Notes on Mathematics

for TMUA Paper 1 and ESAT Mathematics 2

Updated June 2025

Contents

INTRODUCTION

This is worth reading before you use this guide.

We have three aims in writing these notes:

First, we want to set out what we expect you to know for the TMUA and for ESAT Mathematics 2 [we just say TMUA/ESAT in these notes but when we do, we mean TMUA and ESAT Mathematics 2; and recall that Paper 2 of the TMUA requires you to know all of the paper 1 specification too]. We do this by basing these notes on the relevant part of the specification and adding comments throughout when we wish to clarify how to interpret aspects of the specification.

Second, we want to encourage you to think deeply and carefully about mathematics and to develop a good understanding of the topics in the specification. To help with this, we have added a lot of discussion and examples as well as some exercises throughout the notes. We recommend you work through each exercise.

Third, we want to make sure that all candidates have access to a free resource to help them prepare for the TMUA and the ESAT.

How to use this guide

You do not need to work through all this guide as you will find that you know many of the topics in the specification very well already. Use this guide as a resource to help you clarify and review topics that you are less familiar with. We have broken down our discussion to fit exactly with the specification to make things as simple to navigate as possible.

What this guide is not

This guide is not a comprehensive textbook: we do not cover every topic to the same level of detail, and we do not develop every topic from scratch. It is also not a substitute for sustained hard work and preparation. It is a resource to help you and to guide you in the right direction.

These notes are not intended to be a rigorous or perfect guide to the foundations of mathematics. Occasionally, we have simplified topics in a manner that might make a purist squirm a smidgen. We have often, but not always, indicated when we have been deliberately vague or simplified things.

Other resources

As well as this guide, there is the *Notes on Logic and Proof* which we produced for paper 2 of the TMUA. There are also many past papers on the UAT website that you can work through, and all the TMUA past papers have detailed solutions for you to look at as well.

The best way to tackle past papers is to use them under timed conditions as much as possible as that will give you a feel for the real TMUA or ESAT. The UAT website also has links to some practice papers in the on-screen format that will be used for the live papers – make sure you look at these to familiarise yourself with the on-screen style.

Who wrote this guide?

Both the *Notes on Logic and Proof* and this guide were written by the same team of mathematicians who developed, set and oversee the TMUA and Mathematics 1 and 2 in the ESAT.

Should I take a TMUA or ESAT course?

We do not recommend that you take a course and we do NOT endorse any courses. No one from the TMUA or ESAT development teams teaches on any courses. All the resources you need to prepare are available from the UAT website and are entirely free. If you spend the same amount of time studying by yourself or with friends, you will get at least the same benefit as going on a course, but you will save the cost.

TMUA and MATHEMATICS 2 of the ESAT

And finally, a note about Mathematics 2 in the ESAT and the TMUA. In developing both these assessments, we have kept the mathematical knowledge required as accessible and as straightforward as possible. We expect everyone taking the ESAT and TMUA to enter the exam equipped with the same level of mathematics knowledge. Both the TMUA and Mathematics 2 in the ESAT aim to test your ability to *use* your knowledge to answer problems that you are unlikely to have met before. We make sure that our questions do not require vast amounts of working to solve, and we make sure that there is no advantage if you have studied more advanced mathematics topics. Most of the work you will need to do when solving our questions will be in the thinking.

A final note

We have used boxes throughout these notes to help you navigate.

The relevant part of the specification is found in these sorts of boxes:

Specification

Examples in these sorts of boxes

Examples

And exercises [we have not given any answers to our exercises] in these sorts of boxes.

Exercises

We hope to be able to update and, if necessary, correct these notes now and again. Look at the date on the front page to see when these notes were last edited.

MM1. Algebra and functions

MM1.1

Laws of indices for all rational exponents.

Indices [or powers, or exponents, if you prefer] are really a mathematician's method for writing out certain ways of combining numbers without using vast quantities of ink.[1] They are a good example of what a well-chosen notation can do. A well-chosen notation aids thinking and makes calculations and manipulations easier than they might otherwise be. For the TMUA/ESAT you are expected to know all the basic rules of indices – both what the notation means and how to deal with the notation.

In this section, we will introduce the basic rules we expect you to know along with some informal notes to help you start to think about how the ideas fit together.

We start with the very basic idea of an index for a number a multiplied by itself a total of m times [that is, a appears m times in the expression] :

$$\overbrace{a \times a \times a \times \ldots \times a}^{a\ appears\ m\ times\ here}$$

We write this concisely as a^m

This basic idea allows us to work out how we might combine powers when we multiply: we can ask how we might write $a^m \times a^n$. If we write the whole expression out term-by-term and then use the rule that $a \times a \times a \times \ldots \times a$ m-times is written as a^m we arrive at the following:

$$a^m \times a^n = \underbrace{(a \times \ldots . \times a)}_{m\ times} \times \underbrace{(a \times a \times \ldots \times a)}_{n\ times} = \underbrace{a \times a \times a \ldots \times a}_{m+n\ times} = a^{m+n}$$

And so, we have our first rule for our notation – a rule that is really just the direct consequence of how we decided to write $a \times a \times a \times \ldots \times a$ in our notation:

$$RULE\ 1 \qquad a^m \times a^n \equiv a^{m+n}$$

We note that, for the moment, this rule applies when m and n are whole positive numbers. Later we will explain that the rule works for ALL real numbers m and n.

Next, we are going to extend this rule to "invent/derive" and motivate some other notation. We want to be able to use this rule when m and n are not integers and we also want to make sure that our notation is consistent – that is, we don't want to find

[1] Mathematicians like beauty, clarity, precision, elegance, and brevity; saving ink whilst maintaining these virtues is the ideal.

that we introduce definitions and rules that give us different answers depending on how we apply them.

The first thing we are going to decide is that RULE 1 works when a is positive and m and n are ANY rational number – there are good reasons for this decision which we shall talk about later. That means, for instance, we can apply the rule when m is $\frac{1}{3}$ and n is $-\frac{31}{17}$ and so on. But we do need to ask, what does $a^{\frac{1}{3}}$ mean, and what does the minus sign in $a^{-\frac{31}{17}}$ mean? We will motivate our answers using RULE 1 as this will ensure that our use of notation extended to fractional powers is consistent.

Let's start by trying to work out what we would like $a^{\frac{1}{3}}$ to mean. We can use RULE 1 extended to fractions to write:

$$a^{\frac{1}{3}}a^{\frac{1}{3}}a^{\frac{1}{3}} = a^{\left(\frac{1}{3}+\frac{1}{3}+\frac{1}{3}\right)} = a^1 = a$$

And this means that we must interpret $a^{\frac{1}{3}}$ as the cube-root of a [and note, we also used $a^1 = a$]

So, it should be "obvious" that we must interpret $a^{\frac{1}{n}}$ as the n^{th} root of a ; this gives us our second rule:

$$RULE\ 2 \qquad a^{\frac{1}{n}} = \sqrt[n]{a}$$

Next, we are going to look at what happens when we introduce a minus sign into the power. We are going to do this by exploring how RULE 1 might fit together with expression such as $a^3 \times a^{-2}$:

Using RULE 1 extended to negative numbers we obtain the following:

$$a^3 \times a^{-2} = a^{3+(-2)} = a^1 = a$$

And we ask ourselves what we need to multiply a^3 by to get a; and the answer is that we need to multiply a^3 by $\frac{1}{a^2}$ to get an answer of a. This suggests that we should interpret a^{-2} as being the same as $\frac{1}{a^2}$ and leads to our third rule:

$$RULE\ 3 \qquad a^{-m} = \frac{1}{a^m}$$

Again, this only works when a is a positive number.

Next, we will tackle a^0. To do this, we will use RULE 1 and RULE 3:

$$a^2 \times a^{-2} = a^{2+(-2)} = a^0$$

But we can also look at this another way:

$$a^2 \times a^{-2} = a^2 \times \frac{1}{a^2} = \frac{a^2}{a^2} = 1$$

Now recall, we must make sure that all the definitions and rules we use are consistent – that is to say that we get the same answer no matter how we tackle a question using our rules. This means we must have $a^0 = 1$.

There are other ways of deciding or justifying that a^0 must have the value 1 and we will touch upon some of these ideas at the end of this section when we look at how we might extend the rules to cases where the powers are irrational numbers.

We now have RULE 4:

$$RULE\ 4 \quad a^0 = 1$$

Again, this works only when a is a positive number.

You should now have enough information to understand the other rules of indices:

$$RULE\ 5 \qquad a^m \div a^n = \frac{a^m}{a^n} = a^m \times a^{-n} = a^{m-n}$$

$$RULE\ 6 \qquad (a^m)^n = \underbrace{a^m \times a^m \times \ldots \times a^m}_{n\ times} = a^{\overbrace{m+m+\cdots+m}^{n\ times}} = a^{mn}$$

$$RULE\ 7 \quad a^{\frac{m}{n}} = (a^m)^{\frac{1}{n}} = \sqrt[n]{a^m} = \left(a^{\frac{1}{n}}\right)^m = \left(\sqrt[n]{a}\right)^m$$

With one of these rules [RULE 6], it is important to be a little careful as sometimes it is possible to misinterpret the notation: consider the two expressions $(a^m)^n$ and a^{m^n}. It is easy to think that these two expressions mean the same thing as they look very similar and often look almost the same when they are written out on paper. But they mean different things:

$$(a^3)^2 = a^3 \times a^3 = a^6 \qquad a^{3^2} = a^{(3\times3)} = a^9$$

Finally, we have been careful throughout this section to emphasise that the rules only work when a is a positive number [rational or irrational]. And our rules allow us to understand the meaning of a^m when m is any positive or negative rational numbers or zero [RULE 3 and RULE 7 are useful here]. In fact, even though the specification restricts things to rational powers only, the rules work for any positive a and any real [rational or irrational] m and n.

Because the specification says "rational exponents" we are careful in TMUA/ESAT questions to ensure that these issues [that powers are rational, but can be irrational] are not ones that you need to think about; that is to say, even though we do set questions where m and n could be irrational, we are careful to ensure this fact does not get in the way of your ability to answer the question.

The rest of this section is NOT part of the TMUA/ESAT specification, so you can skip it if you want. But it is useful and gives you some insights into how mathematicians think about things.

We are going to answer the question: what happens to our rules when m or n are not rational, and what happens when a is zero or even when a is a negative number? This will help you understand why we only have positive a values but we let m and n be any real number.

The answer is that things get more complicated in some cases but not others. We will look at two cases and make some brief comments on some of the others:

Case 1

What happens when a is positive but m and n are irrational?

The answer is [as we have already mentioned] that all the rules still apply, and we interpret a^m [and a^n] where m is irrational in a clever way. Let's look at how we might interpret $2^{\sqrt{3}}$. We will tackle this by looking at how we deal with graphs of exponential functions. You will probably have met the graph of $y = 2^x$ and sketched it but with no thought about whether x is rational or whether x is irrational. If we try to sketch the graph of $y = 2^x$ only when x is rational, we will get a series of dots rather than an unbroken curve. One dot above each rational x on the x-axis. The dots will be so close together that it will be hard to tell by just looking that our graph of dots and the unbroken curve that you would usually sketch for $y = 2^x$ are slightly different. We then assume that the values of 2^x when x is irrational are exactly those values that "fill the spaces" between the dots on our graph to make the unbroken curve of $y = 2^x$ look "the same" as the broken curve of dots. So, the value of $2^{\sqrt{3}}$ is "between" the values of 2^p and 2^q where p is a rational number a teeny bit less than $\sqrt{3}$ and q is a rational number a teeny bit more than $\sqrt{3}$.

This is a rather loose explanation of what we do to define irrational powers, but it is essentially correct. What we actually do is use the idea of limits and we can illustrate this idea by trying to find the value of 2^0. We do this by looking at $2^{\frac{1}{m}}$ but we try to make this expression as close to 2^0 as possible by making $\frac{1}{m}$ as close to 0 as possible.

And we do that by making m either very big and positive or very big and negative. Let's start with m being very big and positive: then $2^{\frac{1}{m}} = \sqrt[m]{2}$ and if you play around with various roots of 2, you will see that as m gets bigger [and so $\frac{1}{m}$ gets smaller and $2^{\frac{1}{m}}$ gets closer to 2^0] that the value of $\sqrt[m]{2}$ gets close to 1. Similarly, if we look at $2^{\frac{1}{m}}$ when m is large and negative, we see that the value of $2^{\frac{1}{m}}$ also approaches 1.[2] This all suggest that we define $2^0 = 1$. And this idea extends to all a^0 where a is positive. You can also sketch the graphs[3] of $y = a^x$ for various positive values of a to see how things fit together [and notice how the shape of the graph changes depending on whether $a > 1$ or $0 < a < 1$]

Case 2

What happens when a is negative ?

In simple terms, things go wrong very quickly. Consider the value of $(-64)^{\frac{1}{3}}$ our definitions suggest this is just the cube root of -64 which is -4 so all seems well. Now, consider $(-64)^{\frac{1}{2}}$; this is supposed to be the square root of -64, but -64 does not have a square root, or at least it does not have one in the real number system. So, we see that it gets messy and for different values of x, even ones very very close together, we encounter problems. This is why, when you first meet them, index laws are only used for positive a values and rational powers [although we can cope with irrational powers as we saw above]

Something to think about: what happens to a^m when $a = 0$?

[2] The case when m is negative takes a little more care to deal with. Have a think about how it works.
[3] A good website to use to help you understand graph sketching is DESMOS GRAPHING

MM1.2

Use and manipulation of surds.

Simplifying expressions that contain surds, including rationalising the denominator.

For example: simplifying $\frac{\sqrt{5}}{3+2\sqrt{5}}$ and $\frac{3}{\sqrt{7}-2\sqrt{3}}$

What is a surd? Here we consider a surd to be an expression that has a root in it [usually a square root] that cannot be simplified to a rational expression. For instance. $3 + 5\sqrt{2}$ is a surd as we cannot simplify it to a rational expression; whereas $3 + 5\sqrt{64}$ is not a surd as we can simplify it to 43. Surds are a mathematical way of expressing numbers exactly and they help get around the impossibility of expressing certain irrational numbers precisely using decimal expansion – that is, for instance, it is impossible to express $\sqrt{2}$ exactly as a decimal as the decimal bit of the number $\sqrt{2}$ is never ending.

Before we look at the sorts of things we might expect you to know for the TMUA/ESAT, we note that there is a convention with square root signs that we adopt in the TMUA/ESAT [it is a standard maths convention] and that is that $\sqrt{a}$ is always positive. So $\sqrt{64}$ is 8 and <u>not</u> -8. If we want to have both 8 and -8, we write $\pm\sqrt{64}$

What sort of things do we expect you to be able to do with surds in the TMUA/ESAT? Let's look at a few:

Simplifying roots

We expect you to be able to simplify expression such as $\sqrt{50}$ or $\sqrt{40}$ in various ways. You should be comfortable with how the following, and similar, examples work:

$\sqrt{50} = \sqrt{25 \times 2} = \sqrt{25} \times \sqrt{2} = 5\sqrt{2}$

$\sqrt{40} = \sqrt{4 \times 10} = 2\sqrt{10} = 2\sqrt{2 \times 5} = 2\sqrt{2}\sqrt{5}$

Multiplying out expressions that involve surds.

The best way to multiply out expression with surds in them is to treat the square roots like x and then simplify at the very final stage. Here is an example:

$$\left(2 + 3\sqrt{5}\right)^2 = 2^2 + 2 \times 2 \times 3\sqrt{5} + 3^2\sqrt{5}\sqrt{5} = 4 + 12\sqrt{5} + 45 = 49 + 12\sqrt{5}$$

Compared with

$$(2 + 3x)^2 = 2^2 + 2 \times 2 \times 3x + 3^2xx = 4 + 12x + 9x^2$$

Factorise expression with surds in them.

From our earlier discussion, we can see that going from $\left(2+3\sqrt{5}\right)^2$ to $49+12\sqrt{5}$ is quite easy. But it is less easy to start with $49+12\sqrt{5}$ and factorise it to get $\left(2+3\sqrt{5}\right)^2$. Nevertheless, you should be able to factorise expressions with surds in when it might be useful. For instance, if you were asked to find the exact value of $\sqrt{49+12\sqrt{5}}$ you would need to spot that $49+12\sqrt{5} = \left(2+3\sqrt{5}\right)^2$ to get to the answer.

How might you go about factorising expression with surds in them? There are a few "tricks" that might help sometimes – although it is best to practise this yourself and devise your own "tricks" and way of approaching the factorising.

Let's look at $49+12\sqrt{5}$ and compare it with $4+12x+9x^2$ from above:

The first thing we notice is that the $12\sqrt{5}$ is the middle term [and this is usually where we start with this sort of question], so this suggest that our original expression must be $\left(a+b\sqrt{5}\right)^2$. We then notice that the middle term is also written as $2ab\sqrt{5}$. This means we must have $ab=6$ and $a^2+5b^2=49$. Now all we need to do is substitute pairs of numbers that multiply to give 6 into the equation $a^2+5b^2=49$ until we find some that work. It doesn't take long to find $a=2$ and $b = 3$ [you can start with b and work though $b=1,2,3,6$ until you find which one works]

Exercise

Start with some random surd expression and square them and simplify them. Then look at the expressions, maybe after some days, and see if you can factorise them back to their original form.

Exercise

What is $\sqrt{49-12\sqrt{5}}$? Be very careful – you might get the wrong answer[4] !

Rationalising the denominator

This is probably the most common thing you will meet with surds in standard exams. It is about "moving" the surd expression in a fraction from the bottom of the fraction the top. That is, we want to find another expression that has the same value as the

[4] The answer is not $2-3\sqrt{5}$. Why not ?

original, but which has a surd in the numerator [top] of the fraction and no surds in the denominator [bottom].[5]

In simple cases, this is straightforward. For instance, how do we rationalise the denominator for $\frac{1}{\sqrt{5}}$? The answer is we multiply the expression by 1 but we write 1 in an unusual manner: we use $1 = \frac{\sqrt{5}}{\sqrt{5}}$ and this gives:

$$\frac{1}{\sqrt{5}} = \frac{1}{\sqrt{5}} \times 1 = \frac{1}{\sqrt{5}} \times \frac{\sqrt{5}}{\sqrt{5}} = \frac{\sqrt{5}}{\sqrt{5} \times \sqrt{5}} = \frac{\sqrt{5}}{5}$$

In practice, you should be able to convert from $\frac{1}{\sqrt{5}}$ to $\frac{\sqrt{5}}{5}$ almost without thinking about it: $\frac{1}{\sqrt{a}} = \frac{\sqrt{a}}{a}$

What about more complicated expressions such as $\frac{3}{2+\sqrt{5}}$? We use the same idea – that is, multiplying by 1, but by writing 1 in a specific way. We also use the standard difference of two squares formula - this is a formula that crops up everywhere and so you should always be on the lookout for it just in case it might be useful, but also be aware that using it can sometimes lead you down the wrong path.[6]

Recall the difference of two square formula: $a^2 - b^2 = (a+b)(a-b)$

Let's look at using all this to rewrite $\frac{3}{2+\sqrt{5}}$. We need to multiply by 1 written in an appropriate manner. We start by using the difference of two squares and noticing:

$$(2+\sqrt{5}\,)(2-\sqrt{5}\,) = 4 - 5 = -1$$

And also

$$(\sqrt{5}\ +2)(\sqrt{5}\ -2) = 5 - 4 = 1$$

[5] As you will be aware, the top and the bottom of fractions have different roles. The bottom of a fraction [denominator] tells us [denominates] what sort of fraction it is [is it halves, or thirds, or quarters, etc]. The top of a fraction [numerator] enumerates [i.e. tells us how much] of the fraction we have. This seems obvious but it is often misunderstood when fractions are first met – for instance, it is common to see $\frac{1}{2} + \frac{2}{3} = \frac{1+2}{2+3}$. Have a think about how you perform fraction division: usually the rule used is "flip the second fraction and multiply". Can you explain why this method works [hint, make the denominators the same then think about what the denominator tells you] ? And is the following true [and, if so, why]: $\frac{a}{b} \div \frac{c}{d} = \frac{a \div c}{b \div d}$?

[6] Doing mathematics is sometimes a bit like playing chess. When you have an idea, you also need to anticipate what the consequences of applying the idea will be before devoting time to pursuing it. Sometimes, ideas that seem perfect at the start will lead you down the wrong path. Mathematics is also a bit like chess in that the more you practise your mathematics, the better you will become at it.

Looking carefully at both of these, we decide to use 1 written as $\frac{\sqrt{5}-2}{\sqrt{5}-2}$ [we could use $\frac{2-\sqrt{5}}{2-\sqrt{5}}$ but it gives us a negative value and that involves a teeny bit more work]. So, we now rewrite our expression $\frac{3}{2+\sqrt{5}}$ as follows:

$$\frac{3}{2+\sqrt{5}} = \frac{3}{2+\sqrt{5}} \times 1 = \frac{3}{2+\sqrt{5}} \times \frac{\sqrt{5}-2}{\sqrt{5}-2} = \frac{3\sqrt{5}-6}{5-4} = \frac{3\sqrt{5}-6}{1} = 3\sqrt{5}-6$$

Of course, you should be able to rationalise denominators much more quickly than we have done above – we just took our time to explain each step carefully.

As an aside: it is very easy when rationalising denominators to be careless and write $\sqrt{5} \times \sqrt{5} = 25$. Make sure you don't!

Exercise

What is the factorisation of $a^3 - b^3$ and of $a^3 + b^3$? [these factorisations are both useful and you ought to know them]

Can you use these to help you rationalise the denominators in:

$$\frac{7}{\sqrt[3]{36} + \sqrt[3]{6} + 1}$$

$$\frac{x-27}{\sqrt[3]{x} - 3}$$

MM1.3

Quadratic functions and their graphs; the discriminant of a quadratic function; completing the square; solution of quadratic equations.

Quadratics are essentially functions that are written in the form $ax^2 + bx + c$ where $a \neq 0$. The term “quadratic” can mean the function, or the graph, or the expression etc. – it is generally a loose term for all things related to the function.

In this section we will look at the quadratics both algebraically and graphically. It will help if you read this section in conjunction with the one on “graph shifting” below and also have some graphing software to hand [e.g., DESMOS GRAPHING].

Why is there so much talk about quadratics in GCSE and A level mathematics? The simple answer is that they are both simple to deal with and have lots of interesting properties that can be clearly illustrated without adding unnecessary complications. They are the perfect “toy” function to play with to help you start to understand more complicated concepts and graphs as your mathematical knowledge develops.

Before we start exploring some aspect of quadratics, let’s list many of the sorts of things you are expected to be able to do with quadratics:

1. Factorise them when appropriate.
2. Solve quadratic equations using “the formula”.
3. Complete the square for a given quadratic [and it is useful to know how this process relates to the quadratic formula]
4. Understand the relationship between the roots of a quadratic and its graph.
5. Understand the relationship between a quadratic written in completed-square form and its graph.
6. Sketch quadratic curves given the equation [marking x-intercepts (i.e., roots) y-intercept and min /max coordinates]
7. Understand how to use the discriminant of a quadratic to tell you about the quadratic’s roots.
8. Know how to find the coordinates of the min/max by completing the square, or by differentiating.

Generally, if you are asked to solve a quadratic equation, it is usually better to see if you can factorise it before launching into other methods [such as using the quadratic formula or completing the square].

The main topic we shall look at here is the interplay between the algebra of quadratics and their graphs:

Let's start by looking at a simple quadratic $y = x^2 + bx + c$ and write it in "completed square" form:[7]

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

We can see from this that the graph of $y = x^2 + bx + c$ is the same shape as the graph of $y = x^2$ but it is in a different place on the xy plane. In fact, if we start with the graph of $y = x^2$ and "shift" [or translate] it to the left by $\frac{b}{2}$ and then "up" by $-\frac{b^2}{4} + c$ then we get the graph of $y = x^2 + bx + c$. We have merely started with the graph of $y = x^2$ and moved it to a new position on the xy plane; we have not squashed or stretched it. So, all graphs of the form $y = x^2 + bx + c$ are really just the $y = x^2$ graph but translated to another part of the xy plane.

If you look at the "graph shifting" section of these notes, you will see that if $f(x) = x^2$ then $y = f(x)$ on shifting becomes $y - \left(-\frac{b^2}{4} + c\right) = f\left(x - \left(-\frac{b}{2}\right)\right)$ which is the translation we described.

What happens to the minimum of $y = x^2$ [which is at (0, 0)] when it is shifted by the translation? Its x-coordinate moves to $-\frac{b}{2}$ and its y-coordinate moves to $-\frac{b^2}{4} + c$. This is not surprising, and we can get the same result in other ways. For instance, we can find the x coordinate of the minimum by differentiating and setting equal to 0: $\frac{\mathrm{d}}{\mathrm{d}x}(x^2 + bx + c) = 2x + b = 0$ so the minimum occurs at $x = -\frac{b}{2}$ and the y-coordinate of the minimum is [by substituting $x = -\frac{b}{2}$ into $y = x^2 + bx + c$] $-\frac{b^2}{4} + c$ which is exactly what we expect from our discussion of graph shifting as the point (0, 0) would be shifted to $\left(-\frac{b}{2}, -\frac{b^2}{4} + c\right)$.

We can investigate the minimum coordinates of $y = x^2 + bx + c$ a third way: first, recall that we have $y = x^2 + bx + c = \left(x + \frac{b}{2}\right)^2 - \frac{b^2}{4} + c$ and we want to find what x value gives us the least y-value. We notice that $\left(x + \frac{b}{2}\right)^2$ is always positive or zero, to make the y value the least possible, we need to make $\left(x + \frac{b}{2}\right)^2$ equal to zero; and this means we need $x = -\frac{b}{2}$ just like before.

What does our discussion about graph shifting for $y = x^2 + bx + c$ tell us about the roots of the quadratic? First we recall that the roots of the quadratic are the solutions to $x^2 + bx + c = 0$ and so they are the x-values when the y-value is zero, and so they are the x-values when the graph crosses the x-axis [because the x-axis has the equation $y = 0$]. How do we know if the shifted graph has roots, in other words, how

[7] As always, make sure you understand how the "completed square" form works.

do we know if the shifted graph crosses the x-axis? The answer is straightforward and leads us to the discriminant condition for roots: we start by recalling (0, 0) on $y = x^2$ is shifted to to $\left(-\frac{b}{2}, -\frac{b^2}{4} + c\right)$ on $y = x^2 + bx + c$. The only way that the new graph can cross the x-axis is if the y-coordinate of its minimum point is "under" the x-axis[8]; in other words, we require $-\frac{b^2}{4} + c < 0$ for the graph to cross the x-axis [we will deal with touching the x-axis in a moment]. Rearranging this, we get that the quadratic has [two real] roots when $0 < b^2 - 4c$ which is the discriminant of the quadratic [see our discussion below if you are not familiar with this]. And if the graph just touches the x-axis then we know the quadratic has one [repeated[9]] root and also we must have $-\frac{b^2}{4} + c = 0$ which leads to $-\frac{b^2}{4} + c = 0$

Exercise

How does changing the values of b and c in $y = x^2 + bx + c$ affect the position of the graph?

You can use a graph sketching package [e.g., DESMOS GRAPHING] and play around with different b and c values [positive and negative] to see what happens and then make sure you can justify your findings by referring to the discussion above and completing the square.

So far, we have looked at $y = x^2 + bx + c$ and related it to $y = x^2$ but what about the more general quadratic of the form $y = ax^2 + bx + c$ [where $a \neq 0$ to make sure it is a quadratic]?

We can start by completing the square on this [the algebra is a little unpleasant but worth doing all the same[10]].

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

And rewriting this a bit [we have jumped steps but make sure you can fill them in!]:

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

[8] Notice we have also used the fact that the coefficient of x^2 is positive so the quadratic is a U shape.

[9] We say a root is *repeated* if it occurs more than once – e.g. $(x + 2)(x + 2) = 0$

[10] There are a number of equivalent but slightly different approaches to completing the square when the x^2 coefficient is not just 1. We have picked one approach, but you might have seen others in your maths classes. You should be comfortable with a range of approaches here.

We can then use this expression to derive the quadratic formula and to explore the relationship between the formula and aspects of the graph of $y = ax^2 + bx + c$

Let's start with deriving the quadratic formula:

For this, we want to solve $ax^2 + bx + c = 0$ and when we replace the quadratic by its completed-square form this becomes:

$$a\left(x + \frac{b}{2a}\right)^2 - \left(\frac{b^2 - 4ac}{4a}\right) = 0$$

which gives

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

and hence

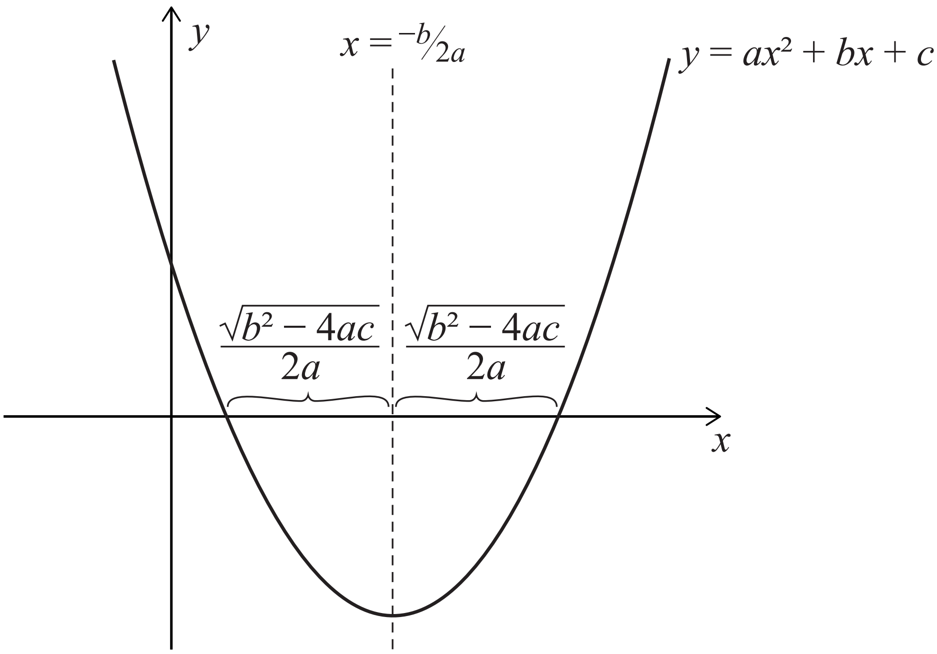
$$x + \frac{b}{2a} = \pm\sqrt{\left(\frac{b^2 - 4ac}{4a^2}\right)}$$

[noticing here we need both the positive and the negative square roots] which rearranges to

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

We can relate this to the sketch of $ax^2 + bx + c$:

The line of symmetry for the graph is at $x = \frac{-b}{2a}$ [make sure you can argue why this is true] and the distance from this line of symmetry to each of the roots is $\frac{\sqrt{b^2-4ac}}{2a}$ and so the distance between the roots must be $\frac{\sqrt{b^2-4ac}}{a}$. See the picture on the next page.



Finally for this section let's remind ourselves of the discriminant condition and how it tells us if the graph cuts the x-axis [and recall that the discriminant is the $b^2 - 4ac$ in the formula]. Here are the three possible conditions and what they tell us:

$b^2 - 4ac > 0$ quadratic has two real distinct roots [cuts x-axis at two distinct points]

$b^2 - 4ac = 0$ quadratic has one repeated root [touches x-axis]

$b^2 - 4ac < 0$ quadratic has no real roots[11] [never cuts or touches the x-axis]

You should also be able to explain clearly why the discriminant condition works – you should be able to explain algebraically using the formula, and also using graph sketching and graph shifting. Spend some time now checking all these different approaches to dealing with quadratic roots fit clearly together in your mind.

[11] This means the quadratic never crosses the x-axis. When you learn about complex numbers, you will discover that the quadratic does have roots, i.e. x values that make the quadratic zero; but it turns out that these roots don't exist in the real numbers, and we have to go to a "larger" number system to find them. There are all sorts of interesting "number systems" [we are deliberately vague here as to what we mean by a "number system" as it can get quite technical and many "number systems" are not really like the numbers you are used to when counting on your fingers and toes.] in maths and some of them have turned out to be essential for dealing with higher level physics too: e.g. complex numbers in elementary quantum theory.

Exercise

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

Divide the expression through by x^2

You now have a quadratic in $\frac{1}{x}$

Use the quadratic formula to show:

$$\frac{1}{x} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2c}$$

And hence that

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

Then show this is the same result as that given by the standard quadratic formula. Are there any conditions on the equivalence of the two formulae [e.g., can we have $c = 0$ for the alternative; and, if not, why not?]?

MM1.4

Simultaneous equations: analytical solution by substitution, e.g. of one linear and one quadratic equation.

In this section, we will look at simultaneous equations. We will look at what simultaneous equations are and then we will investigate how we can solve them. After that, we will look at how the solutions to simultaneous equations fit with the corresponding graphs of the equations.

Let's start by stepping back and exploring how we might think about equations and their graphs in general terms.

We start by considering the simple [linear] equation $y = x + 2$ where x is taken from the real numbers. This equation is really a quick way of writing out a [very large] set of number pairs: for instance, when $x = 1$, $y = 3$ so the number pair (1, 3) is in the set; as is (2, 4) and $(\pi,\ \pi\ + 2)$ and $(\sqrt{2}\ ,\ 2+\sqrt{2})$ etc. In fact, the set of number pairs represented by the equation $y = x + 2$ where x is real is so large that we can never write out all the members of the set.[12] Nevertheless, it is just a very large set of number pairs that all obey the pattern $y = x + 2$.

In other words, we can think of an equation [once we are given all the possible x values – e.g., x is real, or x is real and $x > 0$ etc., etc.] as a quick way of summarising a very large set of number pairs. We also know there is a very clever way of drawing all these number pairs on a diagram – that is we draw the line $y = x + 2$ and each point on the line we draw represents one of the number pairs from the set. This may all seem trivial to you but when the idea of drawing pictures [i.e. graphs] of equations was first introduced it was a major change for the way people could do and understand mathematics – this move from algebra and sets of number pairs to geometry is something we take for granted now but it was not always that way.[13] It is very useful to get used to trying to understand how anything algebraic you meet can be interpreted geometrically and vice versa; sometimes this is not an easy task but when you see connections between algebra and geometry as you progress as a mathematician, you will find your understanding is greatly enhanced.

Now let's turn to look at simultaneous equations and to learn what is meant by solving them and how to interpret their solutions.

Here are two equations:

$$x + 2y\ =\ 5$$

[12] It is "uncountably infinite". If you are interested, look up Cantor's infinities.

