol{M}$:



This diagram shows how the unit vectors $\boldsymbol{i}$ and $\boldsymbol{j}$ have been transformed by $\boldsymbol{M}$ as well as how $\boldsymbol{v}$ has been transformed by $\boldsymbol{M}$. The diagram also shows that $\boldsymbol{Mv} = a\boldsymbol{Mi} + b\boldsymbol{Mj}$. This only works because the grid lines are still evenly spaced and parallel and because the origin is unmoved. This is why those requirements for a transformation to be considered linear were such.

Since $\boldsymbol{v} = \begin{pmatrix} a \\ b \end{pmatrix}$, $\boldsymbol{i} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\boldsymbol{j} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, the above equation can be written as:

$$\boldsymbol{M}\begin{pmatrix} a \\ b \end{pmatrix} = a\boldsymbol{M}\begin{pmatrix} 1 \\ 0 \end{pmatrix} + b\boldsymbol{M}\begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

remember, $\boldsymbol{M}$ is not a constant so cannot be multiplied into the vectors as such. What the above equation means is that if I know what $\boldsymbol{M}\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\boldsymbol{M}\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ are, I can work out $\boldsymbol{Mv}$. In other words, to fully describe some linear transformation $\boldsymbol{M}$ on some arbitrary vector $\boldsymbol{v}$ I need only find out what that transformation does to the unit vector $\boldsymbol{i}$ and $\boldsymbol{j}$.

For example, let's say I have worked out that $\boldsymbol{Mi} = \begin{pmatrix} a \\ b \end{pmatrix}$ and $\boldsymbol{Mj} = \begin{pmatrix} c \\ d \end{pmatrix}$. I can now fully describe the transformation $\boldsymbol{M}$ for any arbitrary vector $\boldsymbol{v} = \begin{pmatrix} p \\ q \end{pmatrix}$.

$$\boldsymbol{Mv} = p\boldsymbol{M}\begin{pmatrix} 1 \\ 0 \end{pmatrix} + q\boldsymbol{M}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = p\begin{pmatrix} a \\ b \end{pmatrix} + q\begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} pa \\ pb \end{pmatrix} + \begin{pmatrix} qc \\ qd \end{pmatrix} = \begin{pmatrix} pa + qc \\ pb + qd \end{pmatrix}$$

This means the transformation $\boldsymbol{M}$ of some arbitrary vector $\begin{pmatrix} p \\ q \end{pmatrix}$ can be fully described by four numbers a, c, b, d. As they are written in this way: $\begin{pmatrix} pa + qc \\ pb + qd \end{pmatrix}$ where p and q relate to the vector and a, b, c and d relate to the transformation, we can write the transformation using the following notation:

$$\boldsymbol{M} = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

This mathematical object is called a "matrix" (plural matrices) and it fully describes a transformation.

It is very important to understand at this stage that if $\boldsymbol{M}$ transforms $\boldsymbol{i}$ into $\begin{pmatrix} a \\ b \end{pmatrix}$ and transforms $\boldsymbol{j}$ into $\begin{pmatrix} c \\ d \end{pmatrix}$ then $\boldsymbol{M} = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$.

Matrix-vector multiplication is defined as the operation which produces the vector obtained by applying the transformation of the matrix on the given vector. This means that if $\boldsymbol{M} = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$ and $\boldsymbol{v} = \begin{pmatrix} p \\ q \end{pmatrix}$ then $\boldsymbol{Mv} =$

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix}\begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} pa + qc \\ pb + qd \end{pmatrix}$$

This is where the seemingly arbitrary (as the topic is usually presented without this back context) definition of matrix-vector multiplication comes from.

Multiple different transformations can be completed in succession. I will call my transformations $\boldsymbol{A}$ and $\boldsymbol{B}$ and my vector will be called $\boldsymbol{v}$. I will define them:

$$\boldsymbol{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

$$\boldsymbol{B} = \begin{pmatrix} e & f \\ g & h \end{pmatrix}$$

If transformation $\boldsymbol{B}$ is applied first followed by $\boldsymbol{A}$ then I must first find $\boldsymbol{Bv}$ and multiply this by $\boldsymbol{A}$. This can be written as $\boldsymbol{ABv}$. Given that, it would be useful to define matrix-matrix multiplication such that the transformed vector could be obtained by either multiplying $\boldsymbol{B}$ by $\boldsymbol{v}$ and then multiplying $\boldsymbol{A}$ by that or by multiplying $\boldsymbol{A}$ by $\boldsymbol{B}$ and then multiplying that by $\boldsymbol{v}$.

Recall before when I said that the only thing you need to fully describe a linear transformation (and hence a matrix) is to know what it does to the unit vectors. I know that matrix $\boldsymbol{B}$ turns $\boldsymbol{i}$ and $\boldsymbol{j}$ into $\begin{pmatrix} e \\ g \end{pmatrix}$ and $\begin{pmatrix} f \\ h \end{pmatrix}$. This means that, to find the matrix $\boldsymbol{AB}$ which does the same thing as $\boldsymbol{B}$ followed by $\boldsymbol{A}$ I need only apply $\boldsymbol{A}$ to these new unit vectors.

$$\boldsymbol{A}\begin{pmatrix} e \\ g \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} e \\ g \end{pmatrix} = \begin{pmatrix} ae + bg \\ ce + dg \end{pmatrix}$$

$$\boldsymbol{A}\begin{pmatrix} f \\ h \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} f \\ h \end{pmatrix} = \begin{pmatrix} af + bh \\ cf + dh \end{pmatrix}$$

This means that the full transformation $\boldsymbol{AB}$ turns $\boldsymbol{i}$ into $\begin{pmatrix} ae + bg \\ ce + dg \end{pmatrix}$ and turns $\boldsymbol{j}$ into $\begin{pmatrix} af + bh \\ cf + dh \end{pmatrix}$.

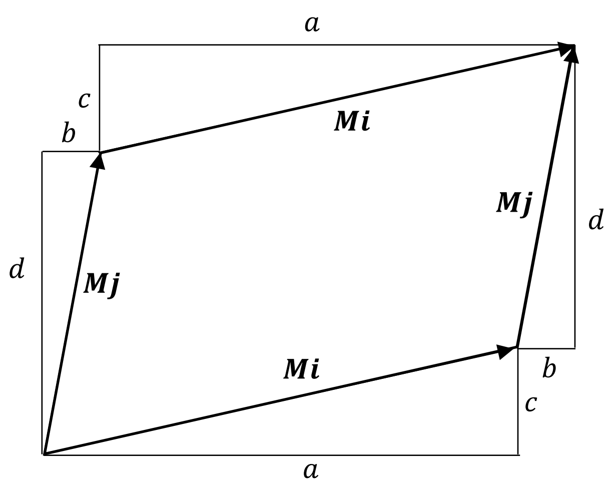
This means that $\boldsymbol{AB}$ can be described as the matrix: $\begin{pmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{pmatrix}$ meaning matrix-matrix multiplication should be defined as follows:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{pmatrix}$$

Which is where the strange (without this context which is typically not provided) definition of matrix-matrix multiplication comes from.

When a linear transformation is applied to all point (more precisely their position vectors) then regions may change their size and shape and therefore may change their area. Since grid lines remain parallel and evenly spaced, each parallelogram on the grid has the same area. These parallelograms could be drawn arbitrarily small such that any region could be approximated using these small parallelograms. This demonstrates that the factor by which the area changes in one region is indicative of how the area changes elsewhere. The simplest area change to find would be that of the unit square. If we draw the grid as a grid of unit squares before the transformation, then each square will have an area of 1. After the transformation, the area of one of the grid parallelograms divided by the area of the original grid unit square will give the scale factor change in area. That is to say the scale factor change in area equals the area of this parallelogram.

For some linear transformation $\boldsymbol{M} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ the unit vectors $\boldsymbol{i}$ and $\boldsymbol{j}$ have been of course transformed into $\begin{pmatrix} a \\ c \end{pmatrix}$ and $\begin{pmatrix} b \\ d \end{pmatrix}$ respectively. This can be used to find the area of on of the grid parallelograms.



The area of the parallelogram can be found by finding the area of the encompassing rectangles (side length $a+b$ and $c+d$) and then subtracting the area of the smaller triangles and rectangles.

$$Area = (a+b)(c+d) - \left(\frac{ac}{2} + \frac{bd}{2} + bc + \frac{ac}{2} + \frac{bd}{2} + bc\right)$$

$$Area = ac + ad + bc + bd - (ac + bd + 2bc)$$

$$Area = ac - ac + bd - bd + bc - 2bc + ad$$

$$Area = ad - bc$$

So, the scale factor describing how areas change with the transformation $\boldsymbol{M} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is given by $ad - bc$. This is called the determinant of $\boldsymbol{M}$ sometimes written as $\det \boldsymbol{M}$ or as $|\boldsymbol{M}|$. There is much more to matrices, including 3×3 matrices for 3 dimensional transformations as well as non-square matrices, but this is all I will cover here.

Laplace transforms, differential equations, and the Gamma Function

A power series is a series of the form $\sum f(n)x^n$, for example the Maclaurin series' for e^x, $\sin(x)$ and $\cos(x)$.

$$\sum_{n=0}^{\infty} \frac{1}{n!} x^n = e^x$$

In this case $f(n) = \frac{1}{n!}$. This could be thought of as a type of transformation, taking a function of n as an input (in this case $\frac{1}{n!}$) and producing a different function of x as an output (in this case e^x). We could define a transformation:

$$transformation\ of\ f(n) = \sum_{n=0}^{\infty} f(n)x^n$$

Where $\sum_{n=0}^{\infty} f(n)x^n$ will be some function of x.