[13] The xy plane, also known as the Cartesian plane, was introduced to mathematics in 1637 by Rene Descartes. He first thought of the idea, so we are told, thanks to his watching a fly on a ceiling.

$$2x + y = 4$$

We are going to find what x and y values make both of these equations true at the same time [i.e., simultaneously] – that is what we mean by “solving” the equations simultaneously. In other words, we are looking for the number pairs that appear in both the set of number pairs for $x + 2y = 5$ and the set of number pairs for $2x + y = 4$. And, if we think geometrically, that must be where the graphs of the equations [which are, recall, pictures of all the number pairs] cross – because the point where they cross must be in both sets.

We will look at three ways of solving these: by substitution; by elimination; and graphically.

By substitution:

We make x the subject of the equation $x + 2y = 5$ to get $x = 5 - 2y$ and then we replace x in the second equation by $5 - 2y$: so, $2x + y = 4$ becomes $2(5 - 2y) + y = 4$. Then simplifying: $10 - 3y = 4$ so $y = 2$ and because $x = 5 - 2y$ we find that $x = 1$. So the solution to the simultaneous equations is $x = 1$ and $y = 2$

By elimination

[This is really like technique 1 but written a little differently]:

We make the number of x’s in each equation the same [we could have chosen to do this for y instead]

$x + 2y = 5 \;\rightarrow \times 2 \rightarrow\; 2x + 4y = 10$

Then we have:

$2x + 4y = 10$
$2x + y = 4$

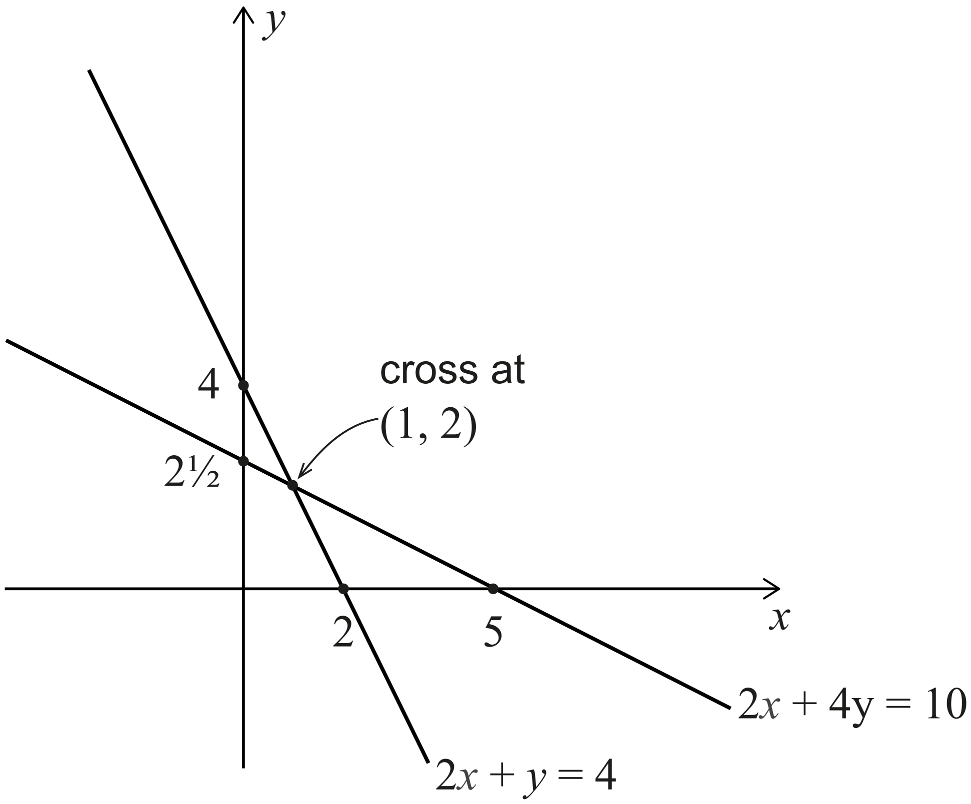
And subtracting the bottom equation from the top one, we get:

$3y = 6$

Which gives $y = 2$ and then $x = 1$ as before.

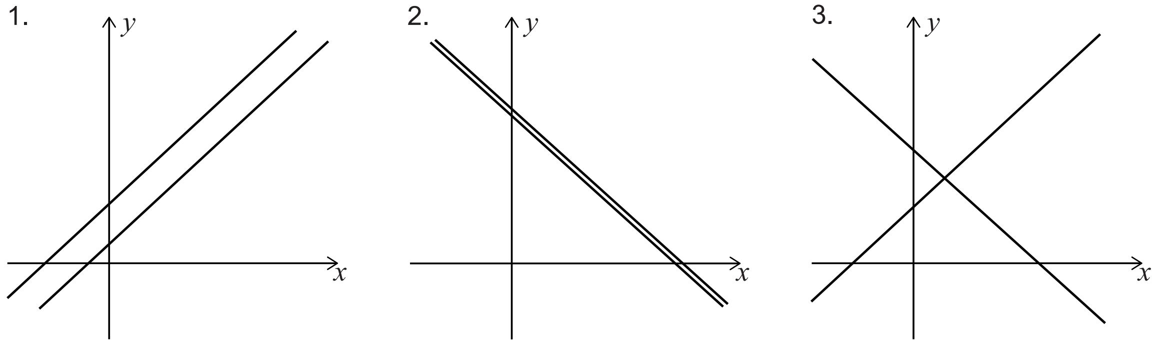
Graphically

We draw the graph of each equation and look at the point where they cross:



Let's look now at simultaneous [linear] equations in more general terms. And recall, linear equations are ones that have graphs which are straight lines. We are going to ask how many solutions are there to a pair of simultaneous linear equations. And we are going to answer the question by looking at the different ways we can draw two straight lines on the xy plane.

We can draw two straight lines on the xy plane in three different ways:



1. Two lines that are parallel and distinct. These lines will not cross and so have no solutions. Let's have a quick look at the equations of two such lines: because they are parallel, they must have the same gradient so the ratio of x to y in each equation must be the same. For example, $y + 2x = 4$ and $y + 2x = 8$. We can see that these lines cannot have any common xy values as there can't be $x\ y$ values such that $y + 2x$ adds to both 4 and to 8 !

2. Two lines are parallel and the same. In this case every xy pair is a solution as we are really only writing the equation of one line twice. For instance: $y + 2x = 4$ and $2y + 4x = 8$.

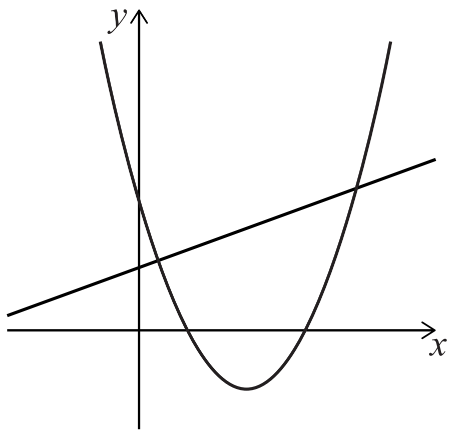
3. Two lines that are not parallel. Here the lines will have to cross exactly once [convince yourself this is obvious [14]] and so these sorts of simultaneous equations will always have one solution.

So far, we have explored linear simultaneous equations What about looking at simultaneous equations where one is linear, and one is a quadratic ?

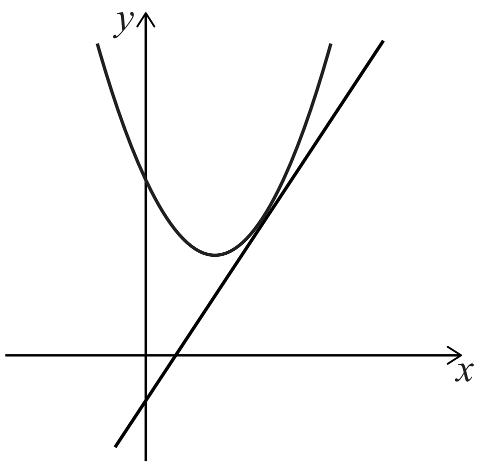
[14] Remember, we are in two dimensions. The situation gets a little more complicated in three dimensions. But we do not ask you to deal with three dimensions in the TMUA. Nevertheless, do think about all the possible cases for two lines in three dimensions.

We can start to get a picture of the possibilities by considering sketches of lines and quadratics. There are three possibilities:

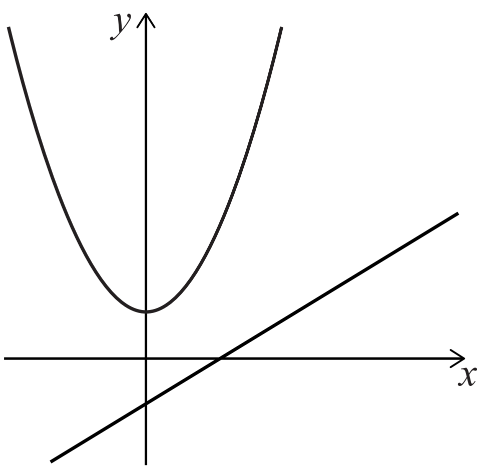
1. Line crosses quadratic giving two distinct solutions.



2. Line is tangent to quadratic giving one [repeated] solution.



3. Line never crosses quadratic giving no [real[15]] solutions.



[15] There will be, in this case, complex solutions but this is outside the scope of the TMUA/ESAT

Solving simultaneous equations where one is linear and one is quadratic is simple: we eliminate the y values [or x values if easier] and solve the resulting quadratic equation using one of the techniques we met earlier in these notes. Factorizing or the discriminant conditions on the resulting quadratic [resulting from the elimination process] will tell you quickly which of the three cases for quadratic plus linear you are dealing with.

Here is an example:

Example

Solve $y = x^2 + 3x + 2$ and $y = x + 1$

To solve these, we eliminate y to give

$x^2 + 3x + 2 = x + 1$

And rearranging

$x^2 + 2x + 1 = 0$

Which factorises to

$(x + 1)^2 = 0$

so there is a single solution of $x = -1$ and $y = 0$ and this must be when the line is tangent to the quadratic.

Exercise

Finally, some things to think about:

What are the possible solutions for a quadratic and a cubic simultaneous equation. Is it possible to have a situation with no real solutions [justify your answer both algebraically and using graphs[16]]?

Find a condition on $m, c, b,$ and d for $y = mx + c$, $m \neq 0, c \neq 0$ to be tangent to $y = x^2 + bx + d, b \neq 0, d \neq 0$

[16] Hint [stop reading if you don’t want a hint]: which graphs generally get steeper faster, quadratics or cubics? Why ?

MM1.5

Solution of linear and quadratic inequalities.

The most important things to be aware of when dealing with inequalities is that they do NOT behave quite the same way as equations with equal signs. With inequalities you can add and subtract on both sides as much as you want, but you cannot multiply and divide both sides [or raise both sides to an even power, or apply certain functions to both sides] without first checking that what you are multiplying or dividing by is positive and/or preserves the inequality:

Here are a few examples :

- $-7 < 5$ is true but multiply both sides by -1 to get $7 < -5$ which is clearly false.
- $-2 < -1$ is true but squaring both sides gives $4 < 1$ which is clearly false.
- $-2 < -1$ is true and cubing both sides gives $-8 < -1$ which is also true.
- $-8 < -4$ is true but dividing both sides by -2 gives $4 < 1$ which is clearly false.
- $30 < 60$ is true but $\cos 30 < \cos 60$ is false while $\sin 30 < \sin 60$ is true.

In general, when dealing with inequalities, it is important to make sure that the way you are manipulating them does not turn a true statement into a false one or generate extra [incorrect] solutions if you are dealing with algebra.

Let's have a look at several examples to illustrate the sorts of things you need to think about as you solve inequalities:

Example 1

Solve $3x + 2 \leq x + 5$

We can answer this by simple rearranging:

$3x - x \leq 5 - 2$

$2x \leq 3$

$x \leq \frac{3}{2}$

Example 2

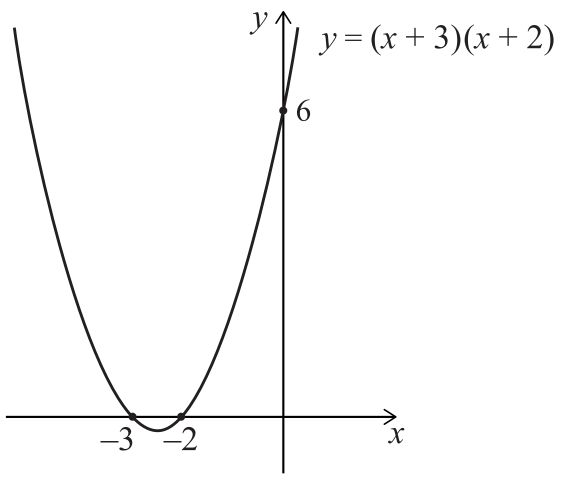
Solve $x^2 + 5x + 6 \geq 0$

We will solve this using graph and some algebra.

First, we factorise the left-hand side:

$$(x + 2)(x + 3) \geq 0$$

Then we roughly sketch the quadratics and find out what x values make the graph sit above the x-axis [that is ≥ 0]



And from the diagram we can see the solution must be $x \leq -3$ or $x \geq -2$

Example 3

Solve:

$$\frac{2x+5}{x+3} > 1$$

It is very tempting [but not a good strategy] to multiply both side of this expression by $x+3$. However, as $x+3$ is negative for some values of x the inequality sign will then be the wrong way around for some values of x and lead to a slight mess. Instead, we need to adjust our algebraic approach slightly and the best thing to do is to multiply both sides by $(x+3)^2$ which we know is always positive for all real x values and so it will not affect the inequality sign:

$$\frac{2x+5}{x+3} \times (x+3)^2 \;>\; 1 \;\times (x+3)^2$$

Giving [note we rearrange rather than multiply out as we have a common factor we can extract] :

$$(2x+5)(x+3) - (x+3)^2 > 0$$

And on simplifying

$$(x+3)(2x+5-x-3) > 0$$

$$(x+3)(x+2) > 0$$

Which is solved as in example 2 above?

Make sure you are familiar with various techniques for solving inequalities – there are lots of ways of approaching things and you might have learnt some that we have not mentioned above – the main approaches tend to be via algebra and graphs or via using number lines.

As an aside, it is useful to know how to write ranges correctly, particularly as it is common to see incorrect expressions.

Let's look at a couple of howlers:

$1 < x < -5$

in this case, it is clear what is meant but we tend to read strings of inequalities as a single "sentence" and this expression is really two separate statements scrunched together [it is presumably meant to be "$1 < x$ or $x < -5$"]. It is, of course, ok to write things like $-3 < x < 5$ as the whole expression makes sense, and it is true also that $-3 < 5$

Here is another one:

$-2 < x$ and $x < -4$

this is [presumably[17]] incorrect as there are no x values that satisfy both $-2 < x$ and $x < -4$ [look at how "and" is used in the *Notes on Logic and Proof* that we have written for TMUA paper 2]. Here, it would have been better to write either: "$-2 < x$ or $x < -4$" and perhaps even better "$x < -4$ or $-2 < x$" and sometimes just "$x < -4,\ -2 < x$". Our preference would be to use "or" .

Let's now turn to look at inequalities involving the modulus sign. We will use some sketches in this section, but we will postpone a detailed look at modulus graph sketching until later.

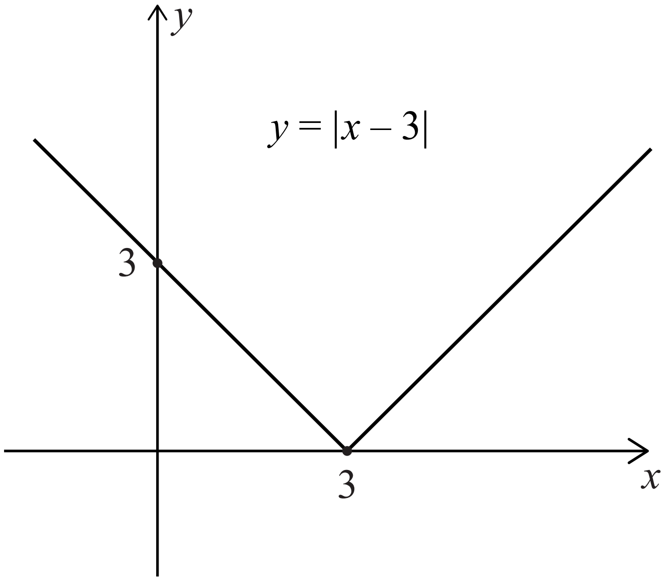
We start by looking at a simple modulus function. We are going to look at.

$$y = |x - 3|.$$

Recall first that $|x - 3|$ means "the positive value of $x - 3$. So when $x = 5$ then $y = 2$ and when $x = -4$ then $y = 7$.

[17] Unless the intention was to write "no solutions" in a silly manner.

We can sketch this[18]:

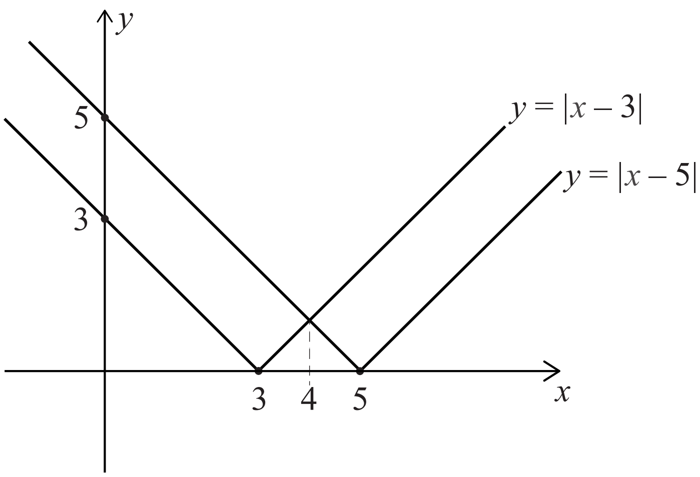


But the best way to think about the meaning of $|x-3|$ is as a measure of the [positive] distance of x from 3 on the number line. Keep this idea in mind as we look at some inequalities involving the modulus.

Let' start by solving the inequality:

$$|x-3|<|x-5|$$

We can solve this graphically by sketching both $y=|x-3|$ and $y=|x-5|$ and working out for which x values the graph of $y=|x-3|$ is below [or the y values are less than] the graph of $y=|x-5|$

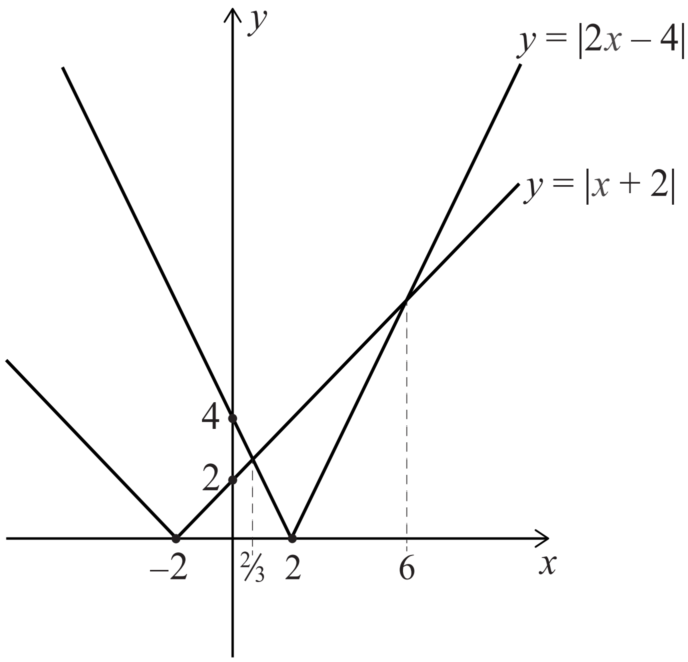


[18] If you are not sure how to sketch modulus functions, the look at the section of graph sketching below.

And from the diagram, we see that the x values that make the inequality correct [or true] are $x < 4$

Another way of thinking of the inequality $|x-3| < |x-5|$ is asking what are the x values that are closer to 3 than they are to 5. And a little thought leads rapidly to the solution $x < 4$

How might we use these ideas to solve slightly more complicated inequalities? For instance, what about the inequality $|2x-4| < |x+2|$ which we can rewrite as $2|x-2| < |x+2|$. We can solve this with a sketch.



The inequality asks for the x values for which the graph of $|2x-4|$ sits "below" [less than] the graph of $|x+2|$ and this gives $\frac{2}{3} < x < 6$

It is worth spending a moment to look at how we found the values $\frac{2}{3}$ and 6 from the graphs. If you are not sure of how we get some of the equations here, you should read the section on modulus graph sketching later in these notes.

To find the value $\frac{2}{3}$ we use the sketch to guide us as it tells us we need to solve $x + 2 = -(2x-4)$ where the minus sign arises because we are looking at the "flipped" part of $y = |2x-4|$. And to find the value 6 we solve $2x - 4 = x + 2$

Or we can think about it as before as asking when is twice the distance of x from 2 less than the distance of x from -2. A little thought [more than before] gives us the answer: $\frac{2}{3} < x < 6$

Comment

There is a wide range of inequalities we could look at in the TMUA/ESAT. What is important is that you understand what inequalities are, and that you have a set of techniques to solve them – relying just on algorithms for solving without understanding is dangerous and limiting. We have looked at a few inequalities above and also looked at some [but not all] techniques for solving them.

MM1.6

Algebraic manipulation of polynomials, including:

a. expanding brackets and collecting like terms

b. factorisation and simple algebraic division (by a linear polynomial, including those of the form $a\,x + b$, and by quadratics, including those of the form $ax^2 + bx + c$)

c. use of the Factor Theorem and the Remainder Theorem

In the TMUA/ESAT we expect you to be able to multiply our brackets and collect like terms; and recall collecting ”like terms” means collecting all the constants together, and separately collecting all the x terms together, and separately the x^2 terms, and separately the x^3 terms, and so on.

You should also be able to factorise simple algebraic expression – certainly quadratics and other expressions with common factors. You should also be able to factorise cubics using the factor theorem [see below].

In addition, make sure you can perform simple algebraic [long] division – you should be able to perform long division by linear and quadratic expressions. Here is an example:

Example

What is $x^4 + 2x^2 + 3x - 4$ divided by $x + 3$?

We write this out in a grid [below this green box] to make things easier to follow – and notice we have a column for each power of x and we have also included $0x^3$ in our grid – this is a good idea as otherwise it is easy to make errors as you work through the long division. We also keep the same powers of x in the same vertical columns throughout [we do this for the top line [the $x^3 - 3x^2 + 11x - 30$ line] for consistency but it is not so important there and you might prefer to shift that line to the left so that the x^3 appears above the x^4 etc.]

		x^3	$-3x^2$	$+11x$	-30
$x+3$	x^4	$+0x^3$	$+2x^2$	$+3x$	-4
	x^4	$+3x^3$	$\downarrow$	$\downarrow$	$\downarrow$
	$0x^4$	$-3x^3$	$+2x^2$	$+3x$	-4
		$-3x^3$	$-9x^2$	$\downarrow$	$\downarrow$
		$0x^3$	$+11x^2$	$+3x$	-4
			$+11x^2$	$+33x$	$\downarrow$
			$0x^2$	$-30x$	-4
				$-30x$	-90
				$0x$	86

Notice that division involves just first terms at each stage, but when we multiply back to get what we need to subtract at each stage, then we use all the terms. We stop the long division when we get 86 as $x + 3$ does not divide into 86. In fact, 86 is our remainder.

This long division tells us that when $x^4 + 2x^2 + 3x - 4$ is divided by $x + 3$ the answer is $x^3 - 3x^2 + 11x - 30$ with a remainder of 86.We can write this as follows:

$$x^4 + 2x^2 + 3x - 4 = (x^3 - 3x^2 + 11x - 30)(x + 3) + 86$$

Study this carefully and make sure you can see exactly how it fits with our grid above and how it fits with your normal understanding of division [e.g., 11 divided by 4 is 2 remainder 3 so that we can write $11 = 2 \times 4 + 3$]

Factor and remainder theorem.

The factor theorem appears in many advanced school mathematics specifications, but the remainder theorem appears in fewer, so you might not have met both theorems in your maths classes. We have decided to keep both theorems in the TMUA/ESAT specification as they are closely related to each other and easy to understand.

We will start by looking at the factor theorem. For this we will use the notation $f(x)$ and you will need to know what is meant by a factor in algebra. Let's start by refreshing our understanding of these two things.

First, what does it mean to be a factor of an algebraic expression? Here are a couple for examples:

Consider $(x+2)(x-3)$ then both $(x+2)$ and $(x-3)$ are factors but, for instance, $(2x+5)$ isn't.

Consider $(2x-7)(x^2+3x+7)$ then both $(2x-7)$ and (x^2+3x+7) are factors.

In general, one algebraic expression is a factor of another of it divides into it exactly - that is, without any remainder. This is just like the idea of factors for numbers.

Next, we remind ourselves that $f(x)$ [or $g(x)$ or $h(x)$ etc] is just another notation for a function of x, or really just an algebraic expression. We often write $f(x)$ instead of y when we are dealing with algebraic expression. So, for instance we could write $f(x) = x^2+3x+2$. This notation is useful as we can then indicate the value of the expression for different $x$values. For instance, $f(2)$ means the value of the expression that is represented by $f(x)$ when $x = 2$; so when $f(x) = x^2+3x+2$, then $f(2) = 2^2 + 3 \times 2 + 2 = 12$.

Now let's use our knowledge of factors and the notation $f(x)$ to set up and explain the factor theorem:

Let's start with $f(x) = (x-2)(x+7)$. We know that both $(x-2)$ and $(x+7)$ are factors of $f(x)$. Let's look at the values of $f(2)$ and $f(-7)$. In both cases, a quick calculation shows that each has a value 0. This leads to the factor theorem which states:

If $\boldsymbol{f(x)}$ is a polynomial in $\boldsymbol{x}$, then $\boldsymbol{f(a) = 0}$ if and only if $\boldsymbol{x-a}$ is a factor of $\boldsymbol{f(x)}$

Before we look at how we can use the factor theorem, we will make a few comments. First, notice that the factor theorem is applicable only to polynomials [and in TMUA/ESAT that means polynomials with real coefficients]. And also notice it is an "if and only if" statement. That means we can use it two ways:

1. We can find that $f(a) = 0$ and then we know that $(x-a)$ is a factor of $f(x)$

2. We can start with $(x-a)$ as a factor of $f(x)$ and then we know that $f(a) = 0$

3. You should realise that you know the factor theorem in a slightly different form: the factor theorem links the roots of the equation [when the graph crosses the x-axis – that is the solution to $f(x) = 0$] with the factors you get when you factorize an expression into brackets. So, for instance, if you are asked to find the roots of $f(x) = (x+4)(x-7)$ [in other words find where $y = (x+4)(x-7)$ crosses the x-axis] you know you have to solve $(x+4)(x-7) = 0$. So you are asking what x values make $f(x) = 0$ and they must be the x values that make $(x+4) = 0$ or $(x-7) = 0$. Read the factor theorem again to make sure you see the connections.

The factor theorem itself should be fairly "obvious". It is useful in form 1 above as it helps us to factorise polynomial expression [mostly quadratics and cubics]. Here are some examples of using the factor theorem :

Example 1

Factorise $f(x) = x^2 + 3x + 2$

We will look at this in a little more detail than usual. First we note that if the factors are $(x-a)$ and $(x-b)$ then $ab = 2$ so we only consider $a = \pm 1$ and $a = \pm 2$ [once we know a we can immediately work out b.][19]

Next, we use the factor theorem to work out what a might be by a sort of "trial and error" process. We want to find an a such that $f(a) = 0$ and we see that as all the bits of the quadratic are positive, then it is best to try negative values for a. We start with $a = -1$ and work out $f(-1) = 0$ so we know that $(x - -1)$ is a factor, ie $(x+1)$ is a factor. This immediately gives $b = -2$ and so the other factor is $(x+2)$

Example 2

Factorise $f(x) = x^3 + x^2 - 5x + 3$

Here, using the same idea as in Example 1, we can see that this probably[11] factorises to $(x+a)(x+b)(x+c)$ with $abc = 3$ [the constant term in the cubic]. So, we should start by using the factor theorem using factors of 3 [i.e., 1, -1, 3 and -3].

We start with looking at $f(1) = 1^3 + 1^2 - 5 \times 1 + 3 = 0$ so we know that $(x-1)$ is a factor.

Now we could continue to check -1, 3 and -3 and if you were to do so, you would find that $(x+3)$ is a factor because $f(-3) = 0$. As there are no other factors of 3 and

[19] Of course, we are assuming it CAN be factorised into nice brackets with integers!!

we have found only two factors then it looks like something has gone wrong. Actually, it hasn't, one factor is repeated, and the answer is $(x-1)(x-1)(x+3)$.

There is an alternative way to proceed. Once we know that $(x-1)$ is a factor we can use long division [or factorising by inspection – we have not included that technique in this notes] to find a quadratic factor of $f(x) = x^3 + x^2 - 5x + 3$ and then factorise that quadratic.

Long division gives:

$$f(x) = x^3 + x^2 - 5x + 3 = (x-1)(x^2 + 2x - 3) = (x-1)(x-1)(x+3)$$

Now we have an idea of what the factor theorem says and how to use it, we will turn to look at the remainder theorem. Recall above when we undertook an algebraic long division, we obtained the following expression

$$f(x) = x^4 + 2x^2 + 3x - 4 = (x^3 - 3x^2 + 11x - 30)(x+3) + 86$$

If we use this to calculate $f(-3)$ we see from the final expression that the answer is 86, the remainder we obtained when we divided $f(x)$ by $(x+3)$. This, in essence, is that the remainder theorem says. In this case it says that the remainder when $f(x)$ is divided by $(x+3)$ is $f(-3)$.

Now we have a rough idea of what the remainder theorem says, let's look in more general terms so we can build up a good grasp of the remainder theorem for a general polynomial. Consider the polynomial $f(x)$ divided by $(x+b)$ for some non-zero [integer[20]] b, the remainder would have to be just a number[21] so we can write a general expression for this division as follows:

$$f(x) \ = \ g(x)\,(x-b) \ + R$$

Where $g(x)$ is some unique polynomial[22] and R is the remainder.[23]

And from this expression we immediately see by putting $x = b$ that

$$f(b) = \ g(b)(b-b) + R = g(b) \times 0 + R = R$$

---

[20] b does not have to be an integer, but it is easier here to assume it is. Most cases of using the remainder theorem will involve a as an integer or a fraction; fraction in the case, for instance, of diving by $(2x+3)$

[21] Think about how algebraic long division works. If you get a $px + q$ during long division when dividing by something like $x - 3$ [for example] then you can always complete a further step in the long division. It is only when you reach a number alone in the long division that you can no longer divide by something like $x - 3$

[22] Can you explain why it is unique?

[23] As an aside, you should be able to argue that the highest power of x in $g(x)$ is one less than the highest power of x in $f(x)$

So, we can construct a version of the Remainder Theorem:

When a polynomial $f(x)$ is divided by $(x-b)$ the remainder is $f(b)$

This also works for division by $(px-q)$. What do you think the reminder is in this case and can you explain why?[24]

We can also look at general division here, for instance, when we divide a polynomial $f(x)$ by a quadratic x^2+bx+c [or even $(x-p)(x-q)$]. Before we explore this briefly, have a think about what you expect the general form of the remainder to be in this case…

Here is the general case for quadratics

$$f(x) \;=\; g(x)\,(x^2+bx+c) \;+\; mx+n$$

And you should be able to explain that the highest power of x in $g(x)$ is two less than the highest power of x in $f(x)$; and you should also be able to explain [and extend the concept] that the highest power of x in the remainder is [at most[25]] one less than the highest power of what we are dividing $f(x)$ by.

What is the relationship between the factor theorem and the remainder theorem? In simple terms, the factor theorem is a special case of the remainder theorem. Simply put: if the remainder on dividing $f(x)$ by $(x-b)$ is zero, then $(x-b)$ must be a factor of $f(x)$. Clever.

---

[24] When a polynomial $f(x)$ is divided by $(px-q)$ the remainder is $f(q/p)$ and this is because we can write $f(x) \;=\; g(x)\,(px-q)\;+R$ so we need to find what x makes $px-q=0$ and then substitute that value into $f(x)$

[25] It could be that $m=0$. And if both m and n are zero then the quadratic is a factor of $f(x)$

Exercise

The degree of a polynomial is the highest power of x that appears there. So, the degree of a quadratic is two, and that of a cubic is three, and so on. Here is a general expression for dividing some polynomial $f(x)$ by a polynomial $h(x)$:

$$f(x) = g(x)h(x) + r(x)$$

What can you say about the degree of $h(x)$, $g(x)$ and $r(x)$ in this expression. Be very careful about how you express your answer.

MM1.7

Qualitative understanding that a function is a many-to-one (or sometimes just a one-to-one) mapping.

Familiarity with the properties of common functions, including $f(x) = \sqrt{x}$ (which always means the 'positive square root') and $f(x) = |x|$

Let's start exploring what is meant by a "function". In simple terms a function is a mapping [or, better, it is a rule] from a set of input values to a set of output values. Not all algebraic expression are functions and, in this section, we will clarify what special features make something a function.

First, let's look at input and output values. We will be a little loose on notation here[26] and so we will use $f(x)$ to denote an algebraic expression and we will usually combine this with an expression for how the rule takes input values [the x values] to output values such as $f(x) = x^2 + 3$. And we will often also need to specify what values we are allowed to input into the function – usually this is just all x values on the x-axis which we call the "real numbers". We can be even more casual about things and write "$y = x^2 + 3$" and whilst this is not strictly perfect, it is ok as mathematicians will know what you are talking about.[27]

Now we have a rough idea of how our notation works, we can ask if ALL different expressions that we can construct using x are functions. So, is $f(x) = x^2$ a function, is $f(x) = \pm\sqrt{x}$ a function , is $f(x) = x^3 + 3x^3 - 17x + 242$ a function?

The answer is that not all these expressions are allowed to be called functions. We need one more restriction: an algebraic expression is a function, if for any given x value [in the collection of permitted inputs] the expression gives only one output value. So $f(x) = x^2$ is a function as any x value only gives one output value. But $f(x) = \pm\sqrt{x}$ is not a function as a single x value leads to more than one output value[28]: for instance, if we use $x = 16$ we get $f(x) = 4$ or -4.

[26] Strictly a function is denoted by f or g etc and the value of the function is denoted by $f(x)$ or $g(x)$ etc ; and sometimes functions are written in a different notation as $f: x \to x^2 + 3, x \in \mathbb{R}$

[27] In TMUA/ESAT we try to be as accurate as possible in the way we word our questions. We also make sure that the care we take does not get in the way of your understanding what we are asking. So we might talk about " the function f defined by…" or we might just talk about "the function f(x) =…" depending on the needs of the question.

[28] There is one exception: when $x = 0$

So here are the things we need to define a function

1. A rule [an algebraic expression] that maps input values to output values

2. A clear list of what we are allowed to input into the rule [this is called the Domain of the function[29]]

3. We need to be sure that for each input value in the domain there is only one output value.

We will look in a little more detail and point 2 and point 3 :

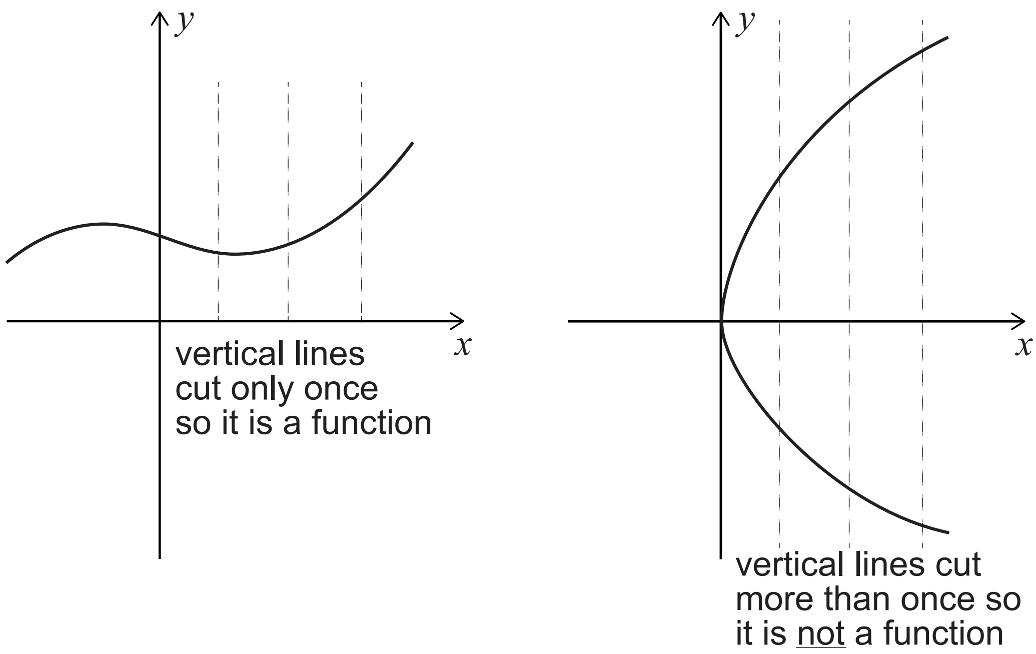
Point 2: usually the Domain of a function is just assumed to be the whole of the x-axis – that is all the real numbers. And often, if it is assumed that the Domain is just the whole x-axis, then it tends not to be mentioned explicitly. Sometimes you are expected to know the domain of functions so they are not mentioned [an example would be the log function - its domain is just positive x values]. And sometimes the domain of a function is deliberately restricted and then it is always mentioned explicitly [for instance, we could say $f(x) = x^2 + 3$ for $x \geq 2$]. In the TMUA/ESAT we tend to mention the domain most of the time and you should spend a moment each time thinking about why the domain is as mentioned [usually it is just to make the maths as precise as we can, but this is not always the reason, so you should always look at the details we put in a question].

Point 3. Whilst there is only one output for each input, it does not follow that two different inputs must have different outputs. For instance, consider $f(x) = x^2$ and $g(x) = x^3$. These are both functions – you should check this for yourself using the above definitions. For $g(x) = x^3$ each x input value gives us a unique output, so no two input values give the same output. But for $f(x) = x^2$ this is not true as both $f(2) = 4$ and $f(-2) = 4$. Even though two different x values give the same output here for $f(x)$ we know that $f(x)$ is a function as each x input value only gives us one output. Functions where each output is unique [occurs only once] are known as one-to-one functions, and any functions that have the same output for [at least some] different input values are known as many-to-one.

[29] The term "domain" for function is not on the TMUA/ESAT specification so you will not see it in the test. Nevertheless, you should know what it means as it is a very common term. Also, we recommend you explore the concept of range [You could also explore the idea of codomain but that can get a bit muddled when you read about it, so it is probably best left alone for the moment.]

There is another way to think about functions which is easier than all the formal stuff above, and that is to consider their graphs. You can tell if an expression is a function from its graph by looking to see if there are any x values that have more than one y value – if they do then they are not functions. In other words, any vertical line drawn on the graph [through an input x value] will only cross the function once.

Here are a few diagrams to help you see what we mean.

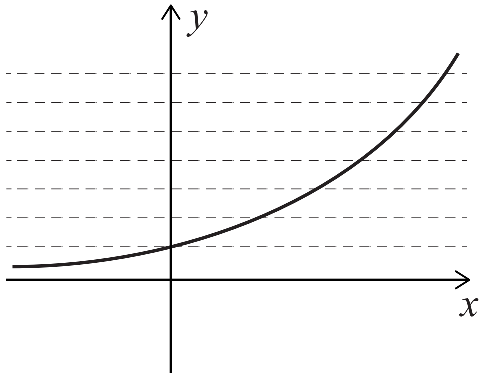


In addition, once we know we have a function using our “vertical line test” we can see if it is one-to-one or many-to-one.[30] If every horizontal line only crosses the function at most in one place then it is one-to-one, but if we can find a horizontal line that crosses the function more than once, then we have a many-to one function.

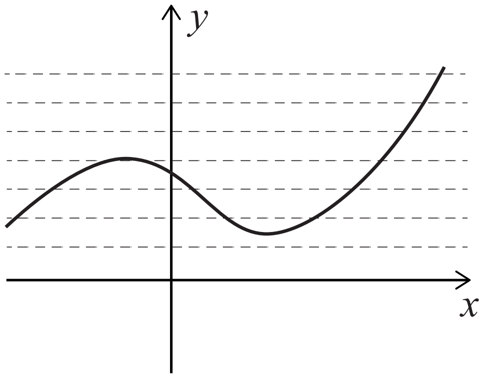
Here are diagrams to help you see what we mean:

[30] The idea of one-to-one and many-to-one is useful when exploring inverses of functions. Only one-to-one functions have inverses because we require the inverse of a function also to be a function. Note that the TMUA/ESAT does not require you to know in general about inverse functions.

one-to-one



many-to-one

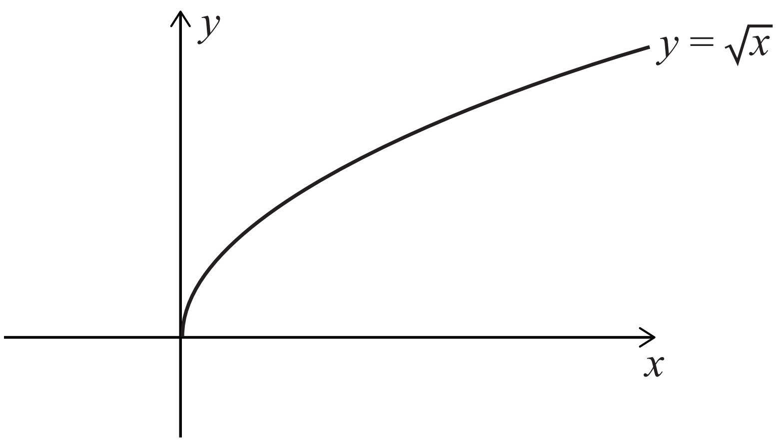


Let's now look at a couple of functions that you will probably meet in the TMUA/ESAT. We will also meet some functions later in these notes and there you should think carefully about domains – for instance, in the logarithm section and in the trigonometry section.

Here we will look briefly at $f(x) = \sqrt{x}$ and $f(x) = |x|$

First $f(x) = \sqrt{x}$ always means the positive square root of x. This is a standard maths convention, and we adopt it without comment in the TMUA/ESAT – whenever we use the square root sign, we will ALWAYS mean the positive square root and we will not comment on that in a question – you are expected to know it! The other thing to notice with this function, and again this is often just assumed, is that the input values will only be the positive real numbers and zero [i.e., $x \geq 0$].[31]

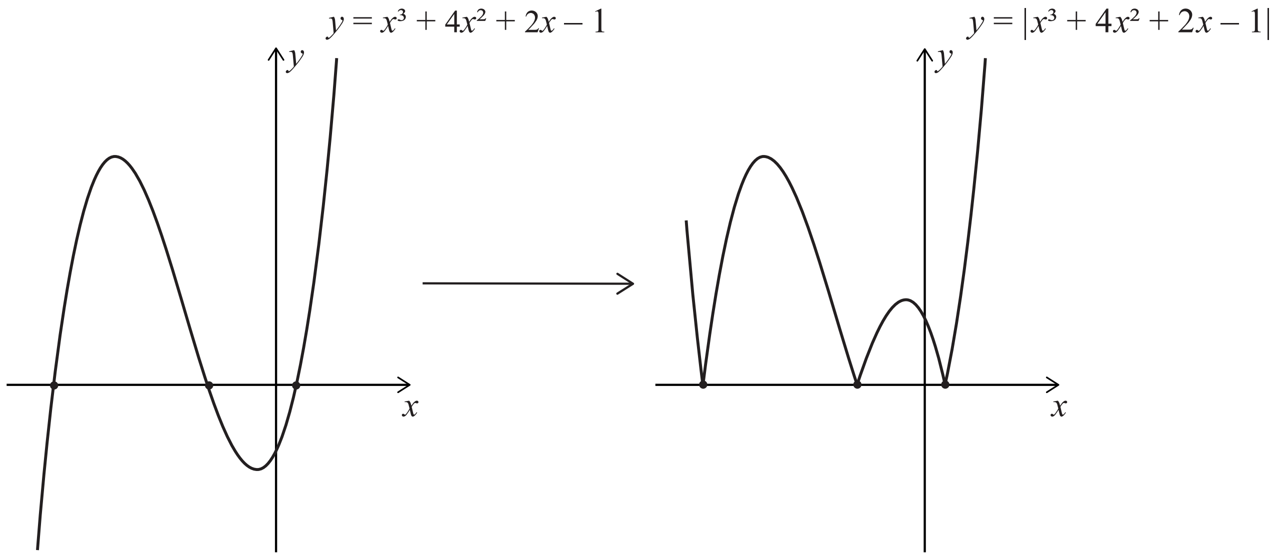
Here is the graph of $y = \sqrt{x}$ [you can see using horizontal and vertical line test that it is one-to-one]



[31] Complex numbers , which you might have met, are not on the TMUA/ESAT specification and we mostly ignore their existence. Occasionally when their existence might add complications to a question, we tend to add a comment to restrict the question to real numbers.

And now $f(x) = |x|$ or "the modulus function". This takes the positive value of whatever is inside the vertical straight lines. So $|7| = 7$, $|-2| = 2$, $|0| = 0$ and so on. You should make sure you can deal with expressions including the modulus both algebraically and graphically.

As an aside: a quick way to sketch $y = |f(x)|$ is to sketch $y = f(x)$ and then reflect everything that is below the x-axis in the x-axis but leave everything above the x- axis alone. Here is a diagram to show what we mean:



MM2. Sequences and series

MM2.1

Sequences, including those given by a formula for the n^{th} term and those generated by a simple recurrence relation of the form $x_{n+1} = f(x_n)$

For this MM2 section, there are three terms that you should know: series, sequence and progression. A sequence is an ordered list of numbers [often going on forever]; a series is a sum of a sequence; and a progression is just a general term that sits somewhere in between series and sequence. In the TMUA/ESAT we tend to use the term “progression” as a “catch-all” term because it is more neutral and helps us word question in way that makes them easier to understand.[32]

We expect you to be able to write out sequences of numbers given simple rules and to spot patterns in these sequences and then use the patterns to make further deductions. The most important thing when writing out a sequence that follows a rule and looking for patterns is knowing how many terms to write and how to tell when a pattern in the sequence is emerging. [33]

For instance, let’s look at the sequence generated by the recurrence relationship:

$x_0 = 1$ and $x_1 = 2$ and $x_{n+2} = |x_n - x_{n+1}| \quad n \geq 0$

We can write out the first four terms: 1,2,1,1.

From this it is a little bit tempting to conclude that the sequence just repeats 1,2,1, 1,2,1,1,2,1…. And it is also a little tempting to conclude that the sequence just goes on as 1,2,1,1,1,1,1,1 . But do we have enough information in the first four terms to justify either of our conclusions? The answer is, of course, “No”. The recurrence relationship we have looked at makes each term in the sequence [from x_2 onwards] depend on the previous two terms, so we need to see repeat in two adjacent terms before we can start to understand what the sequence will do.

[32] We are occasionally loose with our terminology, but it is always very clear what we mean. For instance, we might talk about the first term in a series when we should talk about the first term in a sequence whose sum is the series etc.

[33] There is a whole list of integer sequences: https://oeis.org/ ; the sequence we have used is an example/modification of sequence number A110044 in the list.

If we write a few more terms, we can see what pattern emerges: 1,2,1,1,0,1,1,0,1,1…

The message you need to take away from this brief example is that you must always make sure you have written out enough terms to justify your conclusions and the number of terms you need to write out to spot any patterns that might emerge is dictated by the structure of the recurrence relationship.

As an aside, we could start to shape a TMUA/ESAT question from this sequence. We could ask you to find $\sum_{n=0}^{n=100} x_n$. Let's look briefly at how we might go about this.

When dealing with "sigma" notation there are a number of things you need to do: first pay attention to the limits – in this case we start at x_0 and finish at x_{100} so there are 101 terms in the sequence [careful as it is easy but wrong to assume there are 100 terms[34] !]. Second, it is generally a good idea to write out the first few terms from such a sum to help get a feel for what is being asked [this is actually a very good idea generally, especially when the sum is a more complicated expression. You can see those sorts of expressions frequently in STEP mathematics questions – for past papers: https://www.ocr.org.uk/students/step-mathematics/preparing-for-step/]. And third, because we will be working out the sum using the patterns we have noticed from the recurrence relationship, it is useful to think about what happens at the end of the sum [i.e.,x_{98} , x_{99} , x_{100}]:

$$\sum_{n=0}^{n=100} x_n = \underbrace{1}_{x_0} + 2 + 1 + \underbrace{1}_{x_3} + 0 + 1 + \underbrace{1}_{x_6} + 0 + 1 \ldots\ldots.. + \overbrace{\underbrace{1}_{x_{3k}} + 0 + 1} \ldots\ldots \underbrace{1}_{x_{99}} + 0$$

We have spotted that the sequence can be broken into blocks of 3 [and the first term in each block of 3 is x_{3k}] and the last two terms are the first two terms in the block of 3. We now have enough information to work out the sum fairly easily – and we leave that as "an exercise for the reader".

[34] Explore the "fence post" issue in mathematics. We mention this again below.

MM2.2

Arithmetic series, including the formula for the sum of the first n natural numbers.

In the TMUA/ESAT we expect you to know what an arithmetic series [and sequence] is and recognise one when it appears. We often refer to these as Arithmetic Progressions [APs] in the TMUA. You should also know and understand [i.e., be able to derive[35]] the standard formulae and terminology :

$$\text{first term } = a$$

$$\text{common difference} = d$$

$$(\text{nth term }) \ u_n = a + (n-1)d$$

$$(\text{sum to n terms}) \ \mathrm{S_n} \ = \frac{n}{2}(2a + (n-1)d) = \frac{n}{2}(a + a + (n-1)d) = \frac{n}{2}(u_1 + u_n)$$

The last expression for the sum can be thought of as saying :

$$S_n = n \times \left(\frac{u_1 + u_n}{2}\right) = \ number\ of\ terms\ \times average\ value\ of\ terms$$

And finally, it is useful to note that $u_{n+1} - u_n = d$

We can create new arithmetic series by adding two [or more] series together:

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

$$T_n = A + (A + D) + (A + 2D) + \cdots + (A + (n-1)D)$$

And then we see that the sum is

$$S_n + T_n = a + A + (a + A + d + D) + \big(a + A + 2(d + D)\big) + \cdots + (a + A + (n-1)(d + D))$$

which is the sum of an arithmetic sequence with first term $a + A$ and common difference $d + D$.

[35] We will not test your ability to derive standard formulae in the TMUA/ESAT; but we would recommend, for your general mathematics education, that you know how to derive every formula in the TMUA/ESAT specification as the derivation will give you insights into the structure of the topic. And by “knowing how to derive” we don’t mean that you have just learnt the derivation, the important point is understanding it. These derivations can be found in many textbooks and on the internet too.

Exercise

We could also look at any linear combination of two such sequences: $\alpha u_n + \beta v_n$; verify these are also arithmetic sequences. Does this work for linear sums of more than two sequences [e.g., $\alpha u_n + \beta v_n + \gamma w_n$]?

There are lots of questions on arithmetic sequences in the TMUA/ESAT past papers. We recommend you work through the past papers first [initially under timed conditions] and if you get stuck, study the detailed worked answers we have supplied for each TMUA/ESAT paper.

MM2.3

The sum of a finite geometric series.

The sum to infinity of a convergent geometric series, including the use of $|r| < 1$

In the TMUA/ESAT we expect you to know what a geometric series [and sequence] is and recognise one when it appears. We often refer to these as Geometric Progressions [GPs] in the TMUA. You should also know and understand the standard formulae and terminology :

first term $= a$ [sometimes this is written as ar^0 to fit in with the formula for u_n]

$$\text{common ratio} = r$$

$$(\text{nth term})\ u_n = ar^{n-1}$$

$$\text{sum to n terms } \mathrm{S_n} = \frac{a(1-r^n)}{1-r}$$

$$\text{sum to infinity } \mathrm{S}_\infty = \frac{a}{1-r} \qquad \text{valid when } |r| < 1$$

And you should be comfortable using Σ notation and GPs:

$$S_n = \sum_{k=1}^{n} ar^{k-1} = \sum_{k=0}^{n-1} ar^k$$

Note how we have written the same expression in two different ways using the notation.[36]

There are some additional useful formulae/techniques you should keep in mind:

The ratio of adjacent terms is fixed : $\frac{u_{n+1}}{u_n} = r$ so $\frac{u_{n+1}}{u_n} = \frac{u_n}{u_{n-1}}$ etc

You can derive new progressions from a given GP :

Given $S_n = a + ar + ar^2 + ar^3 + ar^4 + ar^5 \dots \dots \dots$

[36] Whilst it takes some time to get used to Σ notation, you should make some effort to be fluent using it as it will appear frequently during your higher studies and it will usually be assumed you are familiar with it and can manipulate it easily. You might eventually encounter expression with sums of sums $\Sigma\Sigma$ and even sums of sums of sums $\Sigma\Sigma\Sigma$ etc., etc.

Then we can replace r by $-r$ to get

$$S_n(-r) = a - ar + ar^2 - ar^3 + ar^4 - ar^5 \ldots\ldots\ldots..$$

and we note that $S_n(-r)$ is also a GP with first term a and common ratio $-r$ and

$$S_\infty = \frac{a}{1+r}$$

Using this, we can also create other GPs:

Consider this progression:

$$\frac{1}{2}\left(S_n + S_n(-r)\right) = a + ar^2 + ar^4 + ar^6 \ldots\ldots\ldots$$

Which is a GP with $u_1 = a$ and a common ratio of r^2. What is its sum to infinity ?

You can also explore $\frac{1}{2}\left(S_n - S_n(-r)\right)$

And we can generate new GPs by raising each term in S_n to a power Here are some examples:

Example 1

Square every term :

$$S_n(squared) = a^2 + a^2r^2 + a^2r^4 + a^2r^6 + \cdots \ldots\ldots.$$

So $S_n(squared)$ is a GP with $u_1 = a^2$ and a common ratio of r^2.

What is $S_\infty\ (squared)$?

Example 2

Raise every term to power of k [we will assume k is a positive integer]

$$S_n(power\ is\ k) = a^k + a^k r^k + a^k r^{2k} + a^k r^{3k} + \cdots \ldots \ldots$$

Verify that $S_n(power\ is\ k)$ is also a GP. What can you say about $S_\infty(power\ is\ k)$ when $|r| < 1$? What happens if k is any positive number ? And what about if $k < 0$ and $0 < r < 1$?

We can also look at sum of part of a GP :

Given $S_k = a + ar + ar^2 + ar^3 + ar^4 + \cdots + ar^{k-1}$ find an expression for $ar^m + ar^{m+1} + \cdots + ar^n$ [where we assume $n > m$]

When exploring these sorts of expression, you have to be very careful with the ends of the sequence as it can be easy to write the wrong thing when simplifying. For instance , it is tempting to write this expression as $S_n - S_m$ but that would be wrong [why ?]. And also, for instance, we might ask how many terms are there in the sum $ar^m + ar^{m+1} + \cdots + ar^n$? Is it $n - m$ terms[37] ?

There are lots of ways of tackling this sum and how you approach it might depend on how the problem you are dealing with is set up:

Method 1

Treat is as the difference of two sums:

$$ar^m + ar^{m+1} + \cdots + ar^n = S_{n+1} - S_m$$

Note the $n + 1$ in S_{n+1}

Method 2

Factorise out r^m and treat what is in the bracket as a simple GP with $u_1 = a$ and common ratio r and number of terms $n - m + 1$

[37] NO! it is $n - m + 1$ terms. Can you see why ? This is an example of the fence post error – it is very easy to make this error, so explore what it is and think about how you would avoid making it when answering questions [I still use counting on my fingers sometimes to be sure I am getting things correct if I fear I might make a fence-post error.]

$$ar^m + ar^{m+1} + \cdots + ar^n = r^m(a + ar + \cdots + ar^{n-m})$$

Or even better, factorise out ar^m :

$$ar^m + ar^{m+1} + \cdots + ar^n = ar^m(1 + r + \cdots + r^{n-m})$$

Method 3

Treat it as a new GP with first term ar^m and common ratio r and number of terms $n - m + 1$

$$ar^m + ar^{m+1} + \cdots + ar^n = ar^m + (ar^m)r + (ar^m)r^2 + \ \ldots + (ar^m)r^{n-m}$$

Exercise

for each of the above methods, work out the sum and verify all three methods give you the same answer. This is more an exercise in being careful than anything else – you should not skip this exercise thinking you can do it easily !

There are lots of questions on geometric sequences in the TMUA/ESAT past papers. We recommend you work through the past papers first [initially under timed conditions] and if you get stuck, study the detailed worked answers we have supplied for each TMUA/ESAT paper.

MM2.4

Binomial expansion of $(1+x)^n$ for positive integer n, and for expressions of the form $\left(a+f(x)\right)^n$ for positive integer n and simple $f(x)$.

The notations $n!$ and $\binom{n}{r}$.

The binomial theorem is quite rich mathematically and there are lots of different ways we can approach it. For examinations, the best way is usually just to know the formulae and their quirks; but we do not recommend that you ever learn your mathematics in a way that sidesteps understanding. Here we will start by telling you what you need to know about the binomial expansion for the TMUA/ESAT and give you a few tips; then we will look in more detail at how the binomial expansion works.

What we expect you to know: in simple terms, we expect you to be able to do two things: one, calculate values of $\binom{n}{r}$; and two, work out any term in expressions such as $(1+x)^n$ and, more generally, in expressions of the from $\left(a+f(x)\right)^n$.

For instance, we might ask you to find the constant term in the expansion of

$$\left(\frac{1}{x^2}-3x^3\right)^{10}$$

All we have really done here so far is reiterate what the specification says. But let's look at what this entails in a little more detail:

For simple cases such as the expansion of $(1+x)^n$ for smallish n, you will probably have learnt how to use Pascal's triangle. You might even have used Pascal's triangle when tackling slightly more complicated expressions such as $\left(\frac{1}{x^2}-3x^3\right)^{10}$. This approach will always work, but sometimes can be very slow and time consuming; for instance, what if you were asked for the first five terms [in increasing powers of x] of the expansion of $(2+3x)^{17}$? In that case, finding the 17th row in Pascals triangle and writing out the correct expression will take some time. Is there a quicker and slicker method? Yes! : using the Binomial expansion directly. Using the Binomial expansion is easier than it looks once you understand the patterns in the expansion.

Let's start by writing out the Binomial expansion in full and then take it to pieces so we can see how easy it is to use:[38]

$$(a + f(x))^n = \sum_{k=0}^{n} \binom{n}{k} a^k [f(x)]^{n-k}$$

Whilst this looks a little daunting to start with, it is actually very easy to use once you know the patterns you need to think about. Let's work a couple of examples to see how we can use this in practice to find any term in any expansion quickly and effortlessly:

Example 1

Find the term with x^7 in the expansion $(3 + 2x)^8$

We build the answer in stages so you can see how the binomial expansion works in practice:

First as we need x^7 we know we must be looking for the term that has $(2x)^7$ in it so we write that down:

$$(2x)^7$$

Note that both the 2 and the x are raised to the power of 7. That is important.[39]

Then we need to work out what power of 3 goes with $(2x)^7$ and we use the fact that the sum of the powers of the individual terms always add to the overall power which is 8 here. So, we must have 3^1 together with $(2x)^7$ because $1 + 7 = 8$. So, we write

$$3^1(2x)^7$$

And finally, we need to work out what $\binom{n}{r}$ applies to this. That is easy to do. The n value is always the power of the bracket so here $n = 8$, and the r can be either the power of the 3, which is 1, or the power of the $(2x)$ which is 7. That might seem odd that we can use either $\binom{8}{1}$ or $\binom{8}{7}$ but it is ok as both have the same value [think, for the moment, of the symmetry of Pascal's triangle[40]]. So, we can write the answer as

[38] Or we could write this as: $(a + f(x))^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} [f(x)]^k$; check you understand why both forms give the same result.

[39] A common error is to write $2x^7$ without the bracket and so forget that the 2 must also be raised to the power of 7.

[40] Or you can prove it using the definition of $\binom{n}{r}$. Try to prove $\binom{n}{r} = \binom{n}{n-r}$

$\binom{8}{1}\, 3^1(2x)^7$ or $\binom{8}{7}\, 3^1(2x)^7$

You can then work out the numerical value of the coefficient easily.

So, in summary, the patterns you need to recall are

1. The powers of the terms always add to the power that the bracket is raised to.
2. The number n in the top of the $\binom{n}{r}$ is the power that the bracket is raised to.
3. The r value in $\binom{n}{r}$ can be either of the powers appearing in the expression.

Let's look at a second example that illustrates a mistake that students often make when dealing with the Binomial expansion [and you, of course, will not make this mistake!]:

Example 2

Find the coefficient of x^5 in $(2-3x)^7$

Following our rules above it is very tempting [but wrong!] to write the answer as $\binom{7}{5}2^2 3x^5$

This is wrong for two reasons – one serious and one less serious. The less serious reason is that this is not a coefficient as it still has x^5 in it, but that is a forgivable error. The major issue is the way we have written the power-of-5 bit in the expression. What we should have realised is that we need to look at $-3x$ all raised to the power of 5: that is, both the minus sign and the 3 need to be raised to the power of 5 as well. So, the correct expression [that will give us our coefficient] is $\binom{7}{5}2^2(-3x)^5$. We will leave you to work out the answer from there.

How the binomial expansion works.

This section is not something we expect you to know for the TMUA/ESAT so you can skip it if you want. We are going to set out a brief explanation of how the binomial theorem works.

We start by looking at what $\binom{n}{r}$ means and we shall do this using a simple example:

Example

Imagine you have five letters A B C D E. From these five letters you want to find how many different collections of 3 letters you can get but you don't want the order of the letters to be important. So, for instance, ABC is a collection and so is ACB and they count as the SAME collection as they contain the same three letters [think of a collection here as bag of three letters all jumbled rather than the three letters neatly laid one after the other on a table].

How can we work out how many different collections of 3 letters we can get from the 5 letters A B C D E ? [that is how many ways we can chose 3 things from 5 things]. Let's work out how we might go about it. We could start with three boxes and see how many ways we can fill them with three different letters from A B C D E:

| Box 1 | Box 2 | Box 3 |

We have 5 choices of letter for box 1; and then once we have chosen box 1, we have 4 letters left for box 2; and once we have chosen box 2's letter, we have 3 choices left for box 3. So it appears that we have $5 \times 4 \times 3$ choices overall for filling the three boxes. Does that mean we can get $5 \times 4 \times 3 = 60$ different collections of 3 letters chosen from A B C D E? The answer is no, as we will have chosen the same three letters a number of times: for instance, one of choices must be A B C but another choice will be A C B and another choice will be B A C and so on. In other words, we will have multiple copies of each collection of letters amongst our 60 different choices we made. How do we deal with this ?

Let's work out how many times we must have picked out a collection of three letters. We will work this out for the collection A B C. How many ways can we choose the letters A B C in order? In other words, how many ways can we order the letters A B C? We can think of this in a similar way to above: we have 3 choices of the first letter we choose [A or B or C] and then we have 2 choices for the second letter we choose, and 1 choice for the final letter. So, we have $3 \times 2 \times 1$ ways of choosing the letters A B C is $3 \times 2 \times 1 = 6$. We can list these :

ABC ACB BAC BCA CAB CBA

So, let's stop and take a look at what we have worked out. We wanted to see how many ways we can select 3 letters from 5 letters without worrying about the order. We discovered we could select 3 letters in 60 different ways, but we also noticed that the same 3 letters could be chosen 6 different ways as we chose them in order. So, every 6 of the 60 we chose will be the same collection, so we only really have 60 divided by 6 [that is, 10] different collections of 3 letters we can choose [can you list them?]

Let's write the whole calculation out in one go: $\frac{5\times4\times3}{3\times2\times1} = 10$

We notice that the top is almost 5! And the bottom is 3! We can use this to write the expression another way as $\frac{5\times4\times3\times(2\times1)}{(3\times2\times1)\times(2\times1)} = \frac{5!}{3!2!}$ and this is just $\binom{5}{3}$

Often $\binom{5}{3}$ is spoken as "5 choose 3" for obvious reasons – and sometimes it is written at ${}^5C_3$ which can also be read as "5 choose 3" even though the C actually stands for "combination".

We can now look at the general case and work out how many ways we can choose a collection of r things from a collection of n different things. This takes a little time to grasp but the effort is worth it. If you don't like the brief explanation we have given here, you can look elsewhere at the topic of "permutations and combinations" to find an explanation that suits you.[41]

We do exactly the same as above in stages. First, we note that if we have r boxes than we have n choices for the first box, $n-1$ choices for the second box and $n-2$ choices for the third box and so on. If we draw out n boxes, we need to fill the first r of them and leave the remaining boxes empty. There will be $n-r$ empty boxes and r filled boxes.

$$\overbrace{\blacksquare\blacksquare \ldots \blacksquare\blacksquare}^{r\ boxes} \mid \underbrace{\blacksquare\blacksquare \ldots \blacksquare}_{n-r\ boxes}$$

How many choices does this give us in this case: the answer must be

$$n \times n-1 \times n-2 \times \ldots \times n-r+1$$

This is like the $5 \times 4 \times 3$ above.

And we can write this as follows:

[41] The topic of "perms and coms" is often quite tricky to grasp when first met - and different people respond in different ways to different explanations. You might find that you just don't grasp the way we have explained things here, and if that is the case then you will need to search around for other ways of unpacking the ideas. It is worth spending some time thinking it through as it is a very useful topic and one that is worth understanding deeply. Spend time thinking about the concepts and also practise lots of questions until you get the hang of how the ideas work in real examples.

$$\overbrace{\underbrace{\blacksquare\blacksquare \ldots \blacksquare\blacksquare}_{FULL}}^{r\ boxes} \,|\, \overbrace{\underbrace{\blacksquare\blacksquare \ldots \blacksquare}_{EMPTY}}^{n-r\ boxes} = \frac{\overbrace{\blacksquare\blacksquare \ldots \blacksquare\blacksquare|\blacksquare\blacksquare \ldots \blacksquare}^{n\ boxes}}{\underbrace{\blacksquare\blacksquare \ldots \blacksquare}_{n-r\ boxes}} = \frac{n!}{(n-r)!}$$

And this is just like above where we used

$$5 \times 4 \times 3 = \frac{5 \times 4 \times 3 \times 2 \times 1}{2 \times 1} = \frac{5!}{(5-3)!}$$

But we recall that each of these choices contains repeats. We have r objects chosen lots of times but each in a different order – just like above where we had ABC and ACB etc. How many ways have we chosen the same r objects ? The answer is just as above - it must be $r!$ different ways. So, in the choices $\frac{n!}{(n-r)!}$ we have counted each set of r objects $r!$ times so to work out how many different collections we have, we need to divide by $r!$ [just like we had 60 divided by 6 above]

This gives us the number of ways of choosing r objects from n objects :

$$\binom{n}{r} = \frac{n!}{(n-r)!} \div r! = \frac{n!}{(n-r)!\,r!}$$

We can see how each bit of the equation works. The $n!$ divided by $(n-r)!$ is the number of ways we can fill the first r boxes out of a set of n boxes; and the $r!$ is the number of ways we can pick the same collection of r objects in order in those first r boxes.

At this stage it is useful to revisit something we mentioned in an example above. We mentioned that the symbol $\binom{n}{r}$ has some symmetry – so for instance $\binom{7}{2}$ is the same as $\binom{7}{5}$. We can now explore why this is the case using the ideas we have set out here. If you are asked how many ways you can choose 2 objects from 7 objects you have two ways you can work out the answer. You can choose all the different sets of two objects you can find and count them [that will give you $\binom{7}{2}$]; or you can think about how many ways you can choose 5 objects and throw them away to leave two objects behind [that is $\binom{7}{5}$]. A little thought shows you these must be the same. Count how many ways you can choose r objects from n objects and keep them, or how many ways you can choose $n-r$ objects from n object and throw them away. In both cases you are left with all different collections of r objects chosen from n objects .

Now how does this lengthy digression about choosing 3 objects from 9 objects or 2 objects from 7 objects etc. help us understand how the binomial expansion works ? Let's look at an example to understand how they are related.

Consider $(2+3x)^5$. We will write this out in full as

$$(2+3x)(2+3x)(2+3x)(2+3x)(2+3x)$$

Let's work out roughly how we might go about multiplying out these brackets:

If we multiply out term by term, we need to choose either 2 or $3x$ from each bracket and make sure we have got all the different combination [doing this long-hand will take ages – you can try if you want]

When we multiply out term by term, we will have things like:

(2 from 1st bracket) x (2 from 2nd) x (2 from 3rd) x ($3x$ from 4th) x ($3x$ from 5th bracket) = $2^3(3x)^2$

and we can also have

(2 from 1st bracket) x ($3x$ from 2nd) x (2 from 3rd) x ($3x$ from 4th) x (2 from 5th bracket) = $2^3(3x)^2$

and also

($3x$ from 1st bracket) x (2 from 2nd) x (2 from 3rd) x ($3x$ from 4th) x (2 from 5th bracket) = $2^3(3x)^2$

and so on….

In other words, the term $2^3(3x)^2$ comes from all sorts of different combinations – we need to pick three brackets to give us the 2's or two brackets to give us the $3x$'s.