In this case we are using discrete summation, but what would happen if we defined a transformation which did something similar but using continuous summation instead, that is, integration?

I will use t instead of n as n is typically used to represent integers. I will then replace the sigma notion with an integral.

$$\int_{t=0}^{\infty} f\,(t)x^t$$

This integral is uninteresting as it diverges due to the lack of a differential. Since the bounds have t going from 0 to ∞ I will use dt as the differential so that this becomes an integral with respect to t.

$$\int_{t=0}^{\infty} f\,(t)x^t\,dt$$

I could stop here, but this integral could be cleaned up a bit. Typically, when integrating it is convenient to have the dummy variable (in this case t) in the exponent when the base is e. In this case I have x^t instead. This can be solved by writing x^t as $e^{\ln(x^t)} = e^{t \ln x}$

$$\int_{t=0}^{\infty} f\,(t)e^{t \ln x}\,dt$$

This is now better, though it is still rather ugly. The rest of these changes are mainly for aesthetic and convenience as opposed to function.

$t \geq 0$ because the bounds of the integral take t from 0 to ∞. x is some constant with respect to t. If $x > 1$ then $\ln x > 0$ so $t \ln x \geq 0$ so the integral is likely to diverge. If $x \leq 1$ then $\ln x \leq 0$ so $t \ln x \leq 0$ and so the integral is more likely to be convergent. Assuming then that $x \leq 1$, and that $\ln x \leq 0$ I will let $s = \ln x$. This means the integral can be written as:

$$\int_{t=0}^{\infty} f\,(t)e^{st}\,dt$$

which will be a function of s where s is a function of x. The above will only be valid for $s \leq 0$ since $\ln x \leq 0$ meaning it might be nicer to define s as $s = -\ln x$. This way s will always be positive which may be easier to work with. This gives us the final form:

$$\int_0^\infty f(t)e^{-st}dt$$

This is called the Laplace transform of $f(t)$ and was originally created by mathematician Laplace, who we have encountered before. The notation for this is:

$$\mathcal{L}\{f(t)\} = \int_0^\infty f(t)e^{-st}dt$$

This transform has applications in solving differential equations which I will get to later, but one of my favourite uses of this transform is deriving the Gamma and Pi Functions.

I will start by giving an example of a Laplace transform.

$$\mathcal{L}\{e^t\} = \int_0^\infty e^t e^{-st}dt$$

$$= \int_0^\infty e^{t-st}dt$$

$$= \int_0^\infty e^{t(1-s)}dt$$

$$= \left[\frac{1}{1-s}e^{t(1-s)}\right]_{t=0}^\infty$$

$$= \lim_{n\to\infty} \frac{1}{1-s}e^{n(1-s)} - \frac{1}{1-s}e^{0(1-s)}$$

$$= \lim_{n\to\infty} \frac{1}{1-s}e^{n(1-s)} - \frac{1}{1-s}$$

$$= \frac{1}{1-s}\left(\lim_{n\to\infty} e^{n(1-s)} - 1\right)$$

If $1 > s$ then $1 - s > 0$ so the limit diverges to infinity.

If $1 < s$ then $1 - s < 0$ so the limit converges to 0. $s > 1$ will be important in a moment. In this case:

$$\frac{1}{1-s}\left(\lim_{n\to\infty} e^{n(1-s)} - 1\right)$$

$$= \frac{1}{1-s}(0-1)$$

$$= \frac{-1}{1-s}$$

This looks similar to the infinite sum of a geometric series:

$$\sum_{n=0}^{\infty} ar^n = \frac{a}{1-r}$$

This is only true however when $|r| < 1$. In this case that is not true as we said that $s > 1$.

$$\frac{-1}{1-s}$$

$$= \frac{1}{s-1}$$

By dividing the top and bottom by s:

$$= \frac{\left(\frac{1}{s}\right)}{1-\frac{1}{s}}$$

In this case $\frac{1}{s} > 0$ because s is positive and, since $s > 1$, $1 > \frac{1}{s}$ or $0 < \frac{1}{s} < 1$ meaning $\left|\frac{1}{s}\right| < 1$ meaning this is the sum of a geometric series first term $\frac{1}{s}$ and common ratio $\frac{1}{s}$.