How many different ways can we pick three 2's from 5 brackets ? We know the answer from above – it is how many ways can we choose 3 from 5 and that is $\binom{5}{3}$ [or we could look at the number of ways of choosing two lots of $3x$ from five brackets and use $\binom{5}{2}$]

So, the x^2 term in the expansion of $(2+3x)^5$ must be $\binom{5}{3}2^3(3x)^2$

Now we can see how the general term works in the binomial expansion:

$$\overbrace{(f(x)+a)(f(x)+a)(f(x)+a)\ldots(f(x)+a)}^{n\ brackets} = \sum_{k=0}^{n}\binom{n}{k}[f(x)]^k a^{n-k}$$

Let's look at one term on the right hand side $\binom{n}{k}[f(x)]^k a^{n-k}$ This term is, in essence, made up from choosing $f(x)$ from k of the brackets and then using the a from the remaining $n-k$ brackets [hence the powers must sum to n] and then seeing that we get k lots of $f(x)$ from n brackets in exactly $\binom{n}{k}$ different ways so there must be $\binom{n}{k}$ lots of the term $[f(x)]^k a^{n-k}$ in the expansion.

Coordinate geometry in the (x, y) plane.

MM3.1

Equation of a straight line, including

$$y - y_1 = m(x - x_1)$$

$$ax + by + c = 0$$

Conditions for two straight lines to be parallel or perpendicular to each other.

Finding equations of straight lines given information in various forms.

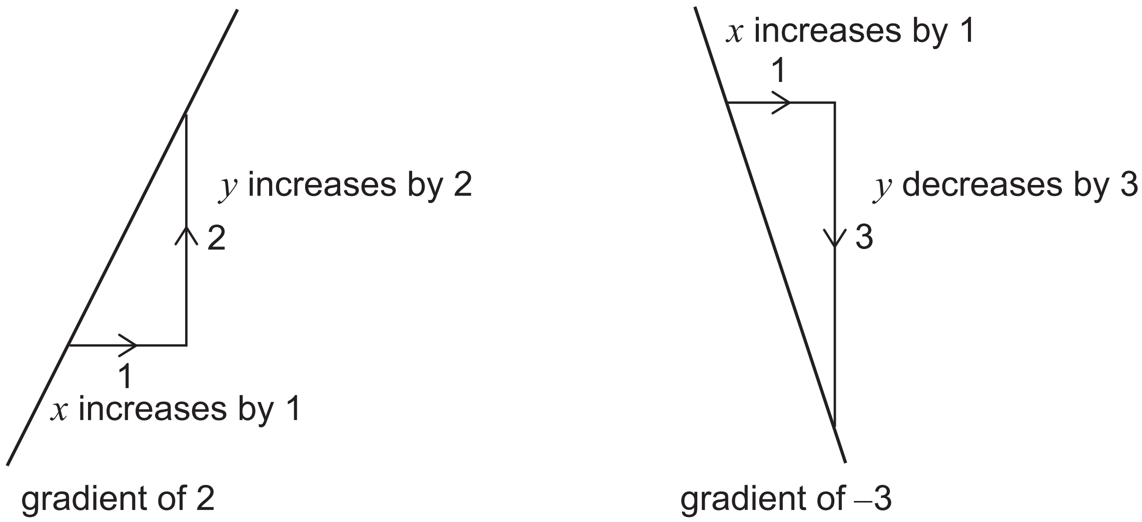
The specification here is self-explanatory as to what you need to know about straight lines. You should be comfortable dealing both algebraically and geometrically [i.e., graphically] with straight lines. In this section we will briefly explore most of the main ideas we expect you to know and add a few things to think about.

Let's start with the classic $y = mx + c$ and remind ourselves of what the m and the c represent for a diagram of a line whilst we explore things in a little more detail.

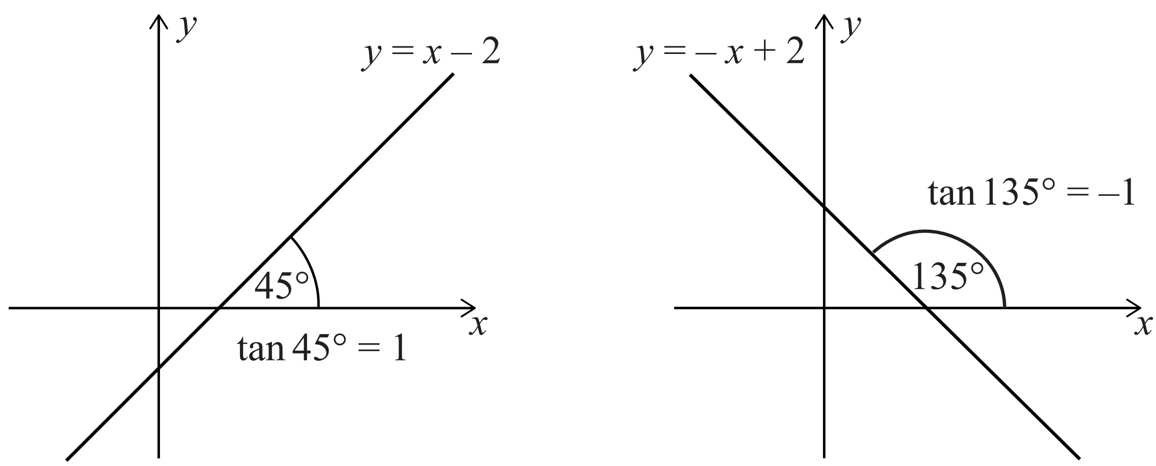
The m in $mx + c$ represents the gradient of the straight line. [You should be able to calculate the gradient of a straight line if you are given two points that it passes through.] But what does the gradient tell us about the line? The usual way of thinking about the m is as a measure of "steepness": the greater the value of m then the steeper the line; and if m is positive the line slopes from bottom left to top right, and if the m is negative is slopes from top left to bottom right.

But what do we really mean by steepness? We can think of steepness in a couple of interrelated ways: we can think of the gradient as telling us how much we have to go vertically to get back on the line for every 1 unit we move horizontally from a point on the line. So, if the gradient is 2 then we need to move vertically up by 2 for every 1 unit we move horizontally from the line; and if the gradient is minus 3 then we need to move down by 3 for every 1 unit we move horizontally from the line [see diagrams below].

Another way of thinking about this [which is more useful when we encounter straight lines as tangents to curves] is that, for instance, a gradient of 2 tells us that the y values are changing twice as fast as the x values – so as x increases by 1 the y values must increase by 2, and if x increase by 5 then y must increase by 10 and so on. Negative gradients, for instance -3, just tell us that as x increases by 1 then the y values *decrease* [the minus sign signals the decrease] by 3. So gradient is a "rate of change" telling us the rate at which y changes relative to x.



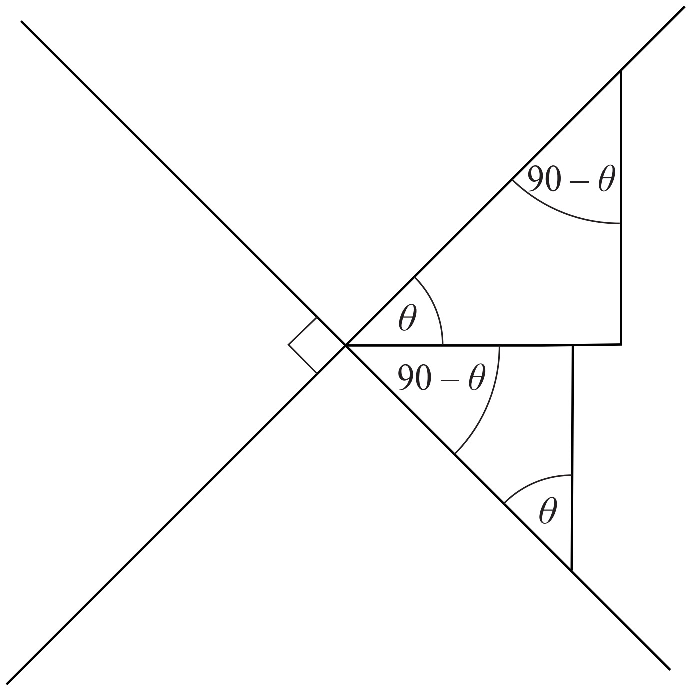
There is yet another way to think about the gradient which can be useful, that is to consider the gradient as the tan [that is the trigonometric tan] of the angle that the line makes with the positive x axis [see diagrams] and of this we assume the scales on the x-axis and the y-axis are the same. You should be able to see why this is the case if you consider how the gradient is calculated. So, the gradient of a line at 45 degrees should be $\tan 45 = 1$ and that is exactly what we expect with a line such as $y = x$. If the angle is 135 then we would expect the gradient to be $\tan 135 = -1$ and that is exactly what we expect with a line such as $y = -x$



There are a couple of special cases for gradients: horizontal lines [which have a gradient of zero] and vertical lines which don't strictly have a gradient but often it is said they have an infinite gradient or negative infinite gradient [this is all a bit of a fudge]. It is best to think of vertical lines [and perhaps horizontal lines] as special cases - vertical lines always have the equation $x = \textit{some number}$ and horizontal lines always have the equation $y = \textit{some number}$. You should also be aware that the x-axis has equation $y = 0$ and the y-axis has equation $x = 0$.

What can you say about the gradients of two parallel lines [we exclude horizontal and vertical lines for this discussion, but you should be able to spot when they come up as special cases]? They have the same steepness so they must have the same gradients. So we can say that lines $y = m_1x + c_1$ and $y = m_2x + c_2$ are parallel if and only if $m_1 = m_2$ [check you know why we write "if and only if" you might need to look at our *Notes on Logic and Proof* for TMUA paper 2]

What about lines that are perpendicular to each other ? Here you need to know [and understand] that the lines $y = m_1x + c_1$ and $y = m_2x + c_2$ are perpendicular if and only if $m_1m_2 = -1$. You should be able to see this easily by looking at the two similar triangles in the diagram below – convince yourself why $m_1m_2 = -1$ using this diagram



Exercise

[This exercise is a little outside of the TMUA/ESAT specification.] Use the idea that the gradient of a straight line is equal to the tangent of the angle it makes with the horizontal x-axis to explain why $m_1m_2 = -1$ if and only if the lines are perpendicular [exclude cases that involve vertical lines]. You can then ask yourself a general question – is there a condition on m_1 and m_2 when the lines meet at some other angles [60 or 45 degrees, for instance] and can you justify your answer?

We have spent some time discussing the gradient of a straight line so now we can turn to ask what the c represents in the $y = mx + c$. The answer is simple, and you should know it: the c is the value of y where the graph crosses the y-axis; this is often called "the y-intercept".

It is also useful to think about straight line graphs using graph-transformations [see later in these notes]:

Exercise

As an exercise, think about what each of the following transformations does to the graph of $y = x$ [y_1 and x_1 are constants; we could have called them something like p and q but we chose to vary the notation a little as you will need to get used to all sorts of notation as you learn more mathematics – text books and teachers are not always consistent in how they use notation. If you introduce notation into your maths that isn't given in a problem, make sure you make it VERY clear what your notation represents!]

$$y = x \quad to \quad y = mx$$

$$y = x \quad to \quad y - y_1 = x$$

$$y = x \quad to \quad y = x - x_1$$

Then combinations of these:

$$y = x \quad to \quad y - y_1 = mx$$

$$y = x \quad to \quad y = m(x - x_1)$$

$$y = x \quad to \quad y - y_1 = m(x - x_1)$$

Finally, you should be able to work out the equation of a line given two bits of information. What two bits of information do you think you will need to uniquely specify the equation of a line ? The answer is you will need:[42]

Case 1

The coordinate of one point on the line and the gradient of the line – i.e., point (x_1, y_1) lies on the line with gradient m.

To work out the equation you need to notice that the gradient of the line joining any point (x, y) on the line with (x_1, y_1) must be constant:

$$\frac{y - y_1}{x - x_1} = m$$

And then rearranging.

Or you can use $y = mx + c$ and substitute the point (x_1, y_1) into the equation to find c

Exercise

Use both the above approaches to find the equation of the line with the following gradients and points:

$m = 4$ point $= (3, 2)$

$m = -5$ point $= (5, 3)$

$m = -2$ point $= (-2, -4)$

Case 2

The coordinates of two distinct points that sit on the line (x_1, y_1) and (x_2, y_2) .

To work out the equation of the line, take a general point (x, y) on the line and calculate the gradient of the line [which is fixed no matter what] in two different ways:

$$\frac{y - y_1}{x - x_1} = \frac{y_1 - y_2}{x_1 - x_2}$$

and then rearrange this equation.

[42] With line questions it is very easy to make sign errors when putting coordinates into equations. Make sure you are careful !

Or you can work out the gradient first using $\frac{y_1-y_2}{x_1-x_2}$ [make sure you get the order of the x and the y on the top and the bottom the same way around – otherwise you will get the wrong sign for your gradient; and also make sure to put y on the top and x on the bottom] and then using $y = mx + c$ and finding c by substituting in one of the points (x_1, y_1) or (x_2, y_2) into the equation.

Exercise

Use both the above approaches to find the equation of the line with the following gradients and points:

$(0,0)$ and $(2,3)$ [is there a shortcut here?]

$(-2,5)$ and $(-4,\ -7)$

$(3,\ -7)$ and $(8,\ -7)$

Final thoughts

In this section we have used the equation of a line in the from $y = mx + c$. however, it is also common to see the equation of a line written in the form $ax + by + c = 0$ [It is unfortunate that c appears in both as its role is different in each equation – do not assume that the c in $ax + by + c = 0$ is the y-intercept!]. You should be able to move from one form to the other easily using algebra. Both forms are useful in different contexts and different education systems might put more emphasis on one form rather than the other.

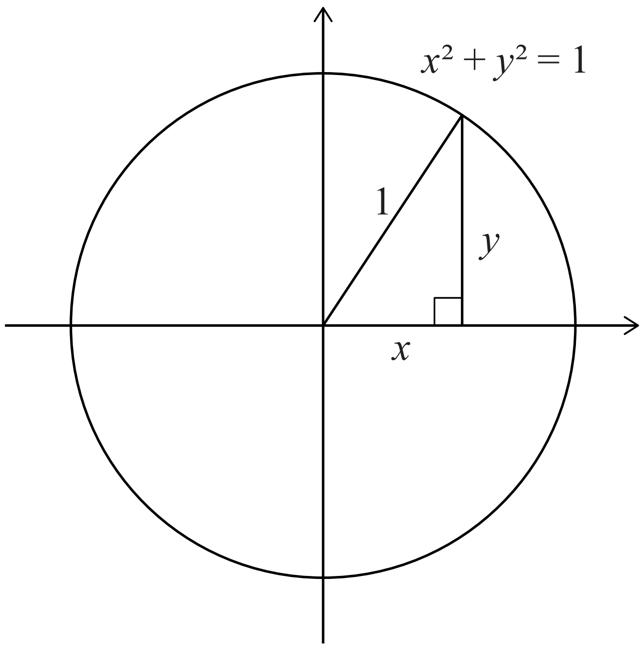
MM3.2

Coordinate geometry of the circle: using the equation of a circle in the forms

$$(x-a)^2+(y-b)^2=r^2$$

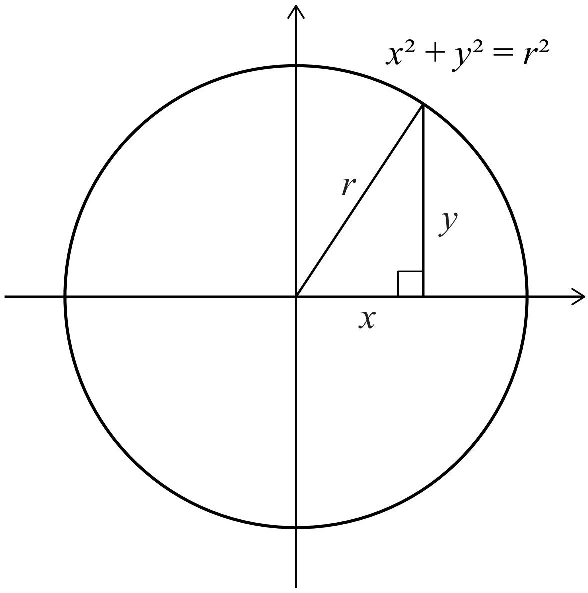
$$x^2+y^2+cx+dy+e=0$$

Imagine drawing a circle of radius 1 on the xy-plane with its centre at the origin and a radius of 1. What can we say about all the points that sit on this circle? They must all be a distance of 1 from the origin. We can use Pythagoras' theorem to express this idea in algebra [see the diagram] and when we do so we get the equation of a basic circle: $x^2+y^2=1^2$. Any (x,y) satisfying this equation is on the circle and any (x,y) not satisfying this equation is not on the circle.[43]



[43] As an aside, we can say all points inside the circle satisfy $x^2+y^2<1^2$, and all points outside the circle satisfy $x^2+y^2>1^2$

What about taking the same circle and making the radius bigger? If the radius is r then we see from Pythagoras' Theorem [and the diagram below] that the equation of this circle must be $x^2 + y^2 = r^2$

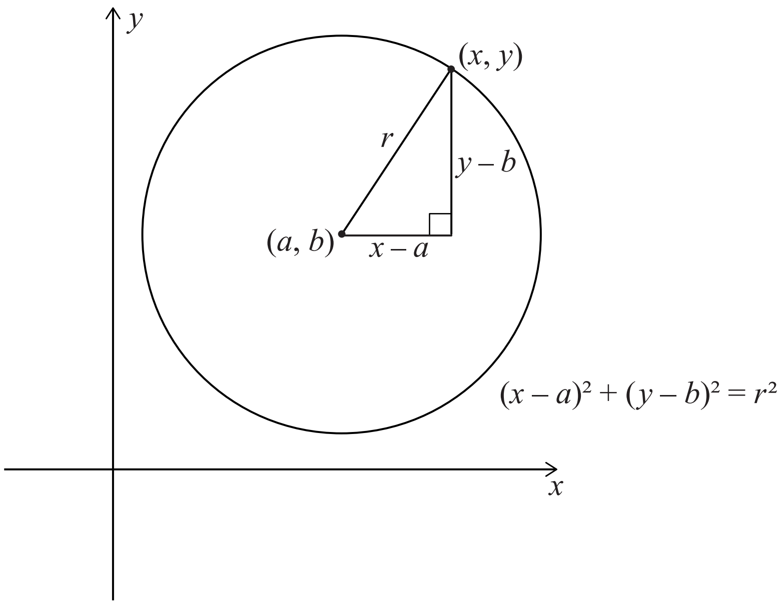


What if we want the centre of the circle to be somewhere else on the xy-plane? We will find the equation of the circle in that case in one of two ways: we either using Pythagoras' or using graph shifting:

Let's find the general equation of a circle of radius with its r centre at the point (a, b):

Pythagoras' tells us that all the (x, y) values that are a distance r from (a, b) will sit on the circle. So, we can use some geometry to work out the equation of the circle [see the diagram too] and we get

$$(x - a)^2 + (y - b)^2 = r^2$$



Or we can start with the circle of radius r centred at the origin which has equation $x^2 + y^2 = r^2$ and shift it horizontally by a and vertically by b [so it moves to have a centre at the point (a, b) and using standard graph shifting. Doing this, we get the equation :

$$x^2 + y^2 = r^2 \qquad \to \qquad (x - a)^2 + (y - b)^2 = r^2$$

You should be able to identify [very quickly] the radius and the centre of any circle equation you are given. For example the circle with equation $(x - 2)^2 + (y - 3)^2 = 25$ has its centre at $(2, 3)$ and its radius is 5 [because $5^2 = 25,$ and be very careful not to say its radius is 25 !!] . The circle $(x - 2)^2 + (y + 4)^2 = 18$ has its centre at $(2, -4)$ and a radius of $\sqrt{18} = 3\sqrt{2}$

There are other ways of writing the equation of a circle and you should learn to recognise them and be able to work out both the radius and centre of a circle given its equation. The most common alternative form of a circle is $x^2 + y^2 + ax + by + c = 0$. We will look at a few examples and use "completing the square" to find the centre and radius:

Example 1

Find the centre and radius of $x^2 + y^2 + 4x + 2y - 12 = 0$

We collect the x terms together and the y terms together:

$$x^2 + 4x + y^2 + 2y - 12 = 0$$

And then complete the square for the x terms and also for the y terms:

$$x^2 + 4x + y^2 + 2y - 12 = (x + 2)^2 - 4 + (y + 1)^2 - 1 - 12 = 0$$

And rearranging gives us our standard equation:

$$(x + 2)^2 + (y + 1)^2 = 17$$

So, the centre is at $(-2, -1)$ and the radius is $\sqrt{17}$

Example 2

Find the centre and radius of $x^2 + y^2 - 4x - 6y + 20 = 0$

We go through the same process as above – collect terms and complete the square to give:

$$(x-2)^2 - 4 + (y-3)^2 - 9 + 20 = 0$$

Which rearranges to give

$$(x-2)^2 + (y-3)^2 = -7$$

OH NO !!!!!

A moment's thought shows something has gone wrong. The left-hand side is the sum of squares so must always be ≥ 0 but the right-hand side is negative. Well, it turns out that there are no real x and y values that make this equation true and so it is not the equation of a circle. The lesson to learn here is that not every equation of the form $x^2 + y^2 + ax + by + c = 0$ is the equation of a circle. You should be able to work out what extra condition we need to place on a,b and c to ensure that the equation $x^2 + y^2 + ax + by + c = 0$ is a circle.

Example 3

Find the centre and radius of $2x^2 + 2y^2 - 4x - 8y - 19 = 0$

Well initially this doesn't look exactly like $x^2 + y^2 + ax + by + c = 0$. But we can easily divide by 2 to get $x^2 + y^2 - 2x - 4y - \frac{19}{2} = 0$ and then proceed as before to get :

$$(x-1)^2 - 1 + (y-2)^2 - 4 - \frac{19}{2} = (x-1)^2 + (y-2)^2 = \frac{29}{2}$$

Which has centre at $(1,2)$ and a radius of $\sqrt{\frac{29}{2}}$

Exercise

Why are the following not equations of circles?

$$2x^2 + 3y^2 - 4x - 8y - 19 = 0$$

$$2x^2 - 2y^2 - 4x - 8y - 19 = 0$$

$$x^2 + 2y + 5 = y^2 + 4x + 7$$

Given the equation $px^2 + qy^2 + ax + by + c = 0$ what conditions on p, q, a ,b and c will make it an equation of a circle ? [this is slightly tricker than it appears – be careful and think of what different cases there might be …]

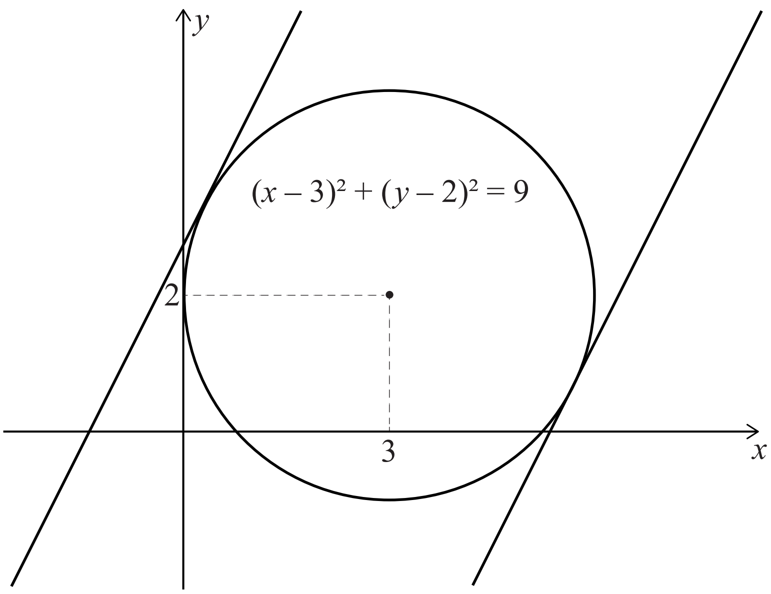
Finally, we will look at a couple of scenarios involving circles that you should be able to deal with. We will do this by working through a couple of examples.

Example 1

When is a line tangent to a circle?

Find the values of c for which $y = 2x + c$ is tangent to the circle with equation $(x-3)^2 + (y-2)^2 = 9$

This can be solved algebraically [or geometrically with some careful work] but it is good to sketch a picture to start with to get an idea of why the question asks for values [not just value] and to get a very rough idea of what these values might be. Here is a sketch:



Now we can start to think about how we should approach this. In simple terms, if a line is tangent to a circle, then it intersects it at only one point. So, if we try to solve the equation of the line and the circle simultaneously, we will need to look for the case that has only one solution. If you think about it for a moment, the number of ways a line can intersect a circle can be twice, or once, or not at all. And when you try to solve the equation of a line simultaneously with the equation of a circle, you will get a quadratic in x [or in y]. So, putting all this together suggests we try to solve simultaneously and then use the discriminant condition to work out the values of c that gives one [repeated] root of the quadratic we obtain [think why we expect the discriminant to be 0 for two different values of c].

Let's do that for this question: substitute $y = 2x + c$ into $(x-3)^2 + (y-2)^2 = 9$ to give $(x-3)^2 + (2x + c - 2)^2 = 9$ and multiplying out we get

$$x^2 - 6x + 9 + 4x^2 + 4(c-2)x + (c-2)^2 = 9$$

Rearranging

$$5x^2 + 2(c-7)x + (c-2)^2 = 0$$

And we want this to have one repeated root, so the discriminant condition requires

$$4(c-7)^2 - 20(c-2)^2 = 0$$

Which then gives

$$c - 7 = \pm\sqrt{5}\,(c-2)$$

And from this you can work out the two c values – you can see we have two lines in the diagram above, one for reach of the c values we found.

Example 2

What is the closest distance between a line and a circle?

Find the closest distance between the line $y = x + 11$ and the circle $(x+2)^2 + (y-3)^2 = 9$

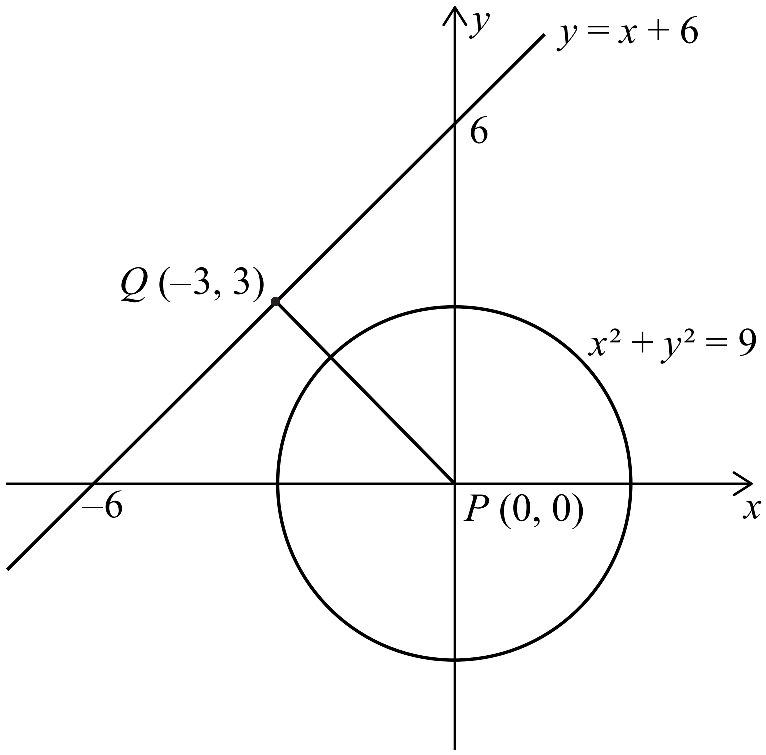
There are lots of ways to approach this question. We are going to use a mix of algebra and geometry. The first thing we are going to do [which is not necessary but just makes things easier to deal with] is to translate the circle so its centre is at the origin and translate the line the same way [convince yourself that this does not change the answer] We are going to replace x by $x-2$ and y by $y+\ 3$ in both equations to give :

$y+3\ =\ x-2\ +11$ and $(x-2+2)^2+(y+3-3)^2\ =\ 9$

Which simplify to

$y=x+6$ and $x^2+y^2=3^2$

We can sketch these and then use some [simple] geometry to find the shortest distance between the line and the circle:



From this diagram we can see that the line between $(0,0)$ and $Q(-3,3)$ will help us find the length that we seek. The length of the line between (0,0) and $Q(-3,3)$ is $3\sqrt{2}$ [it is $\sqrt{3^2+3^2}$] and the length of the radius we know is 3. So, the distance from the line to the circle [which is the value we seek] must be $3\sqrt{2}-3$

Exercise

For example 2, can you find other ways of finding the length required[44]?

How would you solve the problem if the equation of the line had a different gradient – e.g., find the shortest length between the line $y = 2x + 15$ and the circle $(x + 2)^2 + (y - 3)^2 = 9$. Try to solve it using a number of different methods. Which method do you think is best ?

[44] For instance: there is a formula for the shortest length of a point from a line. You could find the distance of the centre of the circle from the line and then subtract the radius length from the result. If you choose to take this approach, make sure you can prove and fully understand the equation that gives the shortest length between a point and a line. We do NOT expect you to know this formula in the TMUA/ESAT. You could also alter the problem in other ways using algebra: for instance, you could ask what r value makes $y = x + 6$ a tangent to the circle $x^2 + y^2 = r^2$ and then the answer to the question would be $r - 3$. Can you see why this would work ?

MM3.3

Use of the following circle properties:

The perpendicular from the centre to a chord bisects the chord;

The tangent at any point on a circle is perpendicular to the radius at that point;

The angle subtended by an arc at the centre of a circle is twice the angle subtended by the arc at any point on the circumference;

The angle in a semicircle is a right angle;

Angles in the same segment are equal;

The opposite angles in a cyclic quadrilateral add to $180°$;

The angle between the tangent and chord at the point of contact is equal to the angle in the alternate segment.

You will notice that we have included these circle theorems twice in the TMUA/ESAT specification. They appear both in the M section and also in the MM section. We have included them twice deliberately. We occasionally ask questions on circle theorems in the TMUA/ESAT and we are aware they are a topic that is often covered [perhaps briefly] in introductory maths classes and then forgotten. We do not want you to encounter a TMUA/ESAT question on circle theorems and then realise you cannot answer it because you have forgotten them. And so we have included them twice – in the M section because they appear in introductory maths classes – and in the MM section to remind you to review your knowledge of the theorems.

We expect you to know and be able to use all the theorems we list. You should also think about the converse[45] of some of the theorems. Whilst we don't expect you to know the proofs of the theorem, you really should make sure you can prove each theorem and that you have a deep understanding of the proof [i.e., you must not try to learn the proof, rather you need to *understand* what is going on – what are your assumptions, how does the proof work, what geometry is being used etc., etc.].

[45] See the *Notes on Logic and Proof*

There are all sorts of techniques you can use when tackling question that link to circle theorems. Here is a list of some techniques you should be comfortable using

1. Angle chasing – filling in all the angles you can using the theorems above and by looking for isosceles triangles [often made up of two radii] or right angles triangles [in the semicircle].

2. Rotating the diagram – this can often help you get insights into the question.

3. Adding lines – sometimes adding a tangent or diameter or some other line [e.g., a chord] helps you to find the solution.

4. Using dynamic methods – learning to move points around on your diagram in a way that does not affect the solution but makes the question easier to solve.

We would recommend you find a site online that helps you play around with circle theorems. We do not endorse any sites [so this is not an endorsement!] but we found Circle theorems very useful, especially for dynamic methods.

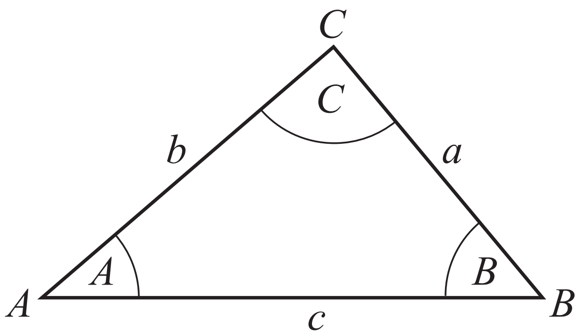
MM4. Trigonometry

MM4.1

The sine and cosine rules, and the area of a triangle in the form $\frac{1}{2}ab\sin$ C.
The sine rule includes an understanding of the ‘ambiguous’ case (angle-side-side).
Problems might be set in 2 or 3 dimensions.

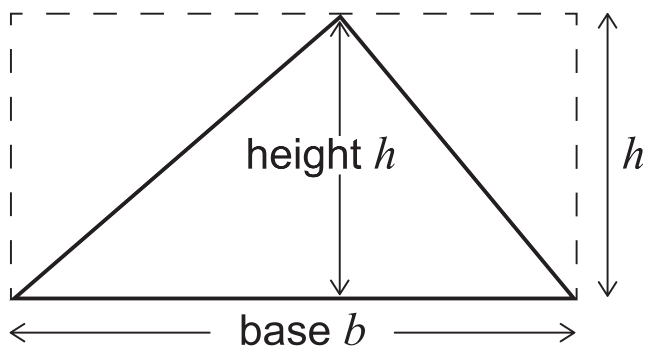
We will start this section by looking at the area of a triangle and then explore the sine rule and the cosine rule.

A quick note on labelling. We tend to label polygons anticlockwise [not always though – and in this section we have varied the labels we have used on triangles in each diagram to keep you on your toes!]. For triangles we label corners [and usually angles in the respective corners] with capital letters, and the sides opposite corners with corresponding lower-case letters:

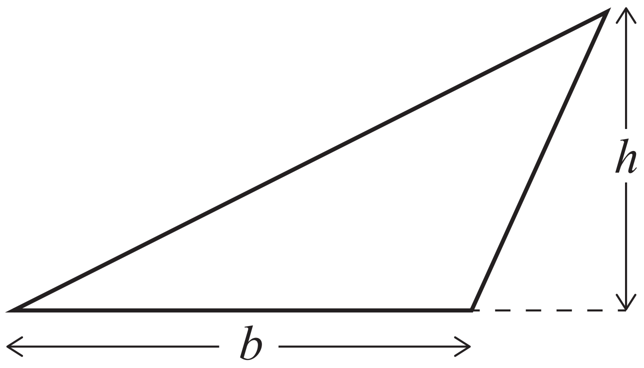


We start with the most basic formula for the area of a triangle. From the diagram below, you can see that the area of a triangle is exactly half that of the rectangle so the area of the triangle must be :

$$area = \frac{1}{2}\ base\ \times vertical\ height = \frac{1}{2}\ bh$$

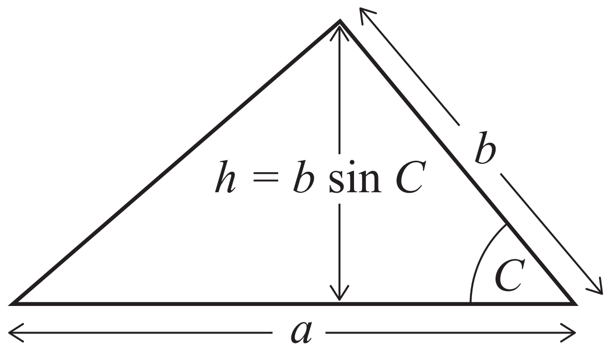


What about for a triangle where the top corner is not above the base?



The area is the same, but you must be careful to use vertical height [as shown] above the horizontal base and not any part of slanted height. Can you convince yourself using geometry that the area is still $\frac{1}{2}$ $base\ \times vertical\ height$ even when the triangle does not fit neatly inside a rectangle[46] ?

Let's now calculate the area of triangle another way using a little bit of trigonometry.



From the diagram you can see that the base of the triangle is length a and the height [using trigonometry] is $b \sin C$. The area of the triangle must be :

$$area = \frac{1}{2} \times base\ \times vertical\ height = \frac{1}{2} ab \sin C$$

This gives us another formula for the area of a triangle using the length of two sides and the angle between the two sides.

[46] Hint: you can add an identical triangle to make a parallelogram then chop up and rearrange the parallelogram slightly. What other ways can you use ?

From this equation we can easily derive the sine rule [which is essentially, as we shall see, a way of saying that the area of a triangle is the same no matter how you choose to calculate it].

Above we showed that the area of a triangle with base a was $\frac{1}{2}a\,b\sin C$. If we rotate the triangle to make the base b, then the area can be written as $\frac{1}{2}b\,c\sin A$ and if we rotate it again to make c the base, then we can see that the area can be written as $\frac{1}{2}c\,a\sin B$. These areas must all be the same, and so :

$$\frac{1}{2}a\,b\sin C = \frac{1}{2}b\,c\sin A = \frac{1}{2}c\,a\sin B$$

Which is essentially the sine rule in disguise. If we multiply by 2 and divide by abc we get :

$$\frac{\sin A}{a} = \frac{\sin B}{b} = \frac{\sin C}{c}$$

Which is the more usual form of the sine rule; it also appears the other way up :

$$\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C}$$

When do we use the sine rule?

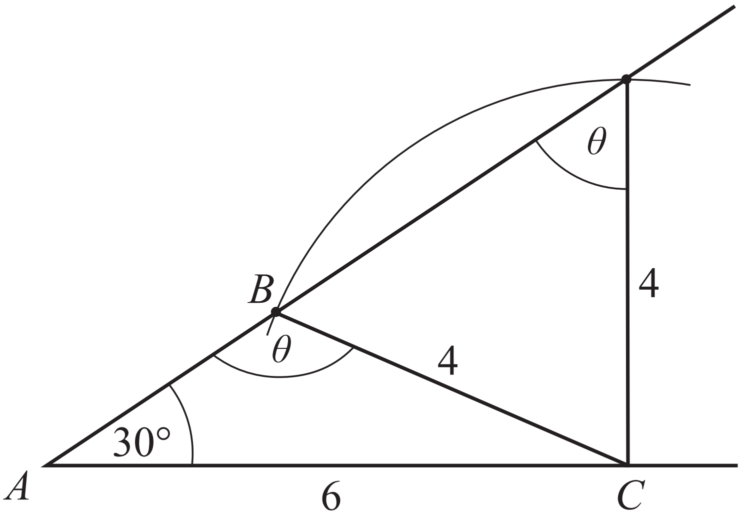
Given	**Find with one use of the sine rule**
Two angles and one side	Any other side [note, once given two angles, you can calculate the third as angles in a triangle add to 180]
Two sides and one angle [not between given sides]	Another angle [see below]

In some cases, there is more than one triangle that will fit a given set of data – the sine rule then gives us an “ambiguous” result. The ambiguity arises in some cases when finding another angle using the sine rule. This is because there are always two angles between 0 and 180 that have the same sine value.

Here is an example of this “ambiguity”:

Example

In triangle ABC, angle $A = 30$, $b = 6$ and $a = 4$. Find angle B



Using the sine rule we can write [note we tend to put the unknown on the top]:

$$\frac{\sin\theta}{6} = \frac{\sin 30}{4}$$

So, we get

$$\sin\theta = \frac{6\sin 30}{4} = \frac{3}{4}$$

We obtain two values for θ : 48.6 or $180 - 48.6$

Both angles are plausible as when added to 30 they do not exceed 180

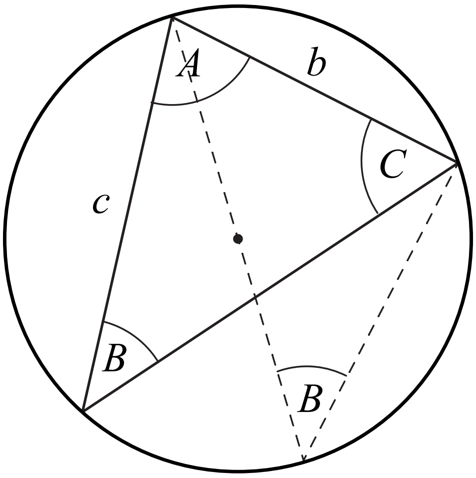
Exercise

Explore when this ambiguity arises – what are the conditions for there to be one answer, two answers, and no answers to a question giving two sides and one angle and asking for the third angle as above?

Exercise

Proof of sine rule using circle theorems …

The diagram below shows a triangle with a circle drawn around it [you can always draw a circle around a given triangle – the centre of the circle will be where the perpendicular bisectors of the sides intersect – can you explain why there is always a circle and why its centre is where we suggest?]. The radius of the circle is R.



Prove using the construction shown as dotted lines [and the appropriate circle theorems] that

$$2R\sin B = b$$

And hence

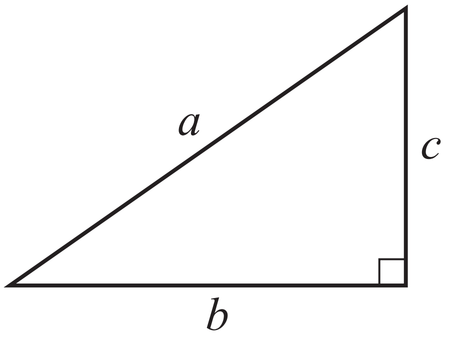
$$2R = \frac{b}{\sin B}$$

Use a similar approach to show that

$$2R = \frac{b}{\sin B} = \frac{c}{\sin C} = \frac{a}{\sin A}$$

Now we will turn to look at the cosine rule. The cosine rule is really a more general from of Pythagoras' theorem for non-right-angled triangles [actually, it works for right-angled triangles too!]

Here is a right-angled triangle



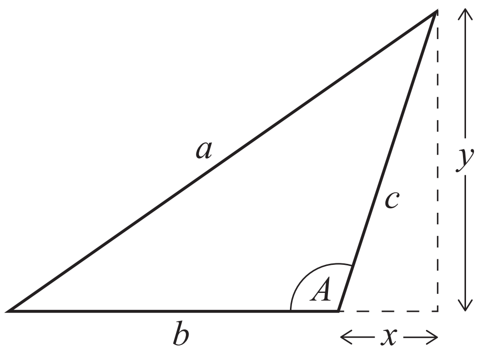
And we know that Pythagoras' theorem tells us that $a^2 = b^2 + c^2$

If we distort the triangle just a little bit and keep b and c the same length, then a^2 won't be the same as $b^2 + c^2$ anymore. We will need a correction term:

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

And we expect the correction term to change as the angles in the triangle change.

We can work out an exact relationship between the three sides and the angle A changes:



From the diagram [and Pythagoras] we can write:

$$(b + x)^2 + y^2 = a^2$$

And

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

Combining these gives:

$$a^2 = b^2 + c^2 - correction = b^2 + c^2 + 2bx$$

But $x = -c \cos A$ [make sure you can see why the minus sign is there]

And so :

$$a^2 = b^2 + c^2 - 2bc \cos A$$

Which is Pythagoras' theorem with a correction term.

Let's explore this equation, which is known as "the cosine rule" a little:

First, we note that the equation applies to any set of three sides and an appropriate angle in the triangle:

$$b^2 = c^2 + a^2 - 2ac \cos B$$

$$c^2 = a^2 + b^2 - 2ab \cos C$$

Make sure you note how a, b and c and A, B and C appear in these equations.[47] And also, how the cosine function takes care of whether the correction term needs to be positive or negative [think about how this works and make sure you understand – think about how the cosine function can be thought of as a "projection" – see the trigonometry section below]

Sometimes the cosine formula is written with the cosine as the subject [we recommend you learn just the first formula above and then manipulate it for each scenario as you need]:

$$\cos A = \frac{b^2 + c^2 - a^2}{2bc}$$

And we can also see that the cosine rule becomes Pythagoras' theorem if we set the angle to 90 [but recall we used Pythagoras' theorem to prove the cosine rule!]

[47] Look up cyclic permutations.

When do we use the cosine formula?

[note, it is almost always better to use Pythagoras’ and trigonometry if you have a right-angled triangle] :

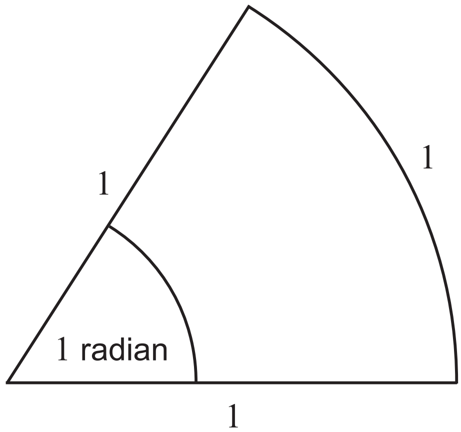
Given	**Find with one use of the cosine rule**
Three sides	Any angle
Two sides and one angle [between given sides]	The third side

MM4.2

Radian measure, including use for arc length and area of sector and segment.

Usually, the first method for measuring angles that you will encounter is to use degrees. And, as you know, one revolution is 360 degrees. There is nothing special about the number 360 [some say it is used as it is roughly the number of days in a year] but any other number would also work. We could have 400 units in a complete revolution [so 100 units in a right angle]. In fact, there is a measure of angles that uses 100 to be a right angle – it is called “Gradians” [you will see a “grad” setting on your calculator]. All these angle measures are a bit arbitrary[48] but there is one measure for angles that is more natural than all the others, and that is “radians”.[49]

So how big is one radian? We take a sector of a circle of radius 1 and arc length also 1 and we define 1 radian to be the angle that is subtended by this arc.



Or, if you prefer, equivalently we can say that one revolution is equal to 2π radians [because the length of the circumference of a circle of radius 1 is 2π]. So, 1 radian is $\frac{360}{2\pi} = 57.298°$ [often we write “rad” for radians but there is a symbol for radians, like

[48] Our decimal system is also a bit arbitrary and based on our having ten fingers. It is a pity that we did not evolve to have 12 fingers [six on each hand] as base 12 would be a nice system to work in. Also, two more arms would be useful, but evolution did not anticipate mobile phones and shopping.

[49] By “more natural” we a mean that it is the most likely measure any alien civilization would use. If you know any Martians, you can check our assertion. There is more to it than that. All the differentials and integrals involving trigonometry you will [or already have] learned in calculus are only true if the angle is in radians. So $\frac{\mathrm{d}}{\mathrm{d}y}\sin x = \cos x$ for x in radians. We do not require you to know about calculus with trigonometry in the TMUA/ESAT but if we did, we would ask questions to test your understanding of this. It is worth your exploring why we need to use radians to make the rules work. And, you should think about what differentiating and integrating would look like if x were in degrees instead of radians. You should assume angles are given in radians unless told otherwise when doing calculus.

there is a symbol for degrees, but it tends not to be used that much. The symbol is a superscript "c" : 1 rad = 1^c]

We can convert from degrees to radians and vice-versa very easily by recalling one revolution is either 360 degrees or 2π rad.

$\boldsymbol{\theta}$ degrees to radians

θ degrees is $\frac{\theta}{360}$ fraction of one full revolution and one revolution is 2π radians, so the conversion must be

$$\theta \text{ degrees} = \frac{\theta}{360} \times 2\pi \text{ radians}$$

$\boldsymbol{\alpha}$ radians to degrees

By the same argument as above, we can convert α radians to degrees

$$\alpha \text{ radians} = \frac{\alpha}{2\pi} \times 360 \text{ degrees}$$

You ought to know some standard conversions [learn them!]

Degrees	**Radians**
30	$\frac{\pi}{6}$
45	$\frac{\pi}{4}$
60	$\frac{\pi}{3}$
90	$\frac{\pi}{2}$
180	π
360	2π

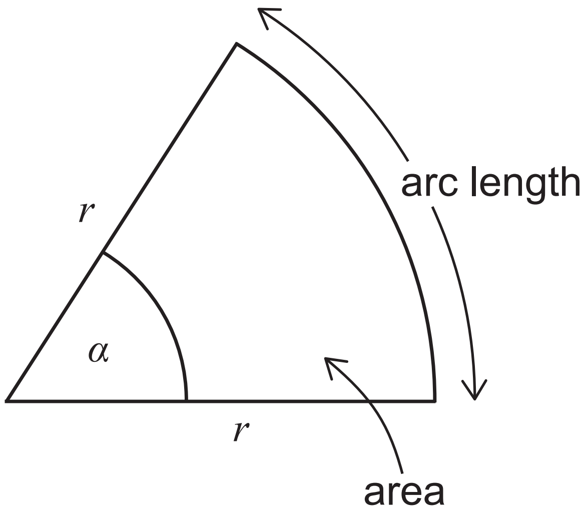
Areas and arc lengths

You need to know [and understand] the formulae for arc length and area for a sector with an angle α radians:

$$arc\ length = r\alpha$$

$$area\ of\ sector = \frac{1}{2}r^2\alpha$$

Make sure you can prove both these formulae [and understand how the proofs work] and that you know how to use both formulae. We will set out the proofs below but think about it before taking a peek.



To prove the formulae in each case we start by recalling that α radians is $\frac{\alpha}{2\pi}$ fraction of a whole circle.

So, the arc length in the picture must be

$$circumference\ \times fraction\ of\ circle\ corresponding\ to\ arc$$

which gives

$$arc\ length = 2\pi r\ \times \frac{\alpha}{2\pi} = r\alpha$$

And the area of the sector must be :

$$area\ of\ whole\ circle\ \times fraction\ of\ circle\ that\ makes\ up\ the\ sector$$

which gives

$$area\ of\ sector = \pi r^2\ \times \frac{\alpha}{2\pi} = \frac{1}{2}r^2\alpha$$

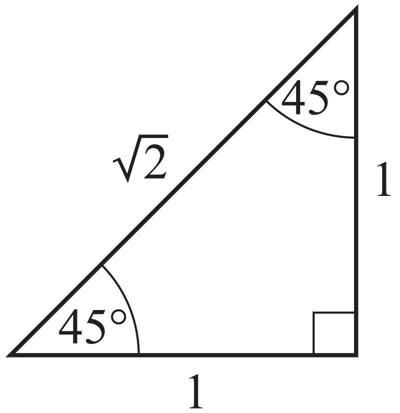
MM4.3

The values of sine, cosine, and tangent for the angles $0°,\ 30°,\ 45°,\ 60°,\ 90°$.

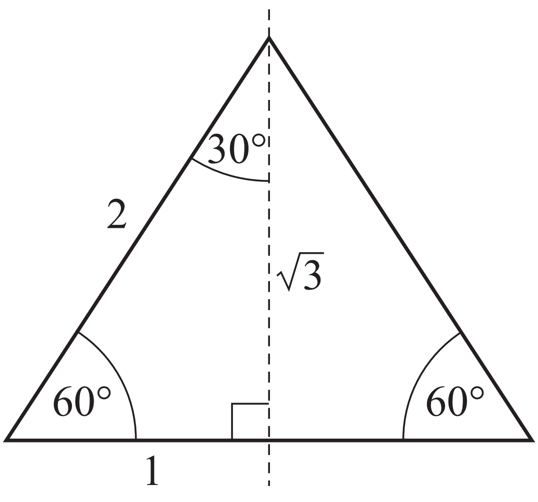
Learn these – standard triangles – and extend them to other ranges [note: tan 90 is not defined]

You should make sure you know the values listed, or that you can work them out quickly by drawing appropriate triangles. You should make sure you can identify them on a sketch of sin, cos or tan.

For 45 you can use the isosceles right-angled triangle:



And for the 30 and 60 ones, you can use half of an equilateral triangle with sides of length 2:



MM4.4

The sine, cosine, and tangent functions; their graphs, symmetries, and periodicity.

You should make sure you know the standard graphs of sine, cosine, and tangent very well – and, as mentioned, all their symmetries and periodicities. Make sure you can sketch them for degrees and for radians. We would recommend you use a good graph sketching package [e.g., DESMOS GRAPHING] and explore each of the graphs and also when they cross each other, what happens if you modify the graphs [see more later under graph sketching].

Exercise

[use a graph sketching tool to check your answers]

Sketch each of the following using a graph sketching package; sketch each in degrees and then radians using the ranges $-720° < x \leq 720°$ and $-2\pi < x \leq 2\pi$ [this is not a typo: these are not the same ranges!]

$$y = \sin x \qquad y = 2\sin x \qquad y = \sin 2x$$

$$y = \cos x \qquad y = 2\cos x \qquad y = \cos 2x$$

$$y = \tan x \qquad y = 2\tan x \qquad y = \tan 2x$$

Sketch on the same axes $y = \sin x$ and $y = \cos x$ for $-2\pi < x \leq 2\pi$. For what values of x in the range does $\cos x = \sin x$? And for what values does $\cos x = -\sin x$

Sketch each of the following for $-2\pi < x \leq 2\pi$ and pay careful attention to what the number 2 does and what the $\frac{\pi}{6}$ does to each graph.

$$y = \sin\left(2x + \frac{\pi}{6}\right)$$

$$y = \sin\left(2x - \frac{\pi}{6}\right)$$

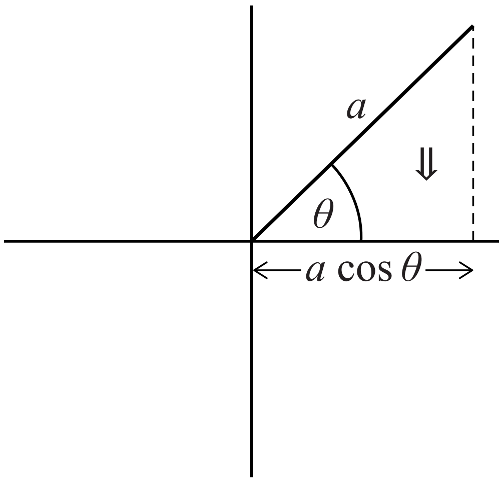
$$y = \sin 2\left(x + \frac{\pi}{6}\right)$$

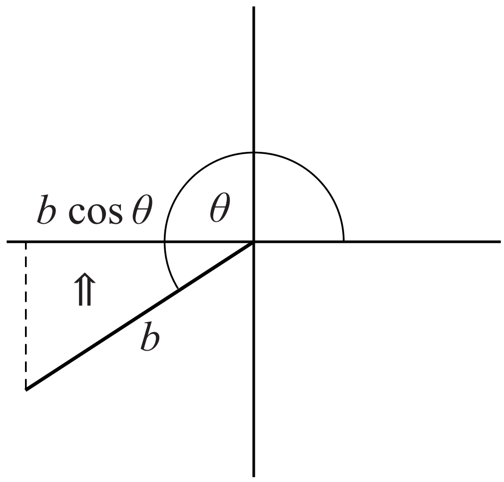
$$y = \sin 2\left(x - \frac{\pi}{6}\right)$$

A useful way to think of sine and cosine is as “projection operators”. They project lines onto the x-axis or the y-axis.

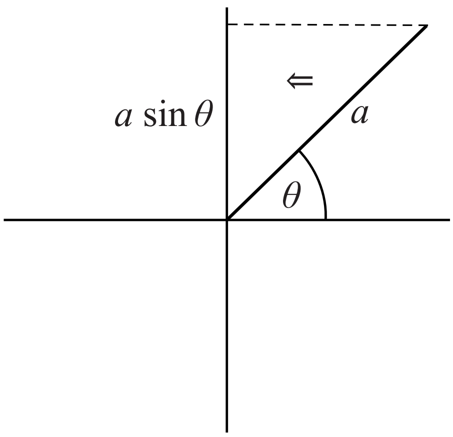
In these diagrams the line coming out at an angle always has positive length. The cosine function projects the line onto the x axis; and the sine function projects the line onto the y axis.

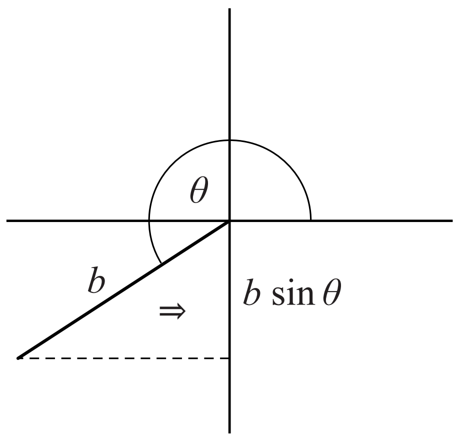
Here are some diagrams to illustrate what we mean:





Above, in the second diagram, $b\cos\theta$ is negative.

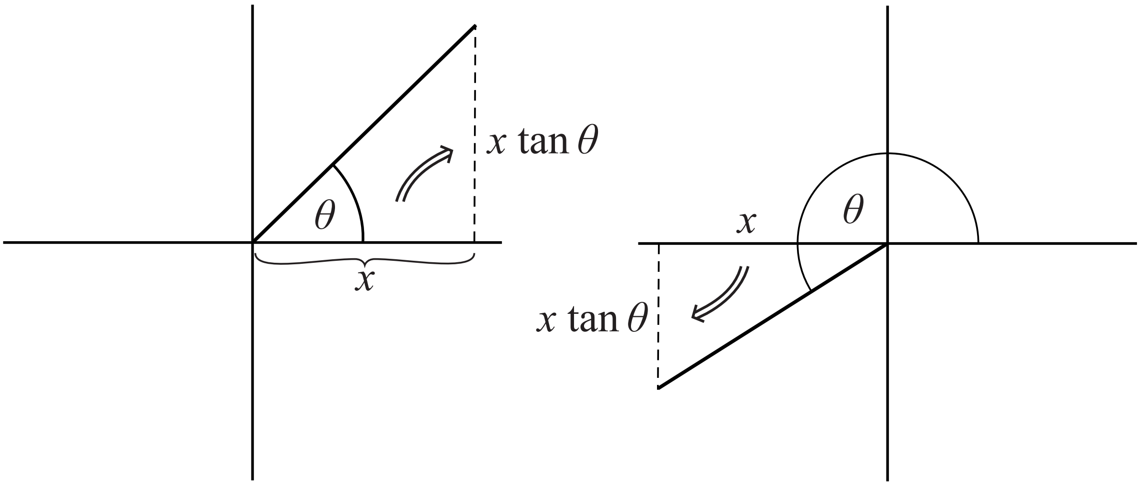




Above, in the second diagram, $b\sin\theta$ is negative.

You can see that the diagrams are essentially the “CAST” diagrams you might have drawn when solving trigonometric equations. They also help you understand why cosine and sine vary in sign for different angles.

Tangent also has an interpretation – it converts from x-axis projections to y-axis projections. Here are a couple of diagrams to illustrate what we mean:



Exercise

What is the sign of $x \tan\theta$ in each of the diagrams above? Can you explain your answer ? [Hint: think carefully about the sign of the x value in each diagram and the sign of the y value in each diagram]. How does your answer fit with the sign of $\tan\theta$ when $90 < \theta < 180$? And what about $270 < \theta < 360$?

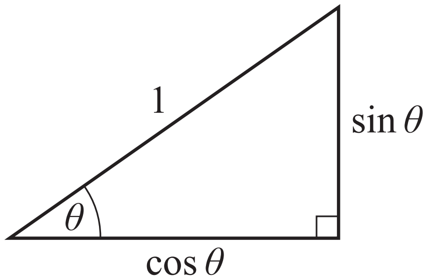
MM4.5

Knowledge and use of

$$\tan\theta = \frac{\sin\theta}{\cos\theta}$$

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

There are many ways to define trigonometric functions. Usually, we first meet trigonometry in relation to right-angled triangles, and if we use this approach then the formula $\sin^2\theta + \cos^2\theta = 1$ is clearly just a version of Pythagoras' Theorem. We have drawn a diagram to illustrate this.



You should notice that whilst we have drawn a triangle with an acute angle θ, the formula applies to ANY angle. Can you justify why it must apply to ANY angle using the "CAST" style diagrams we drew earlier?

In addition, you should know that $\tan\theta = \frac{\sin\theta}{\cos\theta}$ and it is sufficient for the TMUA/ESAT to realise this comes from the standard definition of trigonometric functions you tend to be told when you first learn them:

$\sin\theta = \frac{O}{H}$ $\quad\cos\theta = \frac{A}{H}$ $\quad\tan\theta = \frac{O}{A}$

So, it is "obvious" that

$$\tan\theta = \frac{O}{A} = \frac{O/H}{A/H} = \frac{\sin\theta}{\cos\theta}$$

You should make sure that you understand how the signs of the three trigonometric functions change as the signs of A and O change in CAST diagrams [and note that CAST diagrams always have the H value as positive, usually we set $H = 1$]

MM4.6

Solution of simple trigonometric equations in a given interval (this may involve the use of the identities in **MM4.5**); for example: $\tan x = -\frac{1}{\sqrt{3}}$ for $-\pi < x < \pi$; $\sin^2\left(2x + \frac{\pi}{3}\right) = \frac{1}{2}$ for $-2\pi < x < 2\pi$; $12\cos^2 x + 6\sin x - 10 = 2$ for $0° < x < 360°$.

You should be comfortable solving equations involving trigonometry and using either graphical or CAST type methods to list the full set of solutions. We will look at a couple of examples and make some useful comments as we work through them. There are lots of approaches to solving equations with trigonometry – you will have learnt some – and you might find you prefer methods that are different from the ones we use and that is fine. What is important is that you make sure you are comfortable with a range of methods and that you do not lose or inadvertently add extra solutions in whatever approach you decide to take.

Example

Solve $\sin^2(2x + 60) = \frac{1}{4}$ for $-360 < x < 360$

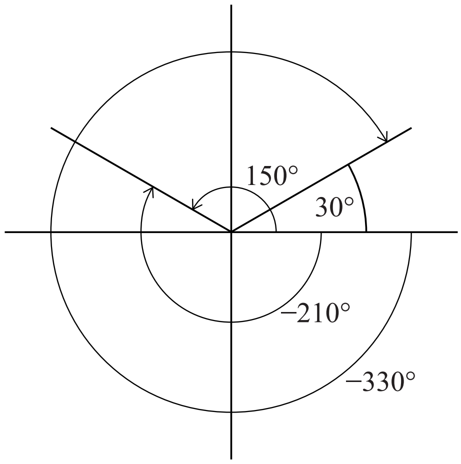
First, we take the square root of both sides and recall that, in this case, we need to consider both positive and negative square roots.[50] We get two equations to solve

$\sin(2x + 60) = \frac{1}{2}$ and $\sin(2x + 60) = -\frac{1}{2}$

Let's look at the first one. We begin by getting the basic solution [essentially, the one that your calculator would give you if you put in $\sin^{-1}\frac{1}{2}$]. We know from earlier that the basic solution here is 30 degrees [or $\frac{\pi}{6}$ radians].

We can illustrate this solution on a graph or on a CAST diagram. WE have also added the next solution of 150 degrees [You should be comfortable using both, even though you might prefer one method over the other]

[50] Don't get confused with what we mentioned about the square root symbol always meaning the positive square root.



Now we need to be careful as it is very easy to make an error and lose some of the solutions. We will first look at what we should do to get the full set of solutions, and then we will look at a common error to see how easy it is to lose solutions by doing things in the wrong order:

We have a basic solution of

$$2x + 60 = 30$$

But as we need all solutions between $-360 < x < 360$ we are going to list some more solutions [BEFORE we do any rearranging] ; and because we will be diving by 2 and subtracting 60 to get x we need to go beyond the range -360 to 360:

We will list all the relevant solutions [and at the ends of our list we might just list ones that will not be relevant just to be sure – in red]

$$2x + 60 = -690\,, -570\,, -330, -210\ \ , 30, 150, 390, 510, 750, 870$$

Rearranging to get x

$$x = -375\,, -315\,, -195, -135\,, -15, 45, 165, 225, 345, 405$$

And those in the range $-360 < x < 360$:

$$x = -315\,, -195, -135\,, -15, 45, 165, 225, 345$$

As we mentioned we will also look at a common error that can mean you lose solutions. If you rearrange the basic solution first and then look for general solutions, you will lose some of the solutions. Let's look at this INCORRECT way of solving:

We start with

$$2x + 60 = 30$$

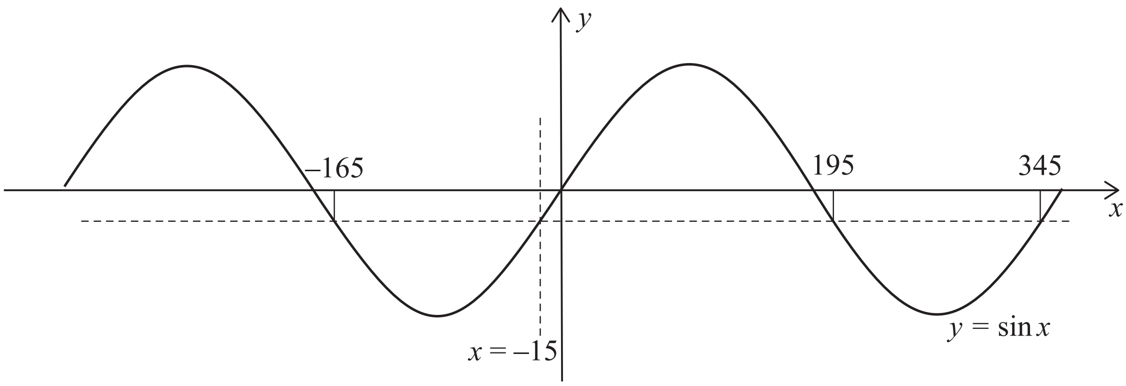
And rearrange to get

$$x = -15$$

And we then use this to generate all the other solutions for x in the range $-360 < x < 360$. When we do this, we get

$$x = -165, -15, 195, 345$$

We can see how we got these using a simple graph:



And you can see things have gone wrong. So better to find all your solutions first and then rearrange to find x at the end.

We have so far only solved half the question, as we also need to deal with $\sin(2x + 60) = -\frac{1}{2}$ but we will leave that as an "exercise for the reader".

Exercise

Repeat the above solutions for the same question but in radians:

$\sin^2\left(2x+\frac{\pi}{3}\right)=\frac{1}{4}$ for $-2\pi < x < 2\pi$

Often trigonometry is mixed with other topics such as quadratics or cubics or inequalities etc. In those cases you often need to aim to get to expressions of the form $\sin ? = number$ or $\cos ? = number$ or $\tan ? = number$

Here is an example:

Example

Solve $12\cos^2 x + 6\sin x - 10 = 2$ for $0° < x < 360°$.

This looks like a quadratic but has both sin and cos in it. The first thing we should[51] do is convert the cos to sin using $\cos^2 x = 1-\sin^2 x$:

$12(1-\sin^2 x) + 6\sin x - 10 = 2$

Rearrange [and put $\sin x = S$ to simplify the "look"]

$$12S^2 - 6S = 0$$

Factorise and solve for S:

$$6S(2S-1) = 0$$

So, we get $S = \sin x = 0$ or $S = \sin x = \frac{1}{2}$ and solutions $x = 30,\ 150,\ 180$

[51] Often mathematicians will think of an idea and check it roughly in their heads first before proceeding to check the idea is likely to work. Sometimes, an idea that seems like it might work well will flounder later and then you will need to start again and rethink. STEP mathematics questions are a useful resource to help you learn to think through different approaches to questions. See https://www.ocr.org.uk/students/step-mathematics/preparing-for-step/

MM5. Exponentials and Logarithms

MM5.1

$y = a^x$ and its graph, for simple positive values of a.

Make sure you know what the graph of $y = a^x$ looks like for different values of a. Look carefully at $0 < a < 1$ and $a = 1$ and $1 < a$ and make sure you can explain their features. What happens to the graphs as a gets bigger and bigger? Use a graph drawing package [e.g., DESMOS GRAPHING] to help you if necessary but make sure you think through the results in each and every case.

Notice that we do not look at cases when $a < 0$. We discussed why we do not look at this case in the discussions earlier when we looked at indices so it might be useful to revisit that discussion.

MM5.2

Laws of logarithms:

$$a^b = c \Leftrightarrow b = \log_a c$$

$$\log_a x + \log_a y = \log_a(xy)$$

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

$$k \log_a x = \log_a(x^k)$$

including the special cases:

$$\log_a \frac{1}{x} = -\log_a x$$

$$\log_a a = 1$$

Questions requiring knowledge of the change of base formula will not be set.

Logarithms are very closely related to indices; in fact, they are really the "inverse of indices". They tell you what power a number has to be raised to rather than raising a number to a power. Let's unpack that idea a little bit to get an idea of how logs work. We will start with a few examples to help you get the feel for things and then we will look at logarithms graphically; and then we will move on to exploring [briefly!] how the logarithm rules work.

But before we begin, a teeny little bit of history is useful. Before calculators existed, doing lots of calculations, especially with big numbers, could be complicated. So, logarithms were invented to make the calculations easier [it is worth looking at the history a little bit to understand how clever and inventive mathematicians can be – see the Wikipedia page on the History of Logarithms. Even if you are not keen on history, the background to logarithms is worth exploring.]

Here are some examples using $\log_{10}$. $\log_{10}$ tells you what power 10 needs to be raised to get a given number:

$\log_{10} 10 = 1$ because 10 needs to be raised to the power of 1 to get 10: $10^1 = 10$

$\log_{10} 100 = 2$ because 10 needs to be raised to the power of 2 to get 100: $10^2 = 100$

$\log_{10} 1000 = 3$ because 10 needs to be raised to the power of 3 to get 1000: $10^3 = 1000$

$\log_{10} 27 = 1.431363764 \ldots \ldots$ because $10^{1.431363764\ldots} = 27$ [you can check this with your calculator]

We can also try using logs to other "bases". $\log_2$ tells you what power 2 needs to be raised to get a given number: here are some other [reasonably obvious] examples:

$\log_2 32 = 5$ because 2 needs to be raised to the power of 5 to get 32 : $2^5 = 32$

$\log_2 \frac{1}{2} = -1$ because 2 needs to be raised to the power of -1 to get $\frac{1}{2}$: $2^{-1} = \frac{1}{2}$

So, we have in general the following relationship between log and powers:

$$\log_a c = b \quad \textbf{says the same as} \quad a^b = c$$

You should make sure you are very familiar and comfortable with this idea.