$$\frac{\left(\frac{1}{s}\right)}{1-\frac{1}{s}}$$

$$= \sum_{n=0}^{\infty} \frac{1}{s} \times \left(\frac{1}{s}\right)^{n}$$

$$= \sum_{n=0}^{\infty} \left(\frac{1}{s}\right)^{n+1}$$

$$= \left(\frac{1}{s}\right) + \left(\frac{1}{s}\right)^{2} + \left(\frac{1}{s}\right)^{3} + \left(\frac{1}{s}\right)^{4} + \cdots$$

So

$$\mathcal{L}\{e^{t}\} = \frac{1}{s-1} = \left(\frac{1}{s}\right) + \left(\frac{1}{s}\right)^{2} + \left(\frac{1}{s}\right)^{3} + \left(\frac{1}{s}\right)^{4} + \cdots$$

We know that e^{t} can be written as a series:

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

So, taking the Laplace transform of this series should yield the same series infinite series as for $\mathcal{L}\{e^{t}\}$.

$$\mathcal{L}\left\{\sum_{n=0}^{\infty} \frac{t^{n}}{n!}\right\}$$

$$= \int_{0}^{\infty} \sum_{n=0}^{\infty} \frac{t^{n}}{n!} \times e^{-st} dt$$

$$= \int_0^\infty \left(\frac{t^0}{0!} + \frac{t^1}{1!} + \frac{t^2}{2!} + \frac{t^3}{3!} + \cdots\right) e^{-st} dt$$

$$= \int_0^\infty \left(\frac{t^0}{0!} e^{-st} + \frac{t^1}{1!} e^{-st} + \frac{t^2}{2!} e^{-st} + \frac{t^3}{3!} e^{-st} + \cdots\right) dt$$

$$= \int_0^\infty \frac{t^0}{0!} e^{-st} dt + \int_0^\infty \frac{t^1}{1!} e^{-st} dt + \int_0^\infty \frac{t^2}{2!} e^{-st} dt + \int_0^\infty \frac{t^3}{3!} e^{-st} dt + \cdots$$

$$= \frac{1}{0!}\int_0^\infty t^0 e^{-st} dt + \frac{1}{1!}\int_0^\infty t^1 e^{-st} dt + \frac{1}{2!}\int_0^\infty t^2 e^{-st} dt + \frac{1}{3!}\int_0^\infty t^3 e^{-st} dt + \cdots$$

$$= \sum_{n=0}^{\infty} \frac{1}{n!} \int_0^\infty t^n e^{-st} dt$$

Perform the following substitution:

$$Let\ x = st$$

$$t = \frac{x}{s}$$

$$dx = sdt$$

$$dt = \frac{1}{s} dx$$

$$\sum_{n=0}^{\infty} \frac{1}{n!} \int_0^\infty t^n e^{-st} dt$$

$$= \sum_{n=0}^{\infty} \frac{1}{n!} \int_0^\infty \frac{\left(\frac{x}{s}\right)^n e^{-x}}{s} dx$$

$$= \sum_{n=0}^{\infty} \frac{1}{n!} \int_0^\infty \frac{\left(\frac{x^n e^{-x}}{s^n}\right)}{s} dx$$

$$= \sum_{n=0}^{\infty} \frac{1}{n!} \int_0^\infty \frac{x^n e^{-x}}{s^{n+1}} dx$$

$$= \sum_{n=0}^{\infty} \frac{1}{s^{n+1}} \frac{1}{n!} \int_0^\infty x^n e^{-x} dx$$

$$Let\ f(n) = \frac{1}{n!} \int_0^\infty x^n e^{-x} dx$$

Because $\int_0^\infty x^n e^{-x} dx$ is a function of n.

$$\sum_{n=0}^{\infty} \frac{1}{s^{n+1}} f(n)$$

$$= \left(\frac{1}{s}\right)f(0) + \left(\frac{1}{s^2}\right)f(1) + \left(\frac{1}{s^3}\right)f(2) + \left(\frac{1}{s^4}\right)f(3) + \cdots$$

$$\left(\frac{1}{s}\right)f(0) + \left(\frac{1}{s}\right)^2 f(1) + \left(\frac{1}{s}\right)^3 f(2) + \left(\frac{1}{s}\right)^4 f(3) + \cdots$$

$$so\ \mathcal{L}\left\{\sum_{n=0}^{\infty}\frac{t^n}{n!}\right\} = \left(\frac{1}{s}\right)f(0) + \left(\frac{1}{s}\right)^2 f(1) + \left(\frac{1}{s}\right)^3 f(2) + \left(\frac{1}{s}\right)^4 f(3) + \cdots$$

$$Since\ e^t = \sum_{n=0}^{\infty}\frac{t^n}{n!}$$

$$\mathcal{L}\{e^t\} = \mathcal{L}\left\{\sum_{n=0}^{\infty}\frac{t^n}{n!}\right\}$$

$$\left(\frac{1}{s}\right) + \left(\frac{1}{s}\right)^2 + \left(\frac{1}{s}\right)^3 + \left(\frac{1}{s}\right)^4 + \cdots = \left(\frac{1}{s}\right)f(0) + \left(\frac{1}{s}\right)^2 f(1) + \left(\frac{1}{s}\right)^3 f(2) + \left(\frac{1}{s}\right)^4 f(3) + \cdots$$