A few things to note about this [things might change when you meet more mathematics but then definitions get honed and changed too!]:

- We only take logs with a positive base number: so $a > 0$ [but $a \neq 1$]
- We can only take the logs of positive numbers: so $c > 0$
- The log of a number can be negative: so b can be any number [even 0]

You should be able to work out why all three of these statements apply – look back to our discussion on indices if you are not sure. It is important to remember that the log function is not defined for negative numbers – that is we must have $c > 0$. [This was important in a TMUA/ESAT question from a few years ago[52]]

[52] Look at TMUA paper 1 2021 question 20

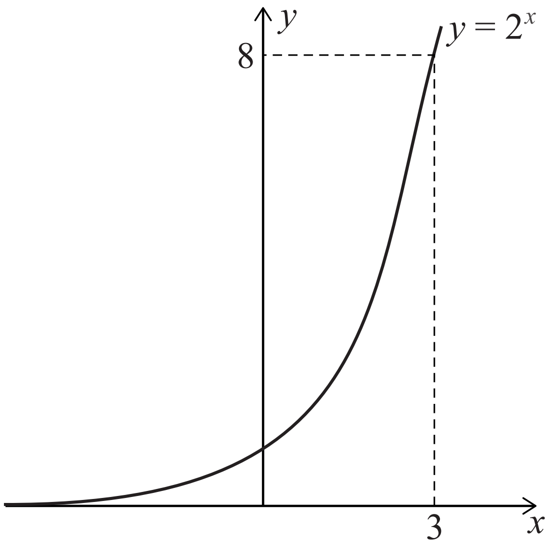
Exercise

work out each of the following:

$\log_5 5^2$; $\log_3 3^5$; $\log_7 \sqrt{7}$

We can take a brief look at logs and graphs. We will work with base 2 as that gives nice graphs[53]

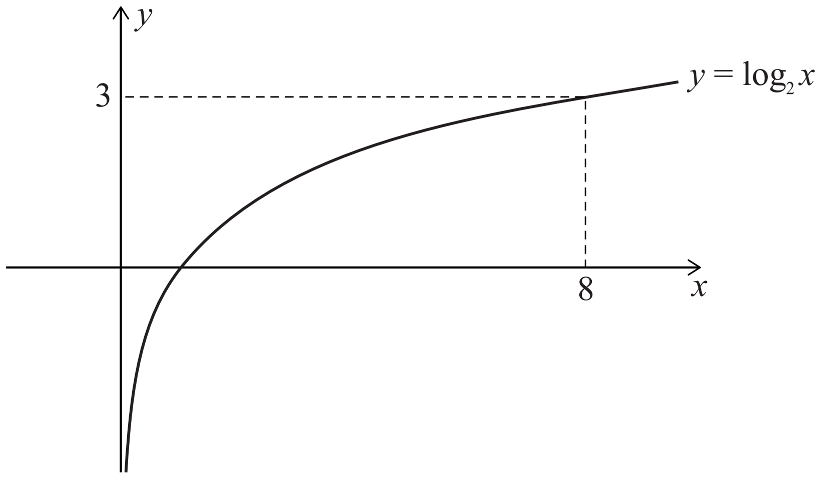
First, we draw $y = 2^x$ and look at a few values



From this, you can see that 2^x takes numbers from the x-axis and gives us numbers on the y-axis : it maps $x \rightarrow 2^x$. And also, if you start on the y-axis [say with 8] and trace back to what number corresponds to it on the x-axis you get the log of the number of the y-axis, namely 3. So going backwards from the y-axis to the x-axis we get $y \rightarrow \log_2 y$

We can now draw the log graph as it is just the graph of $y = 2^x$ with the x and y axes swapped:

[53] Use a graph package to draw $y = 2^x$ and $y = 10^x$ to see how fast they grow [that is exponential growth]



A few things to note here too:

- You can see the graph is only defined for $x > 0$ as we expected.
- The graph of $y = \log_2 x$ crosses the x-axis at 1. Can you explain why?[54]

Now we can look at all the logarithm rules we expect you to know – you should make sure you understand them [i.e. you know how they work and where they come from] and you should make sure you can use them correctly.

We start with $\log_a x + \log_a y = \log_a(xy)$

This is really the logarithm equivalent of $a^p a^q = a^{p+q}$ [think about how this relates to the log equation]

We can see how this equation works as follows:

$$a^{\log_a x + \log_a y} = a^{\log_a x}\, a^{\log_a y} = xy = a^{\log_a(xy)}$$

And make sure you can see what rule we have used at each stage of this. Note we have used one idea that we haven't drawn attention to as yet : $a^{\log_a x} = x$. We hope this idea is "obvious" as it is essentially the very definition of a logarithm.

[54] Because $2^0 = 1$ and so $\log_2 1 = 0$

We can now look at the other rules in the same way :

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

$$a^{\log_a x - \log_a y} \;=\; a^{\log_a x}\, a^{-\log_a y} \;=\; \frac{a^{\log_a x}}{a^{\log_a y}} \;=\frac{x}{y} = a^{\log_a\left(\frac{x}{y}\right)}$$

$$k\log_a x = \log_a(x^k)$$

$$a^{k\log_a x} = (a^{\log_a x})^k = x^k = a^{\log_a(x^k)}$$

$$\log_a \frac{1}{x} = -\log_a x$$

$$a^{-\log_a x} \;=\; \frac{1}{a^{\log_a x}} = \;\frac{1}{x}\; = a^{\log_a \frac{1}{x}}$$

And finally

$$\log_a a = 1$$

which should be “obvious” because $a^1 = a$

The specification mentions the change of base formula and says it will not be examined in TMUA/ESAT. Nevertheless, it is a useful formula and we recommend you have it in your “remembered formulae” maths kit, and make sure you can derive it and understand it too! We will take a brief look at the formula here [but you can skip this section as it is not part of the TMUA/ESAT and we won’t ask questions that depend on it].

The change of base formula allows you to convert from a log with one base to a log with another base; for instance, changing from base 4 to base 7: $\log_4$ to $\log_7$.

Before we explore this idea, have a think about how you might go about this task; for instance, how might you find $\log_4 23$ in terms of $\log_7 23$?

Let’s start with the example we just gave:

Example

find $\log_4 23$ in terms of $\log_7 23$.

Let $\log_4 23 = p$

then $4^p = 23$

take $\log_7$ of both sides: $\log_7 4^p = \log_7 23$

which gives $p \log_7 4 = \log_7 23$

so $p = \log_4 23 = \frac{\log_7 23}{\log_7 4}$

We can use the same method to derive the change of base formula:

Example

Change from $\log_a b$ to $\log_c b$

Let $p = \log_a b$

then $a^p = b$

take $\log_c$ of both sides: $\log_c a^p = \log_c b$

which gives $p \log_c a = \log_c b$

so $p = \log_a b = \frac{\log_c b}{\log_c a}$

and this gives us the change of base formula:

$$\log_a b = \frac{\log_c b}{\log_c a}$$

And as a final note, this leads to one more useful formula if we set $c = b$:

$$\log_a b = \frac{\log_b b}{\log_b a}$$

and as $\log_b b = 1$ we have :

$$\log_a b = \frac{1}{\log_b a}$$

MM5.3

The solution of equations of the form $a^x = b$, and equations which can be reduced to this form, including those that need prior algebraic manipulation; for example, $3^{2x} = 4$ and $25^x - 3 \times 5^x + 2 = 0$.

Example

Solve: $5^{2x} = 27$

We will solve this exactly[55]

We can take logs of both sides, but we need to decide which base is best – we could use base 5 here or [because $27 = 3^3$] we could use base 3. We will try both approaches just for completeness[56]:

Approach 1:

Take $\log_5$ of both sides:

$\log_5 5^{2x} = \log_5 27 = \log_5 3^3$

Simplifying:

$2x = 3\log_5 3$

and so

$x = \frac{3}{2}\log_5 3$

Approach 2:

Take $\log_3$ of both side:

$\log_3 5^{2x} = \log_3 27 = \log_3 3^3$

Simplifying:

$2x\log_3 5 = 3$

And so $x = \frac{3}{2\log_3 5}$

[55] Exactly means we will find an expression for the value of x rather than calculating [using a calculator] and then rounding the answer. You should be aware that many log values are irrational [like surds] and so cannot be expressed precisely as a decimal. That is why we often use surds and log expressions rather than rounded numerical values.

[56] You can check these two approaches give the same value using the change of base formula.

MM6. Differentiation

MM6.1

The derivative of $f(x)$ as the gradient of the tangent to the graph $y = f(x)$ at a point.

In addition:

Interpretation of a derivative as a rate of change;

Second-order derivatives;

Knowledge of notation: $\frac{dy}{dx}$, $\frac{d^2y}{dx^2}$, $f'(x)$, and $f''(x)$.

Differentiation from first principles is excluded.

In order to understand what derivative is, you need to have a good grasp of the notion of a "rate of change". We will explore this idea briefly. You will be used to ideas such as speed, and speed is a rate of change. Speed tells you how fast distance is changing compared to time: a speed of 3m/s means that distance is changing at a rate of 3m for every one second of time that elapses. So, rates of change tell you how fast one measure changes compared to another measure. Usually, we express rate of change as "how many units of one thing change per single unit of another". Speed is "how much distance changes for every single unit change of time". Acceleration is another example of a rate of change: it tells you how much the speed changes [m/s] for every one unit of time [s]: acceleration is usually measured in metres-per-second changed for every second of time and this is usually written [slightly confusingly] as *metres per second per second* or m/s/s and often the "per second per second" is changed to s^{-2}

Gradients are also rates of change. Recall we said earlier that "we can think of the gradient as telling us how much we have to go vertically to get back on the line for every 1 unit we move horizontally from a point of the line". In other words, gradient is the rate of change of y compared to [or, with respect to] x. The gradient tells us how much y changes for every one unit change in x. So gradient is just a rate of change.

You should be able to understand why the gradient of a distance-time graph will give you speed, and the gradient of a speed-time graph will give you acceleration.[57]

[57] We have ignored the vector vs scalar issues here.

Now you have an idea that gradient is just a rate of change, we need to look at what a rate of change for a curve might mean. Actually, we will look at what we mean by the rate of change of a curve *at a point on the curve*; and recall we know what we mean by the rate of change for a straight line – it is its gradient.

We define the rate of change at a point on a curve as being the gradient of the tangent to the curve at that point. Intuitively this definition should make sense – it is worth your spending some time convincing yourself that this is the best definition. In fact, the definition gives rise to the interpretation of differentiation that we shall look at below – and if you have dealt with differentiation from first principles [which is not on the TMUA/ESAT specification] you will have seen how this leads to the interpretation of differentiation that we set out next.

Now we turn to look at how to find the differential of an expression and how we should understand what the differential of an expression means.

Let's start by thinking about what the differential mean. You already know that when you are given an expression such as $y = x^3 + 7x^2 - 3x + 11$ you can find what y value corresponds to a given x value by substituting the given x value into the expression. What then does $\frac{dy}{dx}$ tell you? In this case we have $\frac{dy}{dx} = 3x^2 + 14x - 3$ and this tells you the gradient of the curve [its rate of change at a point] for any given x value. And recall this is gradient of the tangent to the curve at that point.

So, the $y = \cdots$ gives you y values once you have an x value; and the $\frac{dy}{dx} = \cdots$ gives you gradients [or rates of change] of the curve [the tangent to the curve] at the point corresponding to the given x value.

Below we will see why calculating the gradient of a curve at a point is so useful.

As well as knowing what differentiation tells us, you should make sure you are familiar with the various notation used for differentiation. We expect you to know $\frac{dy}{dx}$, $\frac{d^2y}{dx^2}$, $f'(x)$, and $f''(x)$.[58]

Although we do not use it in the TMUA/ESAT, you should also be aware of the "dot" notation for differentiation. The dot notion tends to be used in physics [and mechanics] when differentiating with respect to time [i.e. time is on the x-axis]: for instance $\frac{ds}{dt}$ might be written as $\dot{s}$, and $\frac{d^2s}{dt^2}$ as $\ddot{s}$

You will note that the specification in this section says "second-order derivatives"; we will explain what we expect you to know about these below.

[58] Note where the two 2s go in the notation: $\frac{d^2y}{dx^2}$

MM6.2

Differentiation of x^n for rational n, and related sums and differences. This might require some simplification before differentiating.

For example, the ability to differentiate an expression such as $\frac{(3x+2)^2}{x^{\frac{1}{2}}}$

In the TMUA/ESAT, we expect you to be able to differentiate simple expressions involving sums of powers of x or expressions that can be simplified to sums of powers of x. We do NOT expect you to be able to differentiate trigonometric expressions or use rules like the chain rule, the product rule etc. We have kept to the scope of what we expect you to be able to differentiate both narrow and simple because we want to be able to test your *understanding* of the topics rather than how good you are using standard algorithms.

We expect you to know the rule for differentiating a power of x :

$$\frac{\mathrm{d}}{\mathrm{d}x}\, x^n = nx^{n-1}$$

And notice that we use $\frac{\mathrm{d}}{\mathrm{d}x}$ to mean "differentiate this".

And [if you have looked at differentiation from first principles, which is not on the TMUA/ESAT specification!] you should think about why the result when you differentiate x^n is actually $\binom{n}{1}x^{n-1}$ when n is an integer.

In general, if you are not using some of the more advanced rules such as the chain rule or product rule, the best thing to do when differentiating a given expression is to simplify first to make it into a sum of powers of x and then differentiate term by term. In fact, one thing we have not yet mentioned [because it is usually taken as obvious in a first course in calculus[59]] is that you can differentiate term by term and add up the result to get the differential of an expression. For instance, you can do this:

$$\frac{\mathrm{d}}{\mathrm{d}x}(x^3+7x^2-3x+11) = \frac{\mathrm{d}}{\mathrm{d}x}x^3 + \frac{\mathrm{d}}{\mathrm{d}x}\,7x^2 - \frac{\mathrm{d}}{\mathrm{d}x}\,3x + \frac{\mathrm{d}}{\mathrm{d}x}\,11 = 3x^2+14x-3 + 0$$

[59] It is usually somewhat dangerous to assume something is obvious in mathematics until you have spent time convincing yourself that it is. Many mistakes in more advanced mathematics can arise when a technique or idea is used incorrectly because it seemed obvious to use it that way: for instance, it might seem "obvious" to write something like $(x+y)^2 = x^2+y^2$ but it is, in fact, generally mathematical bunkum. Mathematicians sometimes use words such as "trivial" and "obvious" but what they often mean is that they have thought deeply about it, often for some time [even weeks or months or years], and only after all this deep thought is the idea they are referring to "obvious" etc. So do not worry if something that someone says is "obvious" is not so obvious to you! See also footnote 76.

MM6.3

Applications of differentiation to gradients, tangents, normals, stationary points (maxima and minima only), strictly increasing functions [if $f'(x) > 0$] and strictly decreasing functions [if $f'(x) < 0$]. Points of inflexion will not be examined, although a qualitative understanding of points of inflexion in the curves of simple polynomial functions is expected.

On of the motivations behind this section of the specification is to equip you with enough basic calculus techniques to help you sketch curves given an equation. As you read through this section and as you think about the ideas we meet, make sure you think about how the ideas relate to the shapes of curves. You ought to have a good idea of the general shape of quadratic curves, cubics, quartics, and quintics and some idea how to generalise those shapes to polynomial with higher powers of x [see later – the graph sketching section of these notes].

We expect you to be able to use differentiation to find the gradients of tangents at a given point and find equations of both the tangents and the normal to a curve at a given point. All of these should be topics you have covered extensively during your studies. In these notes, we will assume you know how to deal with these.

You should also be able to identify stationary points [sometime these are called local maxima or local minima as they are not necessarily the very least or the very greatest value of the function in question] and classify them [i.e., tell if they are maxima or minima using calculus or some other methods].

We also expect you to have a general understanding of what a [horizontal] point of inflexion[60] is and some idea about when they might occur when dealing with simple polynomial functions [polynomial functions are just sums of integer powers of x – such as quadratics, cubics, quartics etc.].

Let's start by looking at stationary points and how we can use the second derivative to help tell if they are local maxima or local minima. Stationary points occur when the tangent to a curve is horizontal. And if the tangent is horizontal then the gradient of the tangent must be zero. And we know that the gradient of the tangent is given by $\frac{\mathrm{d}y}{\mathrm{d}x}$. So stationary points occur at the x-values that make

$$\frac{\mathrm{d}y}{\mathrm{d}x} = 0$$

But how do you know if the stationary point you have identified is a maximum, a minimum [or a point of inflexion]? There are a number of techniques you can use to tell:

[60] This is often written as "inflection".

- You can sometimes tell by knowing the general shape of the curve you are sketching. For instance, if you are sketching the cubic $y = 2x^3 - 3x^2 - 12x + 6$, you can differentiate to find the x coordinates of the stationary points [if there are any[61]]:

  $$\frac{\mathrm{d}y}{\mathrm{d}x} = 6x^2 - 6x - 12 = 0 = 6(x^2 - x - 2) = 6(x - 2)(x + 1)$$

  And so stationary points occur at $x = 2$ and $x = -1$.

  This cubic has a positive coefficient for the x^3 term, so it has its maximum to the left of its minimum. This tells us that the maximum has x coordinate of -1 and the minimum has an x coordinate of 2.

- You can look at the y values either side of the x value you have identified for the stationary point. This tends to be a less efficient method but can be useful. Generally, other methods are better than this and when using this method, you need to be careful not to take x values too far away from the stationary point's x value in case you cross over other stationary points.
  In the example above, we found stationary points at $x = 2$ and $x = -1$. If you calculate the y values at $x = 2$ and at $x = 2.1$ and 1.9 you will find the y values either side of $x = 2$ are larger than the y value at $x = 2$. This suggest the stationary point at $x = 2$ is a minimum.

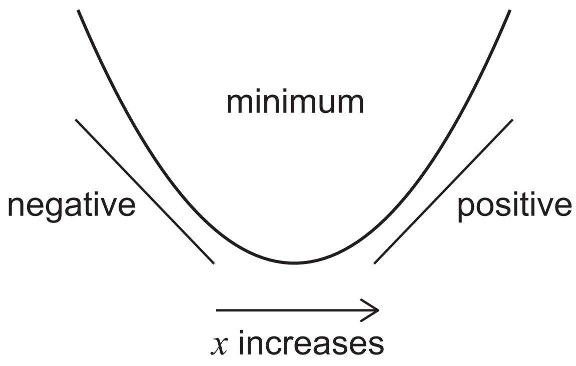
- You can look at the value of the second derivate at the stationary point. If the second derivative at the stationary point is positive, then it is a minimum; and if the second derivative is negative, then the stationary point is a maximum. Check the values for our example $y = 2x^3 - 3x^2 - 12x + 6$ at $x = 2$ and $x = -1$.

  But why is this?

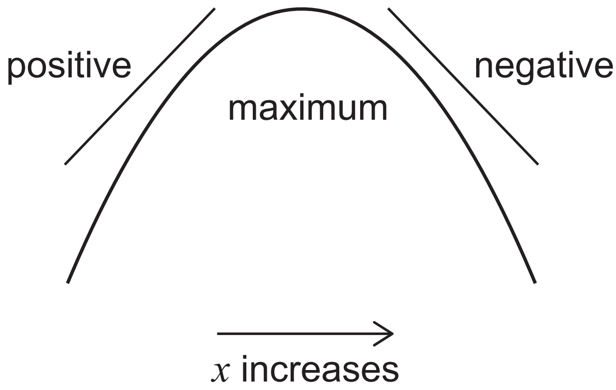
  The second derivative of a function can tell you how the first derivative is changing – usually whether the derivative is increasing or decreasing. If as x increases the first derivative goes from negative to positive [see the diagram below] we have a minimum ; and so, the derivative is usually[62] increasing at a minimum. Hence, we know that if $\frac{\mathrm{d}^2y}{\mathrm{d}x^2} > 0$ at the stationary point – and recall the second derivative tells you how the first derivative is behaving - then there is a minimum.

[61] Some cubic graphs have no stationary points, and some have only a point of inflexion. Make sure you can sketch such curves. Can you work out the equation of a cubic with no stationary points [hint – start with a quadratic with no roots and integrate]

[62]We are being cautious as the derivative at a minimum could also give $\frac{\mathrm{d}^2y}{\mathrm{d}x^2} = 0$ in some cases. See our discussion below



If as x increases the derivate goes from positive to negative [see the diagram below] we have a maximum; and so, the derivative is usually decreasing at a maximum. Hence, we know that if $\frac{\mathrm{d}^2y}{\mathrm{d}x^2} < 0$ at the stationary point, then there is a maximum.



In the last bullet point above we were a little cautious in how we worded things. This is because we need to be careful about the logic of what we have said:

- If $\frac{\mathrm{d}y}{\mathrm{d}x} = 0$ and $\frac{\mathrm{d}^2y}{\mathrm{d}x^2} > 0$ then we have a minimum

- If $\frac{\mathrm{d}y}{\mathrm{d}x} = 0$ and $\frac{\mathrm{d}^2y}{\mathrm{d}x^2} < 0$ then we have a maximum

And so

- the condition $\frac{\mathrm{d}y}{\mathrm{d}x} = 0$ and $\frac{\mathrm{d}^2y}{\mathrm{d}x^2} > 0$ is **sufficient but not necessary**[63] for a minimum.

- the condition $\frac{\mathrm{d}y}{\mathrm{d}x} = 0$ and $\frac{\mathrm{d}^2y}{\mathrm{d}x^2} < 0$ is **sufficient but not necessary**[64] for a maximum.

Why is the condition only a **sufficient** condition and not a **necessary** one ? The answer is that:

- $\frac{\mathrm{d}y}{\mathrm{d}x} = 0$ and $\frac{\mathrm{d}^2y}{\mathrm{d}x^2} = 0$ could occur at a minimum [e.g. at $y = x^4$]

- $\frac{\mathrm{d}y}{\mathrm{d}x} = 0$ and $\frac{\mathrm{d}^2y}{\mathrm{d}x^2} = 0$ could occur at a maximum [e.g. at $y = -x^4$]

In other words:

- at a minimum we have $\frac{\mathrm{d}y}{\mathrm{d}x} = 0$ and we could have either $\frac{\mathrm{d}^2y}{\mathrm{d}x^2} > 0$ or $\frac{\mathrm{d}^2y}{\mathrm{d}x^2} = 0$ but not $\frac{\mathrm{d}^2y}{\mathrm{d}x^2} < 0$

- at a maximum we have $\frac{\mathrm{d}y}{\mathrm{d}x} = 0$ and we could have either $\frac{\mathrm{d}^2y}{\mathrm{d}x^2} < 0$ or $\frac{\mathrm{d}^2y}{\mathrm{d}x^2} = 0$ but not $\frac{\mathrm{d}^2y}{\mathrm{d}x^2} > 0$

[63] See our *Notes on Logic and Proof* for a discussion of necessary and of sufficient
[64] See our *Notes on Logic and Proof* for a discussion of necessary and of sufficient

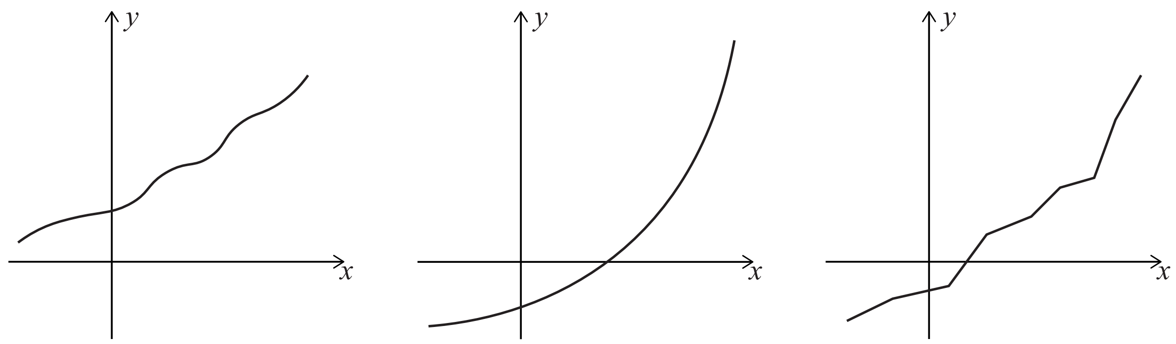
MM6.3.

Applications of differentiation to gradients, tangents, normals, stationary points (maxima and minima only), strictly increasing functions [if $f'(x) > 0$] and strictly decreasing functions [if $f'(x) < 0$]. Points of inflexion will not be examined, although a qualitative understanding of points of inflexion in the curves of simple polynomial functions is expected.

Now we can turn to look at strictly increasing and strictly decreasing functions in relation to differentiation.[65] We will not explore the other topics mentioned in MM6.3 as they are standard topics that we expect you to have fully covered in your mathematics course.

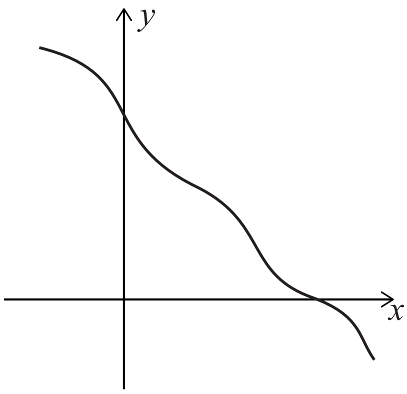
First, note that here the specification has been slightly changed when compared to previous TMUA specifications: *strictly increasing* and *strictly decreasing* has replaced *increasing* and *decreasing*.

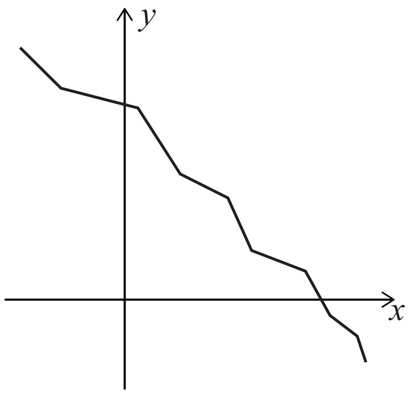
Let's start by developing an intuitive grasp for what a strictly increasing function is. In simple terms, a strictly increasing function is one that is always getting greater [the y value gets greater] as the x value gets greater. We could also say that it always slopes upwards. We have drawn some strictly increasing functions below.



The same idea applies to strictly decreasing functions: they always get less [the y value gets less] as the x value gets greater. We could also say that it always slopes downwards. We have drawn some strictly decreasing functions below:

[65] We have used a slightly narrow definition of strictly increasing and decreasing in the TMUA/ESAT as we wanted to relate the topic to differentiation without adding too many additional ideas to grasp. There are other definitions that take into account some of the subtle issues we discuss in this section – for instance, other definitions do not have to worry about functions with corners etc





There are a few things to note in what we have said so far. First notice that we used the terms “greater” and “less” rather than “bigger” and “smaller”. We did this because there can be some confusion if the y-values are negative: it might not be clear that -3 is “bigger” than -5 , but it is true that -3 is greater than -5. Secondly, notice that some of the pictures in our examples deliberately have corners in them. We put these in to show that a function can increase but there might be place on the function where we cannot find the value of $f'(x)$ – you cannot find the value for $f'(x)$ on a corner as a corner does not have a unique tangent. Both these points are important for when we come to the definition and the logic we use in the TMUA/ESAT in relation to strictly increasing and strictly decreasing functions.

The way we have set out our definition does not capture all cases of strictly increasing/decreasing functions. You should pay careful attention to what the specification says:

if $f'(x) > 0$ then the function is strictly increasing

if $f'(x) < 0$ then the function is strictly decreasing

So, we have given a sufficiency condition for strictly increasing/decreasing but not a necessary one [there are necessary conditions that we could give but we are interested in the TMUA/ESAT at getting a feel for the shape of functions using differentiation so we have restricted what we have expected of you]

Notice that we **<u>CANNOT</u>** say:

If a function is strictly increasing, then $f'(x) > 0$ for all x

If a function is strictly decreasing, then $f'(x) < 0$ for all x

It should be clear we cannot say this – look at the examples we drew above and work our why these two statements are not true.[66]

There is a little bit more we can say about strictly increasing and strictly decreasing functions: notice we have not yet said what x values the conditions apply to. Usually we apply the definition to the whole domain of the function [the domain is all the x values the function is defined on – usually this will be all real x values, that is the whole x-axis, unless otherwise stated, or unless the function is only defined on a subset of the real numbers, e.g. log functions]. We could look at restricted set of x values in an obvious manner:

if $f'(x) > 0$ in $a < x < b$ then the function is strictly increasing in $a < x < b$

if $f'(x) < 0$ in $a < x < b$ then the function is strictly decreasing in $a < x < b$

A final note

In TMUA/ESAT, we will not test you on special cases and subtle technicalities for this topic. We do expect you to understand what is means for a nice graph [continuous – i.e. no breaks in the graph; with no sharp corners] to be strictly increasing or strictly decreasing and we will base our questions around simple functions such as polynomials which we know are well behaved. You won't need to worry about issues such as if the functions we use have corners and that sort of thing, but you do need to make sure you understand the logic of the definition we have decided to use.

Inflexions

For inflexions, we expect you to know what they are in general terms and where they might occur [e.g., in $y = x^3$] so that you can take account of their possibility when thinking about polynomials and their graphs. We do not expect you to have a detailed knowledge of how to identify a point of inflexion using differentiation and we will not ask questions that specifically involve technical issues directly related to points of inflexion.

[66] Think about what happens to $f'(x)$ when graphs have "corners on them".

Integration

Before we start to look at the specification for this section, we have a few introductory remarks to make.

First, there are two ways to think about integration. These two ways are closely related but *how* they relate is not something we expect you to know for the TMUA/ESAT.

Either you can think of integration as the reverse of differentiation – it tells you what must have been differentiated to get the expression you are integrating.

Or, you can think of integration as finding the "area" between a curve and the x-axis; but we have to be very careful by what we mean by "area" and that is something we shall discuss carefully below.

When integration is written as the reverse of differentiation it tends to look like this:[67]

$$\int x^2 \ \mathrm{d}x$$

The expression being integrated appears between the $\int$ and the $\mathrm{d}x$ and in this case the expression is asking "what must be differentiated to get x^2 ?" and the answer is

$$\frac{x^3}{3} + c$$

And note we add a constant term "c" – you should be able to explain why we do this. [recall that when you differentiate a constant, you get 0]

When integration is asking you to find "areas" then we tend to have numbers written at the top and bottom of the $\int$. In these cases, it is called "definite integration", and the numbers are called "limits". We will unpack some aspects of definite integration below. Here is an example of a definite integral:

$$\int_1^3 x^2 \ \mathrm{d}x$$

[67] Sometimes [but not really until some university course on mathematics] integration can be written as $\int \mathrm{d}x \ \ x^2$. This $\int$ is actually an elongated "s" meaning "sum" – it behaves similarly to Σ. The notation we use for calculus is largely derived from Leibnitz – look up "Leibnitz notation".

MM7.1

Definite integration as related to the ‘area between a curve and an axis’. The difference between finding a definite integral and finding the area between a curve and an axis is expected to be understood.

In this section, we will assume you know how to integrate and deal with limits in an integral!

When we talk about areas and integration, we need to be very careful. The term “area” is usually taken to be a positive value and that can lead to some confusion when we talk about definite integration and the area between a curve and an axis. Definite integration is almost a sum of areas but instead it subtracts “areas” that are underneath the x-axis. So, a definite integral calculates all the areas that sit above the x-axis and sums them up and then subtracts all the areas that sit below the x axis; and it does this whole calculation in one go.

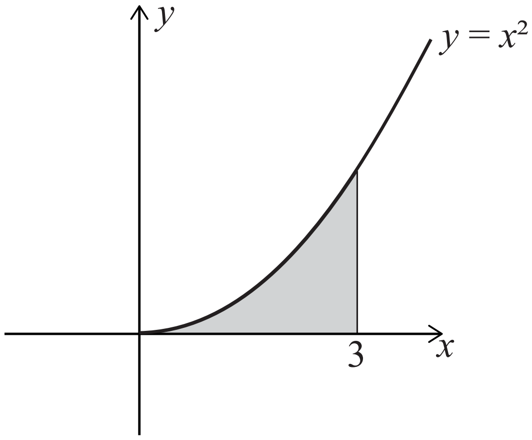
As an aside: There is a good reason for this – we do not expect you to know the details for TMUA/ESAT but you might have met some diagrams in class where integration is presented as sums of very thin rectangle between the curve and the x-axis. The contribution of each of these rectangles to the integral is the y value of the rectangle height [and the y value can be positive or negative] times the teeny width along the x-axis [the $\mathrm{d}x$ bit, which is always taken to be positive]. As rectangles that sit under the x-axis have negative y values, their contributions to the integral are negative – and hence the negative “areas” for those bits of the integral

Let's look at a few examples to help us understand how this works:

Example 1

Calculate $\int_0^3 x^2 \, \mathrm{d}x$ and draw a picture to illustrate the meaning of the answer

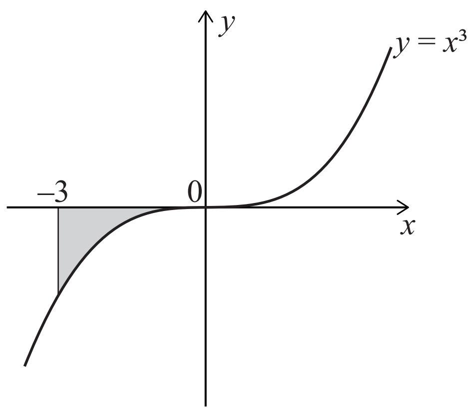
$$\int_0^3 x^2 \, \mathrm{d}x = \left[\frac{x^3}{3}\right]_0^3 = \frac{27}{3} - \frac{0}{3} = 9$$



Example 2

Calculate $\int_{-3}^0 x^3 \, \mathrm{d}x$ and draw a picture to illustrate the meaning of the answer

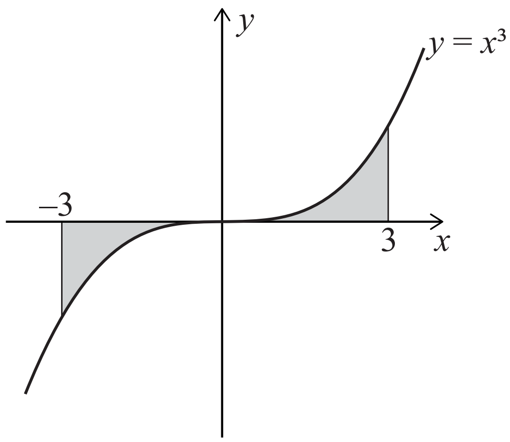
$\int_{-3}^0 x^3 \, \mathrm{d}x = \left[\frac{x^4}{4}\right]_{-3}^0 = \frac{0}{4} - \frac{(-3)^4}{4} = -\frac{81}{4}$



Example 3

Calculate $\int_{-3}^{3} x^5 \, \mathrm{d}x$ and draw a picture to illustrate the meaning of the answer.

$$\int_{-3}^{3} x^5 \, \mathrm{d}x = \left[\frac{x^6}{6}\right]_{-3}^{3} = \frac{3^6}{6} - \frac{(-3)^6}{6} = 0$$



Can you work out why the answer is zero?