Equating coefficients:

$$f(0) = 1, f(1) = 1, f(2) = 1, f(3) = 1, \ldots$$

In general, $f(n) = 1$ where $n \in \mathbb{Z}^+$ (n is a positive integer).

$$f(n) = \frac{1}{n!}\int_0^{\infty} x^n e^{-x} dx$$

$$1 = \frac{1}{n!}\int_0^{\infty} x^n e^{-x} dx$$

$$n! = \int_0^{\infty} x^n e^{-x} dx$$

This integral could be used to extend the factorial function beyond the positive real numbers. This is not a "proof" that the function works for negative or non-integer factorials since in a previous step I said that $n \in \mathbb{Z}^+$. This could however be a strong candidate to extend the function.

This integral is known as the Pi function.

$$\Pi(x) = x! = \int_0^{\infty} t^x e^{-t} dt$$

One main feature of the factorial function is that (by definition) $n! = n(n-1)!$ Also $0! = 1$. The latter is true as:

$$\Pi(0) = \int_0^{\infty} t^0 e^{-t} dt = \int_0^{\infty} e^{-t} dt = [-e^{-t}]_0^{\infty} = \left[\frac{-1}{e^t}\right]_0^{\infty} = (0) - (-1) = 1$$

$$so\ \Pi(0) = 1$$

For the former:

$$\Pi(x-1) = \int_0^\infty t^{x-1}e^{-t}dt$$

Integration by parts:

$$\begin{array}{ccc} & D & I \\ + & e^{-t} & t^{x-1} \\ - & -e^{-t} & \frac{1}{x}t^x \end{array}$$

$$\int_0^\infty t^{x-1}e^{-t}dt = \left[\frac{e^{-t}t^x}{x} + \int \frac{e^{-t}t^x}{x}dt\right]_{t=0}^\infty$$

$$\Pi(x-1) = \lim_{n\to\infty} \frac{e^{-n}n^x}{x} - \frac{e^{-0}0^x}{x} + \int_0^\infty \frac{e^{-t}t^x}{x}dt$$

$$\Pi(x-1) = \lim_{n\to\infty} \frac{e^{-n}n^x}{x} + \frac{1}{x}\int_0^\infty e^{-t}t^x dt$$

$$x\Pi(x-1) = \lim_{n\to\infty} e^{-n}n^x + \int_0^\infty e^{-t}t^x dt$$

$$x\Pi(x-1) = \lim_{n\to\infty} \frac{n^x}{e^n} + \Pi(x)$$

As $n \to \infty$, e^x grows more quickly than n^x so $\frac{n^x}{e^x} \to 0$

$$\Pi(x) = x\Pi(x-1)$$

This means that the Pi function would be a sensible way to extend the factorial function. There is another more famous function, called the Gamma function defined as:

$$\Gamma(x) = \Pi(x-1)$$

$$\text{so } \Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$$

$$and\ \Pi(x) = x\Gamma(x)$$

The Pi function can be used to define the factorial for non-positive integer values. For example, $\left(\frac{1}{2}\right)!$

$$\left(\frac{1}{2}\right)! = \Pi\left(\frac{1}{2}\right) = \int_0^\infty t^{\frac{1}{2}}e^{-t}dt = \int_0^\infty \sqrt{t}e^{-t}dt$$

$$Let\ u = \sqrt{t}$$

$$u^2 = t$$

$$2u\ du = dt$$

$$du = \frac{1}{2u}dt$$

When $t = 0, u = \sqrt{0} = 0$

As $t \to \infty, u = \sqrt{t} \to \infty$

$$\int_0^\infty ue^{-u^2} du$$

$$= \frac{1}{2}\int_{-\infty}^\infty ue^{-u^2} du$$

This is just half of the Gaussian integral which we have already seen.

$$= \frac{\sqrt{\pi}}{2}$$

This can be used to find the factorials of $-\frac{1}{2}, -\frac{3}{2}$ etc., as well as $\frac{3}{2}, \frac{5}{2}, \frac{7}{2}$ etc.,