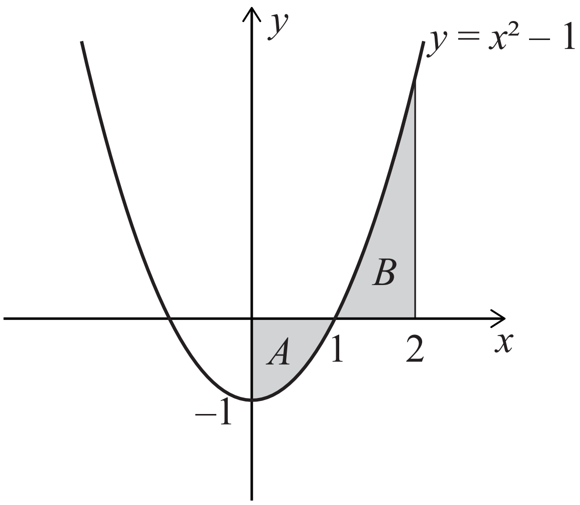
We mentioned above that there can be a subtle difference between being asked to find the area between a curve and an axis and finding a definite integral. We will explore this in a more detail.

We have already discussed that a definite integral can be interpreted as the sum of the areas above the x-axis minus the sum of the areas below the x-axis. So, if you are asked to find the area between a curve and the x-axis between two x values, the definite integral might give you the wrong answer. The answer will be wrong because to find the total area between a curve and the x axis when you are given two x values, you must find the areas that sit above the x axis and ADD them to the *positive* areas that sit below the x-axis; and we know that definite integration will subtract areas that sit underneath the x-axis. How can we calculate the area required? The simple answer is that we calculate each area separately – the areas above the x-axis and the areas below the x-axis – using the definite integral. And then we take the positive values of all these areas and add them together. Here is an example to illustrate what we mean here :

Example

Find the area between the x-axis, the lines $x = 0$ and $x = 2$ the curve $y = x^2 - 1$

Let's start by sketching the area so we can understand what we are being asked to do:



From the diagram you can see that we need to split our calculations into two pieces because we want positive areas – the question asks for areas rather than just asking for an integral. We need to find $\int_0^1 x^2 - 1\,\mathrm{d}x$ and $\int_1^2 x^2 - 1\,\mathrm{d}x$. We expect the first integral to be negative and the second to be positive. Let's calculate the two integrals and then we can see how to combine them to get the answer:

$$\int_0^1 x^2 - 1\,\mathrm{d}x = \left[\frac{x^3}{3} - x\right]_0^1 = \frac{1}{3} - 1 = -\frac{2}{3}$$

$$\int_1^2 x^2 - 1\,\mathrm{d}x \quad \left[\frac{x^3}{3} - x\right]_1^2 = \left[\frac{2^3}{3} - 2\right] - \left[\frac{1^3}{3} - 1\right] = \frac{4}{3}$$

We note that the definite integral for A is negative as expected and the definite integral for B is positive. This tells us that area A is $\frac{2}{3}$ and area B is $\frac{4}{3}$ so the total area required is $\frac{6}{3} = 2$

It is useful to contrast this answer with the definite integral from $x = 0$ to $x = 2$. You should calculate this to see what you get – your answer should be $\frac{2}{3}$

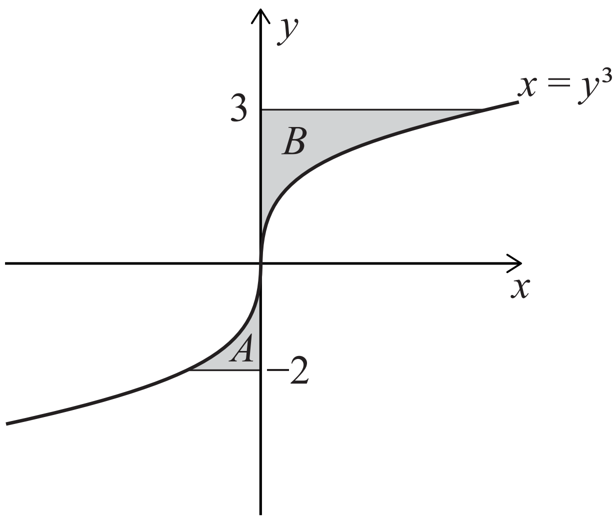
Integration with $\mathbf{d}y$

Although there is no requirement for you to deal with integrals with $\mathrm{d}y$ in them, it is useful to know about them all the same: for instance, $\int_{-2}^{3} y^3 \ \mathrm{d}y$. We call these integrals with respect to [w.r.t[68]] y and we need to make sure that what is being integrated is expressed in terms of y.

These integrals are calculated exactly the same way as those with x in them:

$$\int_{-2}^{3} y^3 \ \mathrm{d}y \ = \left[\frac{y^4}{4}\right]_{-2}^{3} = \frac{3^4}{4} - \frac{(-2)^4}{4}$$

And we can draw a picture to illustrate the region that this integral applies to [note the integral corresponding to A is negative, and to B is positive] :



Integration "tricks"

Finally, there are few "tricks" you should be aware of that can sometimes make integration easier. The ones we will look at here [albeit briefly] are ones that involve using the symmetry or asymmetry of graphs to simplify definite integrals.

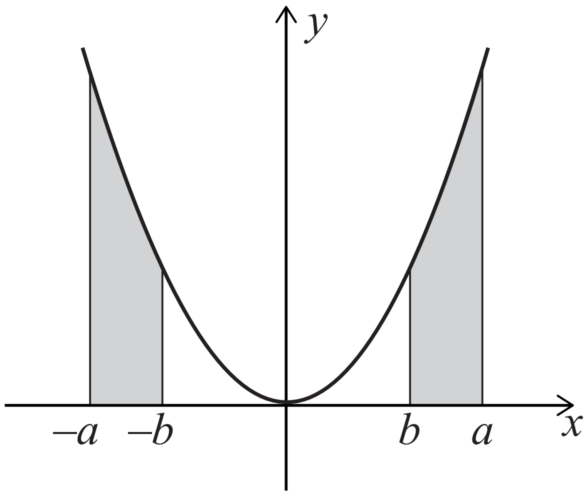
If a graph of a function $y = f(x)$ is symmetric about the y axis [when the graph is reflected in the y axis it looks the same] the following must be true:

$$\int_{-a}^{-b} f(x)\, \mathrm{d}x = \int_{b}^{a} f(x)\, \mathrm{d}x$$

---

[68] "wrt" is a very common abbreviations in mathematics.

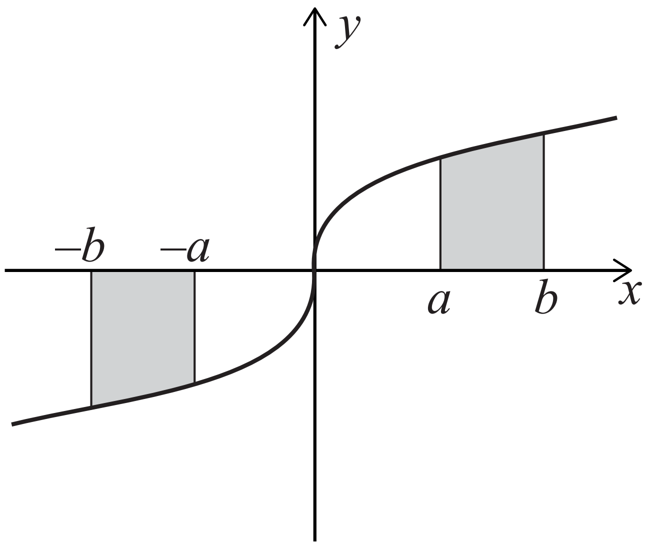
And we can see why using a diagram:



If a graph of a function $y = f(x)$ is antisymmetric [when the graph is reflected in the y axis and then in the x-axis, it looks the same] the following must be true:

$$\int_{-b}^{-a} f(x)\,\mathrm{d}x = -\int_{a}^{b} f(x)\,\mathrm{d}x$$

And we can see why using a diagram:



Exercise

Using these symmetry ideas and your knowledge of integration, you should be able to explain why[69] the following are true:

$$\int_0^{2\pi} \cos x \, \mathrm{d}x = 0$$

$$\int_0^{2\pi} \sin x \, \mathrm{d}x = 0$$

$$\int_{-\pi}^{\pi} \sin x \, \mathrm{d}x = 0$$

$$\int_{-\pi}^{\pi} \tan x \, \mathrm{d}x = 0$$

$$\int_{-10}^{10} x^3 \, \mathrm{d}x = 0$$

$$\int_{-\pi}^{\pi} x^2 \, \mathrm{d}x = 2\int_0^{\pi} x^2 \, \mathrm{d}x$$

$$\int_0^{\pi} \sin x \, \mathrm{d}x = -\int_{\pi}^{2\pi} \sin x \, \mathrm{d}x$$

[69] Although trigonometric integration [and differentiation] is not on the TMUA/ESAT specification, all these examples are expression we could expect you to deduce within the TMUA/ESAT specification.

MM7.2

Finding definite and indefinite integrals of x^n for n rational, $n \neq 1$, and related sums and differences, including expressions which require simplification prior to integrating.

For example: $\int (x+2)^2 \mathrm{d}x$ and $\int \frac{(3x-5)^2}{x^{\frac{1}{2}}} \mathrm{d}x$

In the TMUA/ESAT we expect you to be able to integrate sums of terms in powers of x using the rule :

$$\int kx^n \ \mathrm{d}x = \frac{kx^{n+1}}{n+1} + c \qquad n \neq -1$$

Where k and c are a real constant and n is any real number except -1 . When $n = -1$ the answer falls outside the scope of TMUA/ESAT.[70]

Any integration that you are required to do in the TMUA/ESAT, if it does not require symmetry or other arguments, will be an integration of sums of powers of x. We will not ask question that require more sophisticated methods such as substitution or integration by parts etc., and we are careful to make sure every question we might ask involving integration can be answered equally efficiently using the basic methods outlined here, even if it turns out more advanced methods could also be applied.[71]

In the TMUA/ESAT the expression you are given might not look like powers of x but it will be possible to simplify it into such a sum. Here is an example:

$$\int (x-2)^2 \ \mathrm{d}x = \int x^2 - 4x + 4 \ \mathrm{d}x = \int x^2 \ \mathrm{d}x - \int 4x \ \mathrm{d}x + \int 4 \ \mathrm{d}x$$
$$= \frac{x^3}{3} - \frac{4x^2}{2} + 4x + c$$

Here it is useful to notice something about integration that is often taken for granted without much thought: when you integrate a simple sum, you can integrate term by term and add [or subtract] the individual answers. It is always worth thinking very carefully about the properties of the mathematics you meet to ensure you do not inadvertently perform mathematical moves that might seem right but which are, in fact, invalid.[72]

---

[70] You might have seen: $\int \frac{1}{x} \mathrm{d}x \ \ln|x| + c$. Note that $\ln|x|$ is the natural logarithm of x and this topic is NOT on the TMUA/ESAT.

[71] We are careful in the TMUA/ESAT to ensure that knowing more advanced techniques does not give anyone undue advantage.

[72] For instance, it might be tempting to write $f(x+3) = f(x) + f(3)$ or $f(x^2) = [f(x)]^2$ but these are not generally true. Or you might be tempted to do this [which is **very wrong, so don't do it**]: $\int \frac{f(x)}{g(x)} \mathrm{d}x = \frac{\int f(x) \mathrm{d}x}{\int g(x) \ \mathrm{d}x}$

MM7.3

An understanding of the Fundamental Theorem of Calculus and its significance to integration.

Simple examples of its use may be required in the forms:

$\int_a^b f(x)dx = F(b) - F(a)$, where $F'(x) = f(x)$

$\frac{d}{dx}\int_a^x f(x)dx = f(x)$

You already know both expressions in this section, although you might not realise you know them. We will explore each expression in turn:

First:

$$\int_a^b f(x)\,dx = F(b) - F(a), \text{where } F'(x) = f(x)$$

This links the idea that integration is the reverse of differentiation and the method you are familiar with for calculating definite integrals. $F'(x) = f(x)$ tells you that when you integrate $f(x)$ you get $F(x)$; and $\int_a^b f(x)\,dx = F(b) - F(a)$ tells you how to put limits into the expression you get once you have integrated.

A useful consequence of this expression is the following – essentially if you swap the limits around you introduce a minus sign into the integral:

$$\int_a^b f(x)\,\mathrm{d}x = -\int_b^a f(x)\,\mathrm{d}x$$

You should check you can see why this works.

And the Fundamental Theorem also allows you to write integrals in other ways:

$$\int_a^b f(x)\,\mathrm{d}x = \int_a^c f(x)\,\mathrm{d}x + \int_c^b f(x)\,\mathrm{d}x$$

It is tempting to add the condition $a < c < b$, that is that c must be between a and b; but we have not added that condition. Can you work out why not ? You do have to be a bit careful though: for instance, it might be the case that $f(x)$ is not defined at c or at some other interval [i.e. part of the x axis] that you might have wanted to integrate over by introducing c.

The second expression in the specification $\frac{d}{dx}\int_a^x f(t)\,dt = f(x)$ suggest that if you integrate and then differentiate, you should get back to what you started with [note we had to put an x in the limits because if we had just put two numbers as limits, then the integral would be constant and then the differentiation would give 0]. You need to be a little careful with this expression though, as it is not as simple as it first appears. Have a look at TMUA 2020 paper 2 question 16.

MM7.4

Combining integrals with either equal or contiguous ranges.

For example:

$\int_2^5 f(x)dx + \int_2^5 g(x)dx = \int_2^5 [f(x) + g(x)]dx$

$\int_2^4 f(x)dx + \int_4^3 f(x)dx = \int_2^3 f(x)dx$

This section is really just a reiteration of what we have said earlier in these notes:

- You can integrate term by term or all at once in an integral:

$$\int_2^5 f(x)\,dx + \int_2^5 g(x)\,dx = \int_2^5 [f(x) + g(x)]\,dx$$

- You can use the Fundamental Theorem of Calculus to simplify expressions:

$$\int_2^4 f(x)\,\mathrm{d}x + \int_4^3 f(x)\,\mathrm{d}x = \int_2^3 f(x)\,\mathrm{d}x$$

It is worth checking that you have an intuitive grasp of both of these and also that you have a formal understand of why they work.

We will use the Fundamental Theorem of Calculus to unpack the second statement:

$$\int_2^4 f(x)\,\mathrm{d}x + \int_4^3 f(x)\,\mathrm{d}x = \int_2^3 f(x)\,\mathrm{d}x + \int_3^4 f(x)\,\mathrm{d}x - \int_3^4 f(x)\,\mathrm{d}x$$
$$= \int_2^3 f(x)\,\mathrm{d}x$$

Check you can see what we have done in the middle section.

MM7.5

Approximation of the area under a curve using the trapezium rule; determination of whether this constitutes an overestimate or an underestimate.

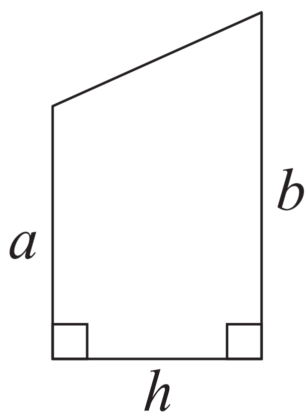
We expect you to be able to use the Trapezium rule to estimate areas under curves [and recall we take area to be positive] or to estimate the values of definite integrals [remember definite integrals take "areas" under the x-axis as negative]. We will make sure that any question we ask in the TMUA/ESAT is very clear about whether it is asking for an estimate of areas between a curve and an axis or whether it is asking for an estimate of a definite integral.

You should either learn the Trapezium rule formula [but make sure you understand it] or you should be able to calculate the result from scratch using your knowledge of the area of a trapezium. We ALWAYS assume that the trapezium rule finds an estimate of an area using equal width strips.

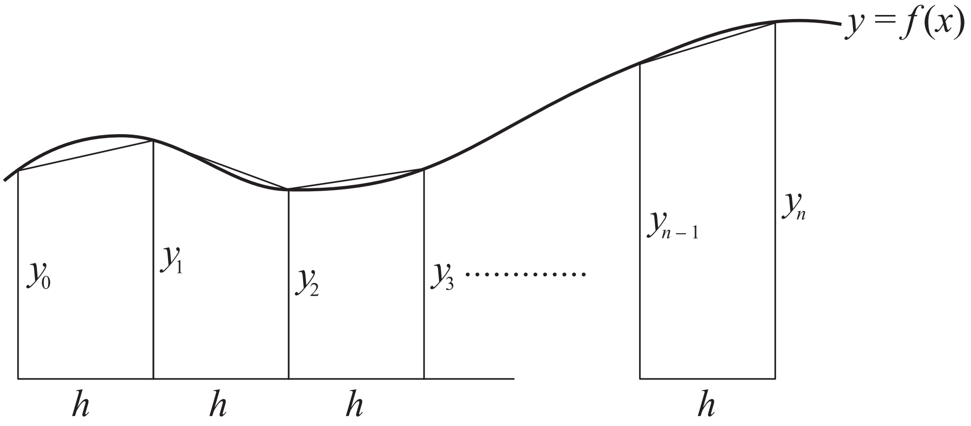
Let's briefly look at the Trapezium rule and how it works:

The area of a trapezium with two right angles in it – as shown - is

$$area = h \times \frac{a+b}{2} = \frac{h}{2}(a+b)$$



Using this formula, we can see how we could estimate the area under a curve using a set of equal width trapezia. We will use n trapezia each of width h and the heights of each trapezium can be calculated from the function $y = f(x)$ whose area we are approximating:



$$approximate\ area = \frac{h}{2}(y_0 + y_1) + \frac{h}{2}(y_1 + y_2) + \frac{h}{2}(y_2 + y_3) + \cdots + \frac{h}{2}(y_{n-1} + y_n)$$

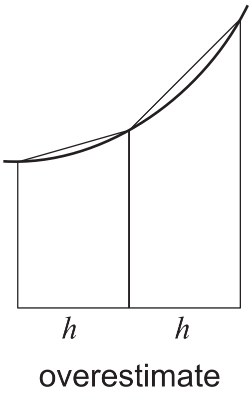
And this simplifies to

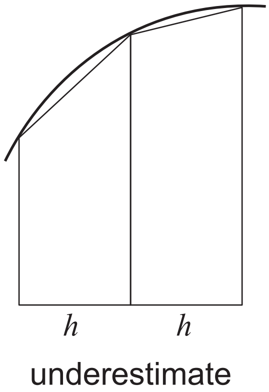
$$approximate\ area = \frac{h}{2}(y_0 + 2y_1 + 2y_2 + 2y_3 \ldots + 2y_{n-1} + y_n)$$

And this is one form of the Trapezium rule. Notice that every y_k appears twice except the first one y_0 and the last one y_n ; this is not surprising as every y_k [except y_0 and y_n] is the shared side for two trapezia.

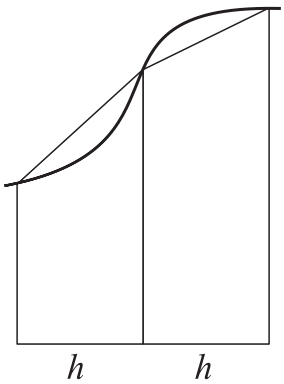
We will not ask you questions that involve complicated calculations as we are interested in checking your understanding of mathematics in the TMUA/ESAT rather than your ability to add lots of numbers together correctly!

You should be able to tell whether the result of the trapezium rule is an overestimate or an underestimate using your understanding of the shapes of curves [some of which we meet in the next section].





Sometimes, it is not possible to tell if the trapezium rule gives an overestimate or underestimate without further work :



MM7.6

Solving differential equations of the form $\frac{dy}{dx} = f(x)$

Solving the expression $\frac{dy}{dx} = f(x)$ is really asking you to find what y is, expressed in terms of x, such that when you differentiate y you get $f(x)$. This is a bit of a mouthful. Another way of saying this is: what do you differentiate to get $f(x)$?

We will look at a couple of examples, firstly without any additional conditions and then with additional conditions [you will see what we mean by additional conditions below]:

Example

Solve $\frac{\mathrm{d}y}{\mathrm{d}x} = 3x^2 + 4x - 3$

To find y we integrate both sides with respect to [wrt] x:

When we integrate $\frac{\mathrm{d}y}{\mathrm{d}x}$ wrt x we get y [see Fundamental Theorem of Calculus above]

And when we integrate $3x^2 + 4x - 3$ wrt x we get $x^3 + 2x^2 - 3x + c$

So, the solution is

$$y = x^3 + 2x^2 - 3x + c$$

A few things to note in this example:

When integrating $\frac{\mathrm{d}y}{\mathrm{d}x}$ we found $\int \frac{\mathrm{d}y}{\mathrm{d}x}\, \mathrm{d}x$ and we used the Fundamental Theorem of Calculus which tells us that integration and differentiation are linked. So, if we differentiate y and then integrate it we should get just y back. Well actually, we should write $y + constant$ but we don't as we tend to assume all the constants generated from integration are combined on the x side of the solution and written as "c"

We can now look at an example with some additional condition attached to it:

Example

Given $y = 5$ when $x = 1$ and $\frac{\mathrm{d}y}{\mathrm{d}x} = 3x^2 + 4x - 3$, find y in terms of x

[Numerical values such as $y = 5$ when $x = 1$ are sometimes called "boundary conditions" or "initial conditions" depending on the circumstances and values given. We won't use these terms in the TMUA/ESAT.]

We can solve this two [entirely equivalent] ways.

Method 1

We can work out the answer as we did above with a constant term c in it and then use the condition $y = 5$ when $x = 1$ to find the value of c by substitution:

$$y = x^3 + 2x^2 - 3x + c$$

So $y = 5$ when $x = 1$ gives

$$5 = 1 + 2 - 3 + c$$

So $c = 5$

And the solution is

$$y = x^3 + 2x^2 - 3x + 5$$

Method 2

We notice that $y = 5$ when $x = 1$ and y is y when x is x, and so we put these values directly as *corresponding* limits [top limits go together, and bottom limits go together] in the integration:

$$\int_{y=5}^{y=y} \frac{\mathrm{d}y}{\mathrm{d}x} \ \mathrm{d}x = \int_{x=1}^{x=x} 3x^2 + 4x - 3 \ \mathrm{d}x$$

Don't worry too much about the expression on the left[73] [but do think a little about what it means]. This leads to:

[73] We have been a little slapdash with our notation on the left-hand side. As we are integrating wrt x then the limits ought to be x values and not the corresponding y values but it is easier just to put the y values in as those are the ones we will end up using when we have integrated.

$$[y]_5^y = [x^3 + 2x^2 - 3x]_1^x$$

Which gives:

$$y - 5 = (x^3 + 2x^2 - 3x) - (1 + 2 - 3)$$

Which gives, as above:

$$y = x^3 + 2x^2 - 3x + 5$$

MM8 Graphs of Functions

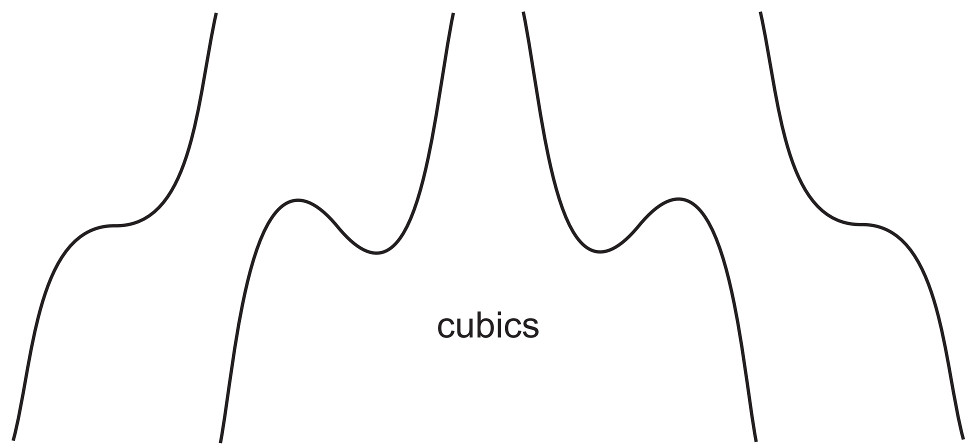
Much of this section has been covered earlier in these notes. And so, we will concentrate mainly on the parts of the specification that have not been covered in detail elsewhere and add a few additional comments when useful.

MM8.1

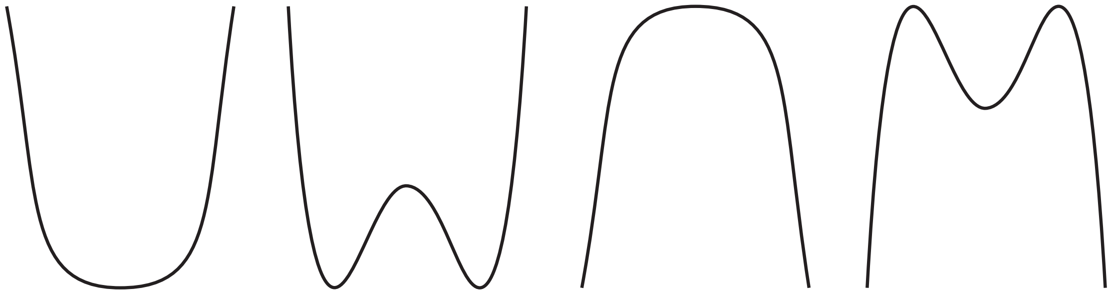
Recognise and be able to sketch the graphs of common functions that appear in this specification: these include lines, quadratics, cubics, trigonometric functions, logarithmic functions, exponential functions, square roots, and the modulus function.

You should be able to sketch any of the following given the corresponding equation: lines, quadratics, cubics, trigonometric functions, logarithmic functions, exponential functions, and square roots.

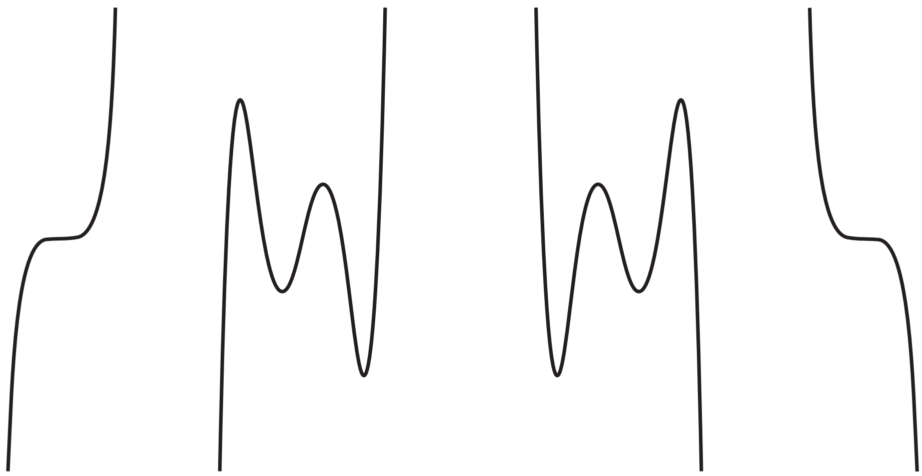
We do expect you to have a good understanding of the possible shapes of quadratics, cubics and also graphs of higher powers of x such as quartics and quintics. We recommend you spend some time experimenting with different equations of such polynomials using a good graph drawing package [e.g., DESMOS GRAPHING]. When drawing cubics, quartics and quintics, pay attention to the sorts of shapes you get and think about the range of the possible shapes for such graphs and what happens when the highest power of x [the x^2 term in a quadratic, the x^3 term in a cubic, the x^4 term in a quartic, etc] has a positive coefficient and when it has a negative coefficient. We have drawn a few examples below to help you get an idea of the sorts of shapes to expect [but, be aware that not all possible shapes are shown!!]:



quartics



quintics



We will now explore the modulus function. We met this earlier in these notes. There we said that the modulus function effectively makes values positive. It is written as two vertical lines, one either side of what it applies to and it can be thought of as saying “take the positive value of this …”.

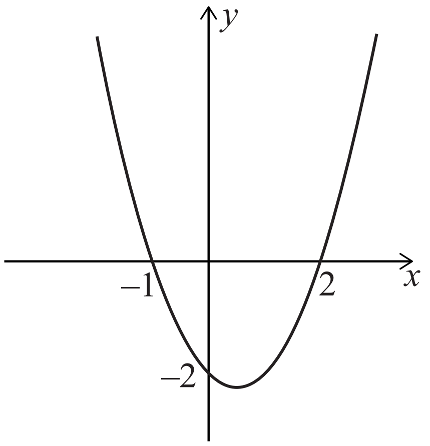
In the TMUA/ESAT we expect you to be able to deal with simple modulus functions both algebraically [which we covered earlier] and also graphically. In this section, we will explore the graphical aspects of modulus functions.

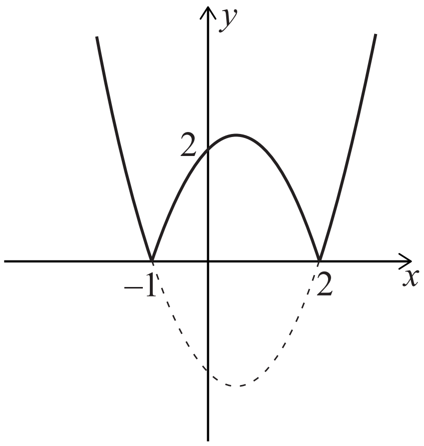
We will start with a couple of examples and then we will summarise how to go about sketching graphs involving the modulus function.[74]

[74] You should also experiment with graphs and using the modules function yourself using a graphical drawing package such as DESMOS GRAPHING

Example

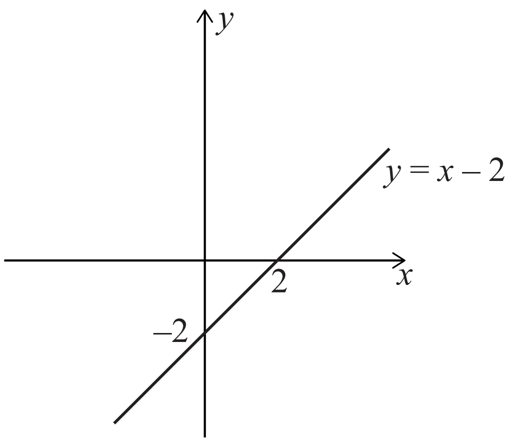
Here are sketches of $y = (x+1)(x-2)$ and then $y = |(x+1)(x-2)|$

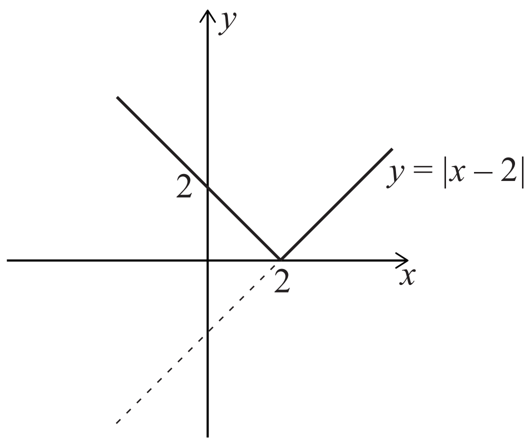




Example

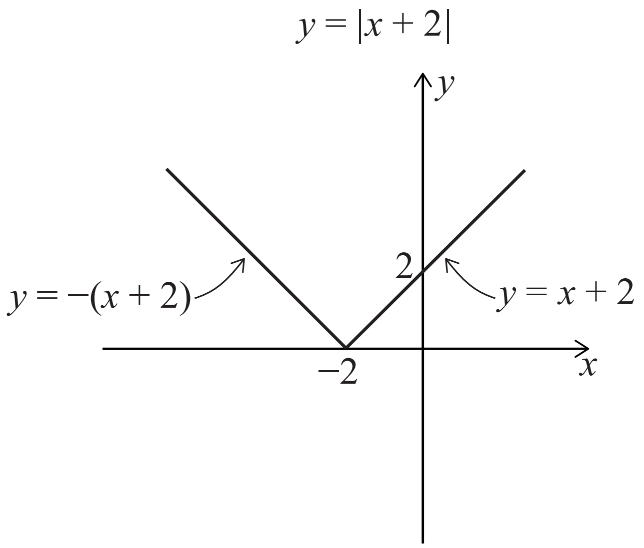
Here are sketches of $y = x - 2$ and $y = |x-2|$

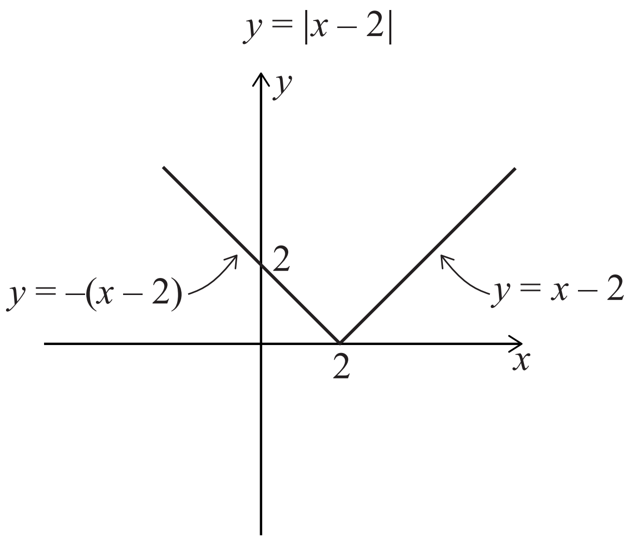


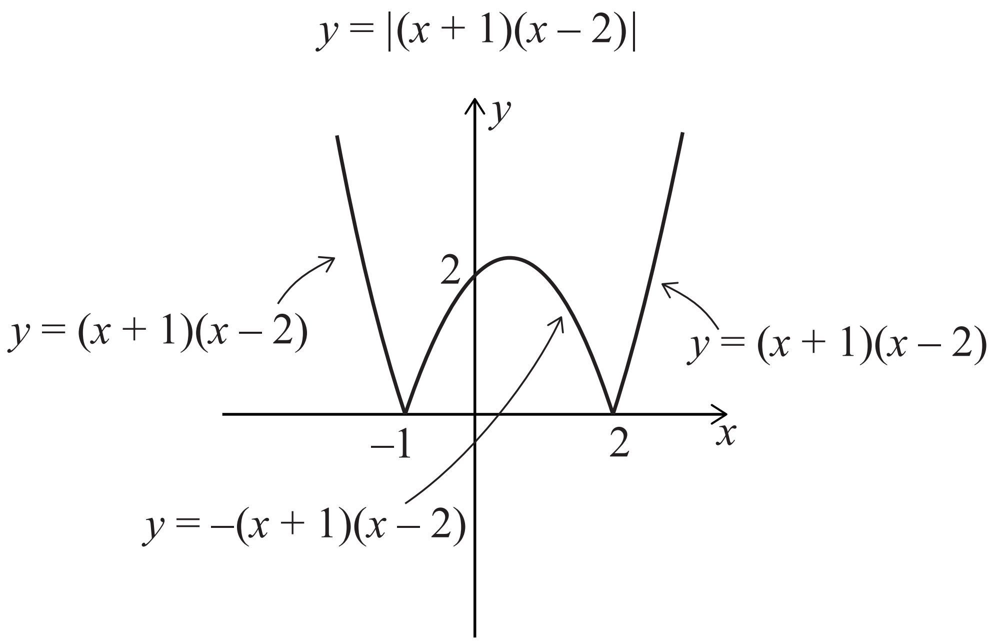


In simple terms, if you need to sketch $y = |f(x)|$ for some $f(x)$ then you first sketch $y = f(x)$ and then reflect everything that sits below the x-axis to be above the x-axis and leave everything that sits above the x-axis where it is unchanged. The result is a sketch of the graph of $y = |f(x)|$ This makes sense because the modulus function tells you to take the positive value of something. When $f(x)$ is positive then nothing needs altering, but when $f(x)$ is negative [i.e. it sits below the x-axis] then you need to find the equivalent positive value – and this means you “flip it” over in the x-axis

It is useful to know the equations of bits of modulus graphs when they are flipped. Here are a couple of examples [again, make sure you understand why the equations alter the way we have indicated: the flipped bits have equation $y = -f(x)$ and the unflipped bits have equation $y = f(x)$]





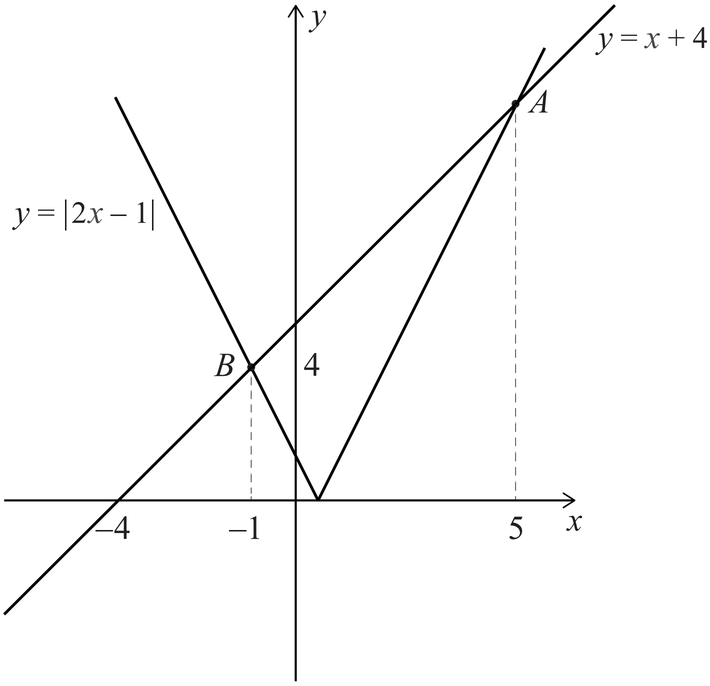


And these are useful when solving equations involving modulus functions using a mixture of algebra and graphs. Here is an example [nest page]:

Example

Solve: $x + 4 \ < \ |2x - 1|$

We sketch both $y = x + 4$ and $y = |2x - 1|$



We find the x-coordinates of where they cross.