Using $\Pi(x) = x\Pi(x-1)$

$$\Pi\left(\frac{1}{2}\right) = \frac{1}{2}\Pi\left(\frac{1}{2} - 1\right)$$

$$\Pi\left(\frac{1}{2}\right) = \frac{1}{2}\Pi\left(-\frac{1}{2}\right)$$

$$\Pi\left(-\frac{1}{2}\right) = 2\Pi\left(\frac{1}{2}\right)$$

$$\Pi\left(-\frac{1}{2}\right) = 2\frac{\sqrt{\pi}}{2}$$

$$\left(-\frac{1}{2}\right)! = \sqrt{\pi}$$

The other use of Laplace transforms if solving differential equations. Equations typically have constants where you need to find the numerical value of one of these constants for the equation to hold true. E.g., for which value of x does $5x + 12 = x^2 - 4x$? A differential equation instead has derivatives, and you need to find the original function. E.g., $\frac{d^2y}{dx^2} - 10\frac{dy}{dx} + 9y = 5t$ where $y(0) = -1$ and $y'(0) = 2$

Before being able to do this I must first cover a few more Laplace transform examples. First, the Laplace transform of the derivative of some generic function $f(t)$.

$$\mathcal{L}\{f'(t)\} = \int_0^\infty f'(t)e^{-st} dt$$

Using integration by parts:

$$\begin{array}{ccc} & D & I \\ + & e^{-st} & f'(t) \\ - & -se^{-st} & f(t) \end{array}$$

$$\mathcal{L}\{f'(t)\} = \left[e^{-st}f(t) - \int -se^{-st}f(t)dt\right]_0^\infty$$

$$\mathcal{L}\{f'(t)\} = \left[e^{-st}f(t) + s\int e^{-st}f(t)dt\right]_0^\infty$$

$$\mathcal{L}\{f'(t)\} = \lim_{n\to\infty} e^{-sn}f(n) - e^{-sn}f(0) + s\int_0^\infty f(t)e^{-st}dt$$

$$as\ n \to \infty, e^{-sn} \to 0$$

$$\mathcal{L}\{f'(t)\} = -e^{(0)s}f(0) + s\int_0^\infty f(t)e^{-st}dt$$

$$\mathcal{L}\{f'(t)\} = s\int_0^\infty f(t)e^{-st}dt - f(0)$$

Shorthand: $Let\ F(s) = \mathcal{L}(f(t))$.

$$\mathcal{L}\{f'(t)\} = sF(s) - f(0)$$

Next, the Laplace transform of the second derivative:

$$I\ Know: \mathcal{L}\left\{\left(f(t)\right)'\right\} = s\mathcal{L}\{f(t)\} - f(0)$$

$$\mathcal{L}\{f''(t)\}$$

$$= \mathcal{L}\{(f'(t))'\}$$

$$= s\mathcal{L}\{f'(t)\} - f'(0)$$

$$= s\left(s\mathcal{L}\{f(t)\} - f(0)\right) - f'(0)$$

$$= s^2\mathcal{L}\{f(t)\} - sf(0) - f'(0)$$

$$= s^2F(s) - sf(0) - f'(0)$$

One more of these:

$$\mathcal{L}\{f'''(t)\}$$

$$= \mathcal{L}\{(f''(t))'\}$$

$$= s\mathcal{L}\{f''(t)\} - f''(0)$$

$$I\ know\ that\ \mathcal{L}\{f''(t)\} = s^2F(s) - sf(0) - f'(0)$$

$$So\ s\mathcal{L}\left\{f''^{(t)}\right\} - f''^{(0)}$$

$$= s\left(s^2F(s) - sf(0) - f'(0)\right) - f''(0)$$

$$= s^3F(s) - s^2f(0) - sf'(0) - f''(0)$$

In general:

$$\mathcal{L}\{f^{(n)}(t)\} = s^nF(s) - s^{n-1}f(0) - s^{n-2}f'(0) - s^{n-3}f''(0) - \cdots - f^{(n-1)}(0)$$

$$or$$

$$\mathcal{L}\{f^{(n)}(t)\} = s^nF(s) - \sum_{i=1}^{n} s^{n-i}f^{(i-1)}(0)$$

There are some more things I must discuss before tackling the example question.

$$k\mathcal{L}\{f(t)\}$$

$$= k\int_0^\infty f(t)e^{-st}dt$$

$$= \int_0^\infty \big(kf(t)\big)e^{-st}dt$$

$$= \mathcal{L}\{kf(t)\}$$

So, constants can be brought inside the Laplace transform.

$$\mathcal{L}\{f(t) + g(t)\}$$

$$= \int_0^\infty \big(f(t) + g(t)\big)e^{-st}dt$$

$$= \int_0^\infty f(t)e^{-st} + g(t)e^{-st}dt$$

$$= \int_0^\infty f(t)e^{-st}dt + \int_0^\infty g(t)e^{-st}dt$$

$$\mathcal{L}\{f(t)\} + \mathcal{L}\{g(t)\}$$

So, Laplace transforms can be split up in this manner.

$$\mathcal{L}\{k\} = \int_0^\infty ke^{-st}dt$$

$$= k\int_0^\infty e^{-st}dt$$

$$= k\left[-\frac{1}{s}e^{-st}\right]_{t=0}^{\infty}$$

$$= k\left(0 - \left(-\frac{1}{s}\right)\right)$$

$$= k\left(\frac{1}{s}\right)$$

$$= \frac{k}{s}$$

So, the Laplace transform of a constant k, is $\frac{k}{s}$.

$$\mathcal{L}\{t\} = \int_0^\infty te^{-st}dt$$

Integration by parts:

$$\begin{array}{ccc} & D & I \\ + & t & e^{-st} \\ - & 1 & -\frac{1}{s}e^{-st} \end{array}$$

$$\mathcal{L}\{t\} = \left[-\frac{t}{s}e^{-st} + \int \frac{1}{s}e^{-st}dt\right]_{t=0}^{\infty}$$

$$\mathcal{L}\{t\} = 0 - (0) + \frac{1}{s}\int_0^\infty e^{-st}$$

$$\mathcal{L}\{t\} = \frac{1}{s}\left[-\frac{1}{s}e^{-st}\right]_0^\infty$$

$$\mathcal{L}\{t\} = \frac{1}{s}\left(0 - \left(-\frac{1}{s}\right)\right)$$

$$\mathcal{L}\{t\} = \frac{1}{s}\left(\frac{1}{s}\right)$$

$$\mathcal{L}\{t\} = \frac{1}{s^2}$$

$$So\ \mathcal{L}\{t\} = \frac{1}{s^2}$$

$$\mathcal{L}\{e^{at}\} = \int_0^\infty e^{at}e^{-st}dt$$

$$= \int_0^\infty e^{at-st}dt$$

$$= \int_0^\infty e^{t(a-s)}dt$$

$$= \left[\frac{1}{a-s}e^{t(a-s)}\right]_{t=0}^{\infty}$$

$$when\ a < s, a - s < 0\ then$$

$$\left[\frac{1}{a-s}e^{t(a-s)}\right]_{t=0}^{\infty}$$

$$=\left(\frac{0}{a-s}\right)-\left(\frac{1}{a-s}\right)$$

$$=\frac{1}{a-s}$$

$$=\frac{1}{s-a}$$

I will use this knowledge to answer the previous example.

$\frac{d^2y}{dx^2}-10\frac{dy}{dx}+9y=5t$ where $y(0)=-1$ and $y'(0)=2$

Writing the above as:

$$y''-10y'+9y=5t$$

$$y(0)=-1$$

$$y'(0)=2$$

$$Let\ \mathcal{L}\{y\}=Y$$

First, Laplace transform both sides:

$$\mathcal{L}\{y''-10y'+9y\}=\mathcal{L}\{5t\}$$

$$\mathcal{L}\{y''\}+\mathcal{L}\{-10y'\}+\mathcal{L}\{9y\}=\mathcal{L}\{5t\}$$

$$\mathcal{L}\{y''\}-10\mathcal{L}\{y'\}+9\mathcal{L}\{y\}=5\mathcal{L}\{t\}$$

$$s^2Y-sy(0)-y'(0)-10\big(sY-y(0)\big)+9Y=5\left(\frac{1}{s^2}\right)$$

$$s^2Y-s(-1)-2-10\big(sY-(-1)\big)+9Y=\frac{5}{s^2}$$

$$s^2Y+s-2-10sY-10+9Y=\frac{5}{s^2}$$

$$s^2Y+s-12-10sY+9Y=\frac{5}{s^2}$$

$$s^2Y-10sY+9Y=\frac{5}{s^2}-s+12$$

$$Y(s^2-10s+9)=\frac{5-s^3+12s^2}{s^2}$$

$$Y=\frac{5-s^3+12s^2}{s^2(s^2-10s+9)}$$

$$\mathcal{L}^{-1}\{Y\} = \mathcal{L}^{-1}\left\{\frac{-s^3 + 12s^2 + 5}{s^2(s^2 - 10s + 9)}\right\}$$

$$y = \mathcal{L}^{-1}\left\{\frac{-s^3 + 12s^2 + 5}{s^2(s^2 - 10s + 9)}\right\}$$

I now need to work out the inverse Laplace transform of $\frac{5-s^3+12s^2}{s^2(s^2-10s+9)}$. I will start by writing it in a different form by using a technique called partial fractions.

$$\frac{-s^3 + 12s^2 + 5}{s^2(s^2 - 10s + 9)} = \frac{-s^3 + 12s^2 + 5}{s^2(s-9)(s-1)}$$

$$Let\ \frac{-s^3 + 12s^2 + 5}{s^2(s-9)(s-1)} \equiv \frac{A}{s} + \frac{B}{s^2} + \frac{C}{s-9} + \frac{D}{s-1}$$

$$where\ A, B, C\ and\ D\ are\ constants\ to\ be\ found.$$

The $\equiv$ symbol has been used as I want the above to be true for all s.