Point A is a where $x + 4 \ = 2x - 1$ so at $x = 5$

Point B is where $x + 4 \ = \ -(2x - 1)$ so at $x = \ -1$

And then we can solve the inequality – which is asking "when does the graph of $y = x + 4$ sit below the graph of $y = |2x - 1|$?"

The answer must be when $x < -1$ or $x > 5$

MM8.2

Knowledge of the effect of simple transformations on the graph of $y = f(x)$ with positive or negative value of a as represented by $y = af(x)$, $y = f(x) + a$, $y = f(x + a)$, $y = f(ax)$

Compositions of these transformations. Knowledge and use of the notation $f\big(g(x)\big)$.

This topic tends to be poorly understood. Usually, when these ideas are first met, students tend to learn the rules without much understanding of what is going on. It is made more tricky by the fact that some of the ways graphs shift tend to be exactly opposite of what you might first expect; for instance, $y = f(x + a)$ looks like it ought to shift [translate] the graph of $y = f(x)$ to the right [in the positive x direction] by a distance a BUT THAT IS WRONG!!

We will look at each of the case in turn and explain what transformation of $y = f(x)$ each represents and then we will look at a couple of examples.

The first thing to get clear is what the notation means. You are already familiar with the notation $y = f(x)$. This notation tells you that the y value "above" [i.e. the y value for the point on the curve corresponding to the given x value] any x value is calculated using $f(x)$. So, as an example, for $y = x^2 + 3$ the y value above $x = 2$ must be $y = 2^2 + 3 = 7,$ and similarly the y value above $x = 4$ is 19, and so on.

We can use this to understand what the notation in this section means and then use our understanding to deduce how the graphs are related to the graph of $y = f(x)$.

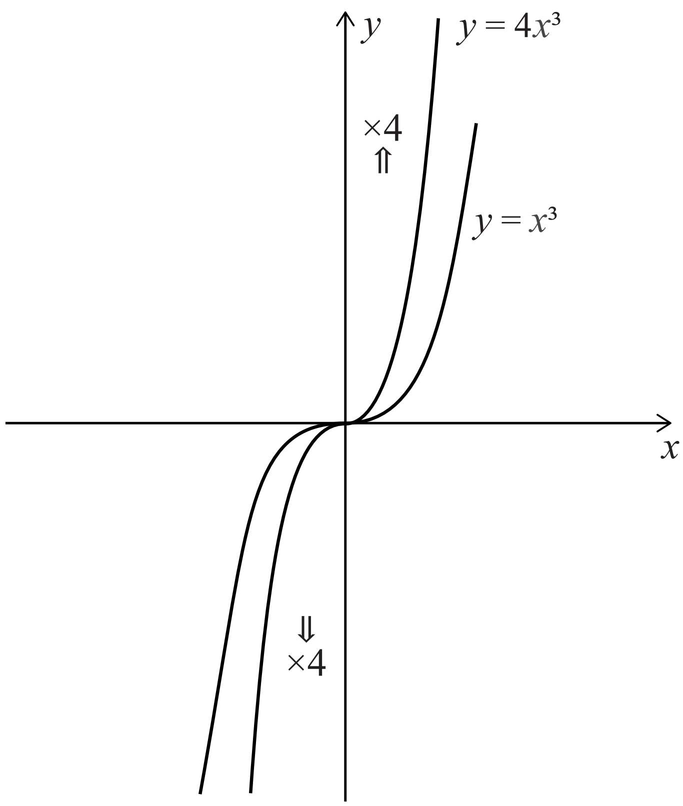
$y = af(x)$

We will look at specific example: We will take $y = f(x)$ to be $y = x^3$ and we will take $a = 4$

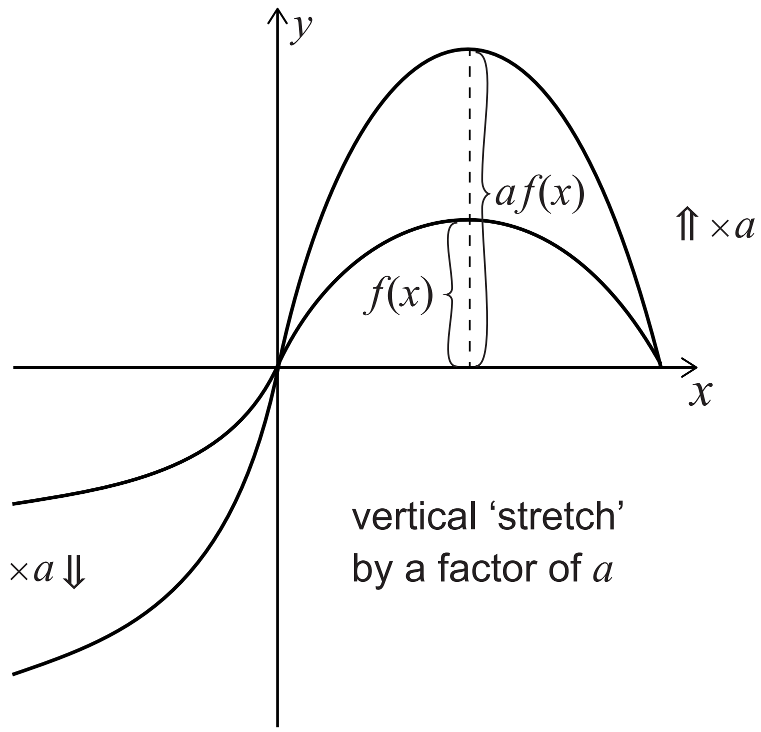
We want to compare the graphs of $y = x^3$ [$y = f(x)$] with $y = 4x^3$ [$y = 4f(x)$]

Here the transformation is reasonably straightforward to grasp: each y value is four times as big for $y = 4x^3$ than it is for $y = x^3$. This is like taking the graph of $y = x^3$ and stretching it vertically by a factor of 4, and the vertical stretching is away from the x axis up [when the y values are positive] or down [when the y values are negative]

Here is a diagram of the before and after:



And here is a general diagram



And note that if $0 < a < 1$ then the graphs effectively get less tall : so, if $a = \frac{1}{2}$ then the graphs of $y = \frac{1}{2}f(x)$ would be half the height of $y = f(x)$

And if a is negative the heights change by $|a|$ and the graphs are flipped by the minus signs.

Exercise

Use a graph drawing package [e.g., DESMOS GRAPHING] to explore the following pairs of functions:

$y = f(x) = x^2$ and $y = af(x) = ax^2$ for $a = 2, 3, -1, -2, \frac{1}{2}, -\frac{1}{2}$

$y = f(x) = \cos x$ and $y = af(x) = a\cos x$ for $a = 2, 3, -1, -2, \frac{1}{2}, -\frac{1}{2}$

What do you notice when $0 < a < 1$?

What do you notice when $a < 0$?

$y = f(x) + a$

We will look at specific example: we will take $y = f(x)$ to be $y = x^2$ and we will take $a = 3$

We want to compare the graphs of $y = x^2$ [$y = f(x)$] with $y = x^2 + 3$ [$y = f(x) + a$]

If we sketch both graphs [you can do this on DESMOS GRAPHING], we can see that going from $y = x^2$ to $y = x^2 + 3$ all the y values go up by 3 units. In other words, the graph is shifted up by 3 units ["up" means parallel to the y-axis]. We can say this more formally by saying that we translate the graph by $\binom{0}{3}$

In general, $y = f(x) + a$ takes the graph of $y = f(x)$ and translates it by $\binom{0}{a}$

Exercise

Use a graph drawing package [e.g., DESMOS GRAPHING] to explore the following pairs of functions:

$y = f(x) = x^2$ and $y = f(x) + a = x^2 + a$ for $a = 2, 3, -1, -2, \frac{1}{2}, -\frac{1}{2}$

$y = f(x) = \cos x$ and $y = f(x) + a = a + \cos x$ for $a = 2, 3, -1, -2, \frac{1}{2}, -\frac{1}{2}$

What do you notice when $a < 0$?

An aside: note here we have written $a + \cos x$ instead of $\cos x + a$ as this latter expression is ambiguous – it could mean either $(\cos x) + a$, which is what we intend, or it could mean $\cos(x + a)$ which is not what we intend. This sort of issue is not uncommon with trigonometry so you need to be a little careful with how you write things and how you interpret things – there are usually conventions that all mathematicians follow [for example contrasting $\cos^2 x$ with $\cos x^2$]. A common example is the inverse trigonometric functions, which are often written as, for instance, $\cos^{-1} x$. Here the -1 is **<u>not</u>** taken to mean $\frac{1}{\cos x}$ [as we might initially expect from our discussion of indices above] but instead the convention is that it means "the inverse of cos" which is sometimes written as arcos. Later in your maths courses, you will probably learn things like "the secant of x" [or $\sec x$] etc, which are the specific symbols used by mathematicians for $\frac{1}{\cos x}$. In the TMUA/ESAT, we are very careful with the way we use notation to ensure these sorts of ambiguities do not arise; and if there is any potential ambiguity we make sure we clarify things carefully in the way we phrase a question, or in the way we set out the mathematics.

$y = f(x + a)$

This particular transformation is often poorly understood and can lead to errors. Errors and misunderstandings arise because it seems [perhaps intuitively at first glance?] that if you ADD something to an x then things should "shift to the right"; whereas, in fact, the opposite happens – curves shift "to the left" [when a is positive]. Of course, you could just learn what happens for this transformation, but it is [much much] better, as always, to understand things. We will unpack this transformation in following discussion: take your time working through our discussion to make sure you develop a good understanding.

There are two things to unpack here. One is how to work out an expression for $f(x + a)$ given $f(x)$; and the other is to work out how the transformation relates to the graphs of $y = f(x)$ and $y = f(x + a)$.

First, let's tackle how to work out an expression for $f(x + a)$ given $f(x)$. This is straightforward and we can look at a couple of examples to see how it works:

Example

Given $f(x) = x^2 + 2x - 5$ find an expression for $f(x + 3)$

We do this as follows: every x in the expression $f(x)$ is replaced by $x + 3$:

$x^2 + 2x - 5 \ \rightarrow \ (x + 3)^2 + 2(x + 3) - 5$

So $f(x + 3) = (x + 3)^2 + 2(x + 3) - 5$

Example

Given $f(x) = \cos(2x)$ find an expression for $f(x - \frac{\pi}{2})$

We do this as follows: every x in the expression $f(x)$ is replaced by $x - \frac{\pi}{2}$:

$$f\left(x - \frac{\pi}{2}\right) = \cos 2\left(x - \frac{\pi}{2}\right) = \cos(2x - \pi)$$

We could simplify this further, but the mathematics needed to do so is outside the scope of the TMUA/ESAT. If you sketch $y = \cos(2x - \pi)$ you might be able to work out what it could simplify to.

In this example, it is quite easy to forget that the 2 in $\cos(2x)$ multiples everything that we replace x by: so we must have $\cos 2\left(x - \frac{\pi}{2}\right) = \cos(2x - \pi)$ and NOT $\cos 2x - \frac{\pi}{2}$

Now you know how to work out an expression for $f(x+a)$ given $f(x)$, we turn to look at how the graphs of each of $y = f(x)$ and $y = f(x+a)$ relate to each other.

First, we need to be very clear what the notation is telling us:

$y = f(x)$ tells us that the y value directly above a given x is calculated using $f(x)$

$y = f(x+a)$ tells us that the y value directly above a given x value is calculated using $f(x+a)$

Let's check this is clear using an example.

Example

Let's look at:

$f(x) = 2^x$ and $f(x+3)$

If we were to sketch $y = f(x)$ we can calculate the y values for some x values:

When $x = 2$, $y = 2^2 = 4$

When $x = 5$, $y = 2^5 = 32$

So if we sketch $y = f(x)$ we would find that

when x is 2 the y value directly above it would be 4

when x is 5 the y value directly above it would be 32

Now let's look at what happens if we were to sketch $y = f(x+3)$

When $x = 2$, the y value directly above it must $f(2+3)$ which is $f(5)$ and we worked that out to be $y = 2^5 = 32$

When $x = 5$, the y value directly above it must $f(5+3)$ which is $f(8)$ and we can work that out to be $y = 2^8 = 256$

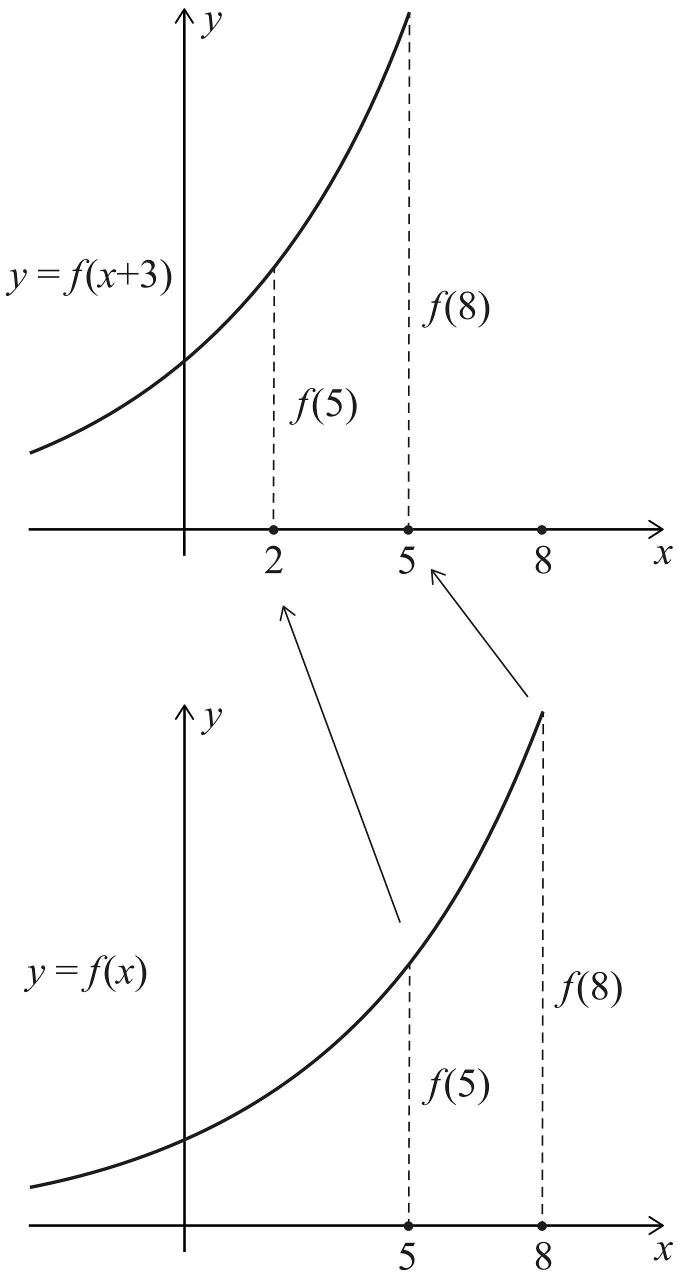
So [and think about this carefully] the y value above a given x value in $y = f(x+3)$ comes from the y above the x value that is three units further along on the sketch of $y = f(x)$:

The y value above $x = 2$ is actually $f(2+3) = f(5)$

The y value above $x = 5$ is actually $f(5+3) = f(8)$

We have to translate the graph of $y = f(x)$ to the left to make sure that the y value above $x = 2$ in the graph of $y = f(x+3)$ is the one from $f(5)$ and that the y value above $x = 5$ is the one from $f(8)$.

We can draw a sketch of this to show what is happening:



So, we can now look at this in general terms. If you look at the graph of $y = f(x + a)$ then the y value above an x value is actually $f(x + a)$ and this is the y value that is above $x + a$ on the original graph.

This means that the graph of $y = f(x + a)$ is the same as the graph of $y = f(x)$ when it is translated backwards parallel to the x axis a distance of a. We can say:

$\boldsymbol{y = f(x + a)}$ **is the same as the graph of** $\boldsymbol{y = f(x)}$ **translated by** $\boldsymbol{\binom{-a}{0}}$

And note that if a is negative [e.g., $y = f(x - 4)$] then the graph shifts "to the right" by 4, that is a translation of $\binom{4}{0}$

It is worth thinking carefully about this transformation – it can seem a little complicated with lots of $f(x)$ and $f(x + a)$ and y values flying about. But once you have grasped what is going on, it can all seem very easy and "obvious".[75]

We strongly recommend you do the following exercise to get used to this transformation.

Exercise

Sketch the following in pairs of functions without using a graphing package, and then check your answers using a graphing package [such as DESMOS GRAPHING]. Think about how each pair of graphs relates to what we discussed above when unpacking the transformation from $y = f(x)$ to $y = f(x + a)$

$y = f(x) = x^2$ and $y = f(x + 2) = (x + 2)^2$

$y = f(x) = x^2$ and $y = f(x - 3) = (x - 3)^2$

$y = f(x) = \cos x$ and $y = f\left(x + \frac{\pi}{3}\right) = \cos\left(x + \frac{\pi}{3}\right)$

$y = f(x) = \cos x$ and $y = f\left(x - \frac{2\pi}{3}\right) = \cos\left(x - \frac{2\pi}{3}\right)$

[75] See footnote 59.

$y = f(ax)$

This transformation is very similar to the one we have just looked at. It is initially counter-intuitive but the reason it "squashes" a graph by a factor of a is "obvious"[76] once you have thought it through with some examples.

We will first look at some examples that illustrate how to find the expression for $f(ax)$ given an expression for $f(x)$; and then we will look at how the graphs of $f(ax)$ and $f(x)$ relate.

Example

Given $y = f(x) = x^2$ find an expression for $f(3x)$

To find $f(3x)$ we replace x by $3x$: $f(3x) = (3x)^2 = 9x^2$

And note that we do NOT write $f(3x) = 3x^2$

Example

Given $y = f(x) = \cos(2x + 30)$ find an expression for $f(4x)$

As before, we replace every x in $f(x)$ by $4x$ to get

$f(4x) = \cos(2(4x) + 30) = \cos(8x + 30)$

Now, let's turn to look at how the graphs of $y = f(x)$ and $y = f(ax)$ are related. We will do this using simple examples. You will note that the discussion is similar to that we set out for $f(a + x)$ above.

[76] See footnote 75.

Example

Let's look at:

$f(x) = x^3 - 3x^2 + 2$ and $f(2x)$

If we were to sketch $y = f(x)$ we can calculate the y values for some x values:

When $x = -1$, $y = -2$

When $x = 0$, $y = 2$

When $x = 1$, $y = 0$

When $x = 2$, $y = -2$

So, if we sketch $y = f(x)$ we would find that

when x is -1 the y value directly above it would be -2

when x is 2 the y value directly above it would be -2

….and so on

Now let's look at what happens if we were to sketch $y = f(2x)$

When $x = -1$, the y value directly above it must $f(2 \times -1)$ which is $f(-2)$ and this is -18

When $x = 0$, the y value directly above it must $f(2 \times 0)$ which is $f(0)$ and this is 2

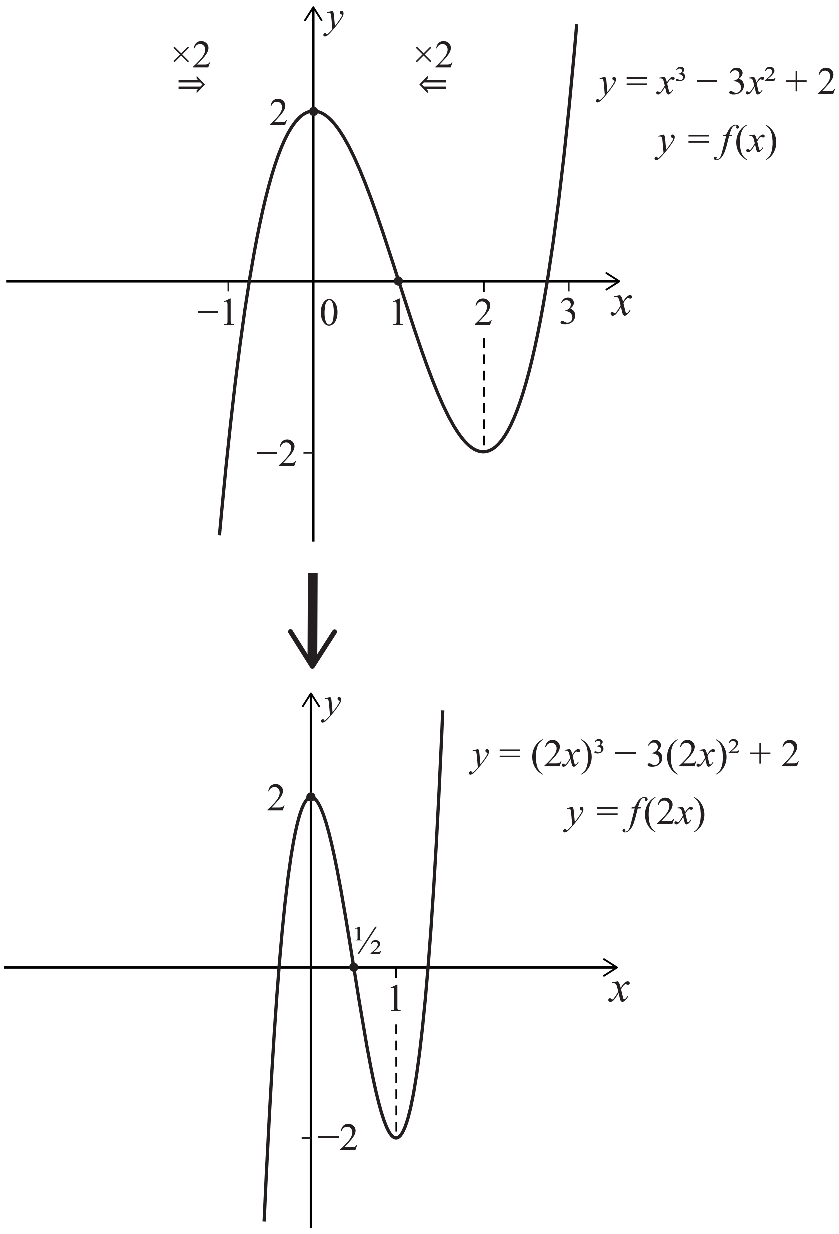
When $x = 1$, the y value directly above it must $f(2 \times 1)$ which is $f(2)$ and this is -2

When $x = 2$, the y value directly above it must $f(2 \times 2)$ which is $f(4)$ and this is 18

So [and think about this carefully] the y value above a given x value in $y = f(2x)$ comes from the y above the x value that is twice as large on the x-axis on the sketch of $y = f(x)$

We have to "squash" the graph of $y = f(x)$ towards the y-axis by a factor of 2.

We can draw some diagrams to show what is happening here:



×2
⇒
×2
⇐
y
2
−1
0
1
2
3
x
−2
y = x³ − 3x² + 2
y = f(x)
y
2
½
1
x
−2
y = (2x)³ − 3(2x)² + 2
y = f(2x)

Exercise

Using a graphing package [such as DESMOS GRAPHING], draw the following pairs of functions:

From the example above: $f(x) = x^3 - 3x^2 + 2$ and $f(2x) = (2x)^3 - 3(2x)^2 + 2$

Then look at $f(x) = x^3 - 3x^2 + 2$ and $f\left(\frac{1}{2}x\right) = \left(\frac{1}{2}x\right)^3 - 3\left(\frac{1}{2}x\right)^2 + 2$

What does this tell you about $y = f(ax)$ when $0 < a < 1$?

Now look at $f(x) = x^3 - 3x^2 + 2$ and $f(-2x) = (-2x)^3 - 3(-2x)^2 + 2$; in this, what does the minus sign do in $f(-2x)$ and what does the 2 do in $f(-2x)$? Can you explain your answers?

Let's summarise the transformations we have looked at.

$af(x)$	Vertical [parallel to y-axis] stretch away from x-axis by a factor of a. If $a < 0$ [i.e., negative] there is a reflection in the x-axis too.
$f(x) + a$	Translation by $\binom{0}{a}$
$f(x + a)$	Translation by $\binom{-a}{0}$
$f(ax)$	Horizontal [parallel to x-axis] squash away from y-axis by a factor of a. If $a < 0$ [i.e., negative] there is a reflection in the y-axis too.

Exercise

Sketch $y = |f(x)|$ and $y = f(|x|)$ for $f(x) = \sin x$ and for $y = \cos x$

You can use DESMOS GRAPHING to help you, but try to sketch the graphs without using any graphing packages

Combining these transformations

In this section we will look at some examples where more than one of the transformations listed above is used.

It is important to be very careful when using certain combinations of the transformations above because the order in which you interpret the transformation must be correct.

We will look at this using a test case:

Consider $y = f(x) = \cos x$ and $y = f\left(2x + \frac{\pi}{6}\right) = \cos\left(2x + \frac{\pi}{6}\right)$

If you were asked to sketch $y = \cos x$ and use your sketch to deduce a sketch of $y =$ $= \cos\left(2x + \frac{\pi}{6}\right)$, it would be tempting to suggest it is

1. a "horizontal squash" by a factor of 2 followed by a translation by $\begin{pmatrix} -\frac{\pi}{6} \\ 0 \end{pmatrix}$;
2. or perhaps it is tempting to suggest that it is a translation by $\begin{pmatrix} -\frac{\pi}{6} \\ 0 \end{pmatrix}$ followed by a "horizontal squash" by a factor of 2 .

Before reading on, which of the two suggested transformations is the one you would choose? Or would you propose something else instead? You can use a graph sketching package to explore before we look at the answer.

To answer this, we can look at the transformations suggested in stages:

1. "horizontal squash" by a factor of 2 followed by a translation by $\begin{pmatrix} -\frac{\pi}{6} \\ 0 \end{pmatrix}$:

   $\cos x \ \rightarrow \ \cos 2x \ \rightarrow \cos 2\left(x + \frac{\pi}{6}\right) = \cos\left(2x + \frac{\pi}{3}\right)$ OH NO !!!

2. translation by $\begin{pmatrix} -\frac{\pi}{6} \\ 0 \end{pmatrix}$ followed by a "horizontal squash" by a factor of 2

   $\cos x \ \rightarrow \ \cos\left(x + \frac{\pi}{6}\right) \ \rightarrow \cos\left(2x + \frac{\pi}{6}\right)$

We can see here that 2 gives the correct final answer and 1 gives the incorrect answer. Can you explain why?[77]

We can write the function slightly differently to get another perspective on the transformation:

[77] If we replace x by $2x$ first, then the translation by $\frac{\pi}{6}$ is also affected by the 2 in the $2x$

3. "horizontal squash" by a factor of 2 followed by a translation by $\begin{pmatrix} -\frac{\pi}{12} \\ 0 \end{pmatrix}$:

$$\cos x \;\; \rightarrow \;\; \cos 2x \;\; \rightarrow \cos 2\left(x + \frac{\pi}{12}\right) = \cos\left(2x + \frac{\pi}{6}\right)$$

Exercise

Consider the graph of $y = f(x) = 3^x$

Sketch both $y = f(x + 2)$ and $y = 9f(x)$

What do you notice? Explain your answer.

Now consider $y = f(x) = \log_{10} x$

Sketch both $y = f(10x)$ and $y = 1 + f(x)$

What do you notice? Explain your answer.

The notation $\boldsymbol{f(g(x))}$.

We will look briefly at the notation $f(g(x))$ [which we tend to say as "f of g of x"]

We have been using the ideas connected to this notation already.

Let's take a simple case to illustrate how to unpack his notation:

Let's take $g(x) = 2x$ and $f(x) = x^2 + 3x - 2$

The x in $f(x)$ is just a label to tell you what to do with what you input into the function. That is once you are given an $input$ for $f(x)$, the output is given as $(input)^2 + 3 \times (input) - 2$

So, you can guess what $f(g(x))$ might mean: it means that you take $g(x)$ as the input for $f(x)$. We can write this out:

1. Replace x [which labels the input to $f(x)$] in $f(x)$ by $g(x)$:

$$f(g(x)) = [g(x)]^2 + 3 \times g(x) - 2$$

2. And then replace $g(x)$ by $2x$

$$f(g(x)) = f(2x) = [2x]^2 + 3 \times 2x - 2 = 4x^2 + 6x - 2$$

But it is easier to put $2x$ in immediately and skip step 1.

Exercise

Given

$f(x) = x^2$

$g(x) = x^3 - 3$

$h(x) = \frac{2x+3}{x-4}$

Find simplified expression for

$f(g(x))$

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

$g(h(x))$

$h(g(x))$

What do you notice?

Is it true that $f\big(g(x)\big) = g\big(f(x)\big)$?

MM8.3

Understand how altering the values of m and c affects the graph of $y = mx + c$.

To understand how m and c affects the graph of $y = mx + c$ consider either of the following sequence of transformation. You will need to use your knowledge of graph transformations that we discussed above and make use of a graph sketching package too to enhance your understanding:

$$y = x \rightarrow y = mx \rightarrow y = m\left(x + \frac{c}{m}\right) = mx + c$$

$$y = x \rightarrow y = x + c \rightarrow y = mx + c$$

Exercise

How did you deal with $y = x \rightarrow y = x + c$ above? Did you use $f(x) + c$ or $f(x + c)$?

Pick some pairs of values for m and c and then follow though the graph transformations above step by step using a graph sketching package.

MM8.4

Understand how altering the values of a, b and c in $y = a(x+b)^2 + c$ affects the corresponding graph.

You should be able to work out what each letter does by breaking down the expression into various graph transformations [we tend to assume $a \neq 0$]:

Here is one way of doing this [when $a > 0$]:

$$y = x^2 \;\rightarrow\; y = \left(\sqrt{a}\,x\right)^2 = ax^2 \;\rightarrow\; y = a(x+b)^2 \;\rightarrow\; y = a(x+b)^2 + c$$

And here is another subtly different way

$$y = x^2 \rightarrow y = ax^2 \rightarrow y = a(x+b)^2 \rightarrow y = a(x+b)^2 + c$$

In the first we used $f(\sqrt{a}\,x)$ [horizontal squash when $a > 0$] and in the second we used $af(x)$ [vertical stretch] Check that you get the same graph in both cases by trying some simple numbers and using a graph sketching package for a, eg $a = 4$.

As an aside: note that sometimes we can only use one method, e.g. when $a = -4$ we cannot use $\sqrt{-4}$. But we could introduce an extra step:

$$y = x^2 \rightarrow y = \left(\sqrt{|a|}\,x\right)^2 = |a|x^2 \rightarrow y = -|a|x^2 \rightarrow y = a(x+b)^2 \rightarrow y = a(x+b)^2 + c$$

But this is a bit cumbersome.

Exercise

Pick some sets of values for a, b and c and then follow though the graph transformations above step by step using a graph sketching package.

The last three sections of the specification have been covered in earlier sections:

MM8.5

Use differentiation to help determine the shape of the graph of a given function; including finding stationary points (excluding inflexions); and when the function is increasing or decreasing.

MM8.6

Use algebraic techniques to determine where the graph of a function intersects the coordinate axes; appreciate the possible numbers of real roots a general polynomial can possess.

MM8.7

Geometric interpretation of algebraic solutions of equations; relationship between the intersections of two graphs and the solutions of the corresponding simultaneous equations.