Multiply both sides by $s^2(s-9)(s-1)$

$$-s^3 + 12s^2 + 5 \equiv \frac{s^2(s-9)(s-1)A}{s} + \frac{s^2(s-9)(s-1)B}{s^2} + \frac{s^2(s-9)(s-1)C}{s-9} + \frac{s^2(s-9)(s-1)D}{s-1}$$

$$-s^3 + 12s^2 + 5 \equiv s(s-9)(s-1)A + (s-9)(s-1)B + s^2(s-1)C + s^2(s-9)D$$

The above is true for all s so I can let s be any value in order to find the values of A, B, C, and D.

$$Let\ s = 1$$

$$1 - 1 = 0$$

So, all terms with coefficients of $(s-1)$ will become 0.

$$-(1)^3 + 12(1)^2 + 5 = (1)^2(1-9)D$$

$$16 = -8D$$

$$D = -2$$

$$Let\ s = 9$$

$$9 - 9 = 0$$

So, all terms with coefficients of $(s-9)$ will become 0.

$$-(9)^3 + 12(9)^2 + 5 = (9^2)(9-1)C$$

$$248 = 648C$$

$$C = \frac{31}{81}$$

$$Let\ s = 0$$

So, all terms with coefficients of s will become 0.

$$-(0)^3 + 12(0)^2 + 5 = (0-9)(0-1)B$$

$$5 = 9B$$

$$B = \frac{5}{9}$$

I will choose some arbitrary value to find the last constant.

$$Let\ s = 3$$

$$-(3)^3 + 12(3)^2 + 5 = 3(3-9)(3-1)A + (3-9)(3-1)B + 3^2(3-1)C + 3^2(3-9)D$$

$$86 = -36A - 12B + 18C - 54D$$

Using the known values of $B = \frac{5}{9}, C = \frac{31}{81}$ & $D = -2$:

$$86 = -36A - \frac{20}{3} + \frac{62}{9} + 108$$

$$36A = \frac{200}{9}$$

$$A = \frac{50}{81}$$

Putting these constants back in:

$$\frac{-s^3 + 12s^2 + 5}{s^2(s-9)(s-1)} \equiv \frac{A}{s} + \frac{B}{s^2} + \frac{C}{s-9} + \frac{D}{s-1}$$

$$\frac{-s^3 + 12s^2 + 5}{s^2(s-9)(s-1)} \equiv \frac{\left(\frac{50}{81}\right)}{s} + \frac{\left(\frac{5}{9}\right)}{s^2} + \frac{\left(\frac{31}{81}\right)}{s-9} + \frac{-2}{s-1}$$

$$y = \mathcal{L}^{-1}\left\{\frac{\left(\frac{50}{81}\right)}{s} + \frac{\left(\frac{5}{9}\right)}{s^2} + \frac{\left(\frac{31}{81}\right)}{s-9} + \frac{-2}{s-1}\right\}$$

$$y = \frac{50}{81}\mathcal{L}^{-1}\left\{\frac{1}{s}\right\} + \frac{5}{9}\mathcal{L}^{-1}\left\{\frac{1}{s^2}\right\} + \frac{31}{81}\mathcal{L}^{-1}\left\{\frac{1}{s-9}\right\} - 2\mathcal{L}^{-1}\left\{\frac{1}{s-1}\right\}$$

$$I\ know: \mathcal{L}\{1\} = \frac{1}{s}$$

$$\mathcal{L}\{t\} = \frac{1}{s^2}$$

$$\mathcal{L}\{e^{at}\} = \frac{1}{s-a}$$

$$and\ \mathcal{L}\{e^t\} = \frac{1}{s-1}$$

This means that $\mathcal{L}^{-1}\left\{\frac{1}{s}\right\} = 1, \mathcal{L}^{-1}\left\{\frac{1}{s^2}\right\} = t, \mathcal{L}^{-1}\left\{\frac{1}{s-a}\right\} = e^{at}\ and\ \mathcal{L}^{-1}\left\{\frac{1}{s-1}\right\} = e^t$

Using these:

$$y = \frac{50}{81} + \frac{5}{9}t + \frac{31}{81}e^{9t} - 2e^t$$

Conclusion

This 160-page document has hardly scratched the surface of what the world of mathematics has to offer. With entire fields of mathematics which I have not so much as mentioned (probability, combinatorics, topology just to name a few), as well as many mathematical techniques (e.g., Feynman integration, Weierstrass substitution using matrices to solve linear simultaneous equations). This project has only dipped its toe into each of the topics discussed here, and there is far more to learn, know and understand. Go forth and do so.