Calculus

IB Course Notes

Joshua Corneck

Contents

1 Introduction to Differentiation

As we said above, differentiation (differential calculus) is all concerned with instantaneous rates of change. If we have some function $f(x)$, differentiation will tell us how the output of our function f is changing with respect to our input variable x.

This is actually a very physically important concept. Acceleration is simply how fast your speed is changing, i.e. it is the *rate of change* of speed with respect to time. It follows that we would say that the derivative of speed is acceleration. Similarly, your speed is just how much distance you are covering in a given time interval, and so speed is the derivative of distance.

We are now going to explicitly derive, from first principles, the definition of a derivative.

1.1 Differentiation from First Principles

We are going to derive an expression for the derivative of a function $f(x)$ with respect to its independent variable input x. Please bear in mind that there is nothing special about x: we could just as well have used t or y or any other independent variable.

Back in GCSE maths, we learned how to find the *gradient* of a straight line. This mysterious gradient gave us a value for how the y-value of our function (the output) was changing with respect to the x-value (the input). Let's consider the straight line $y = 5x + 2$. In the table below we put some of the y-values for different x inputs:

x	1	2	3	4	5
y	7	12	17	22	27

We find our gradient by taking two separate points on the line and taking the ratio of the difference in the y values to the difference of the x values at the two points; "rise over run". So, for this line, the gradient m is

$$m = \frac{22 - 17}{4 - 3} = 5$$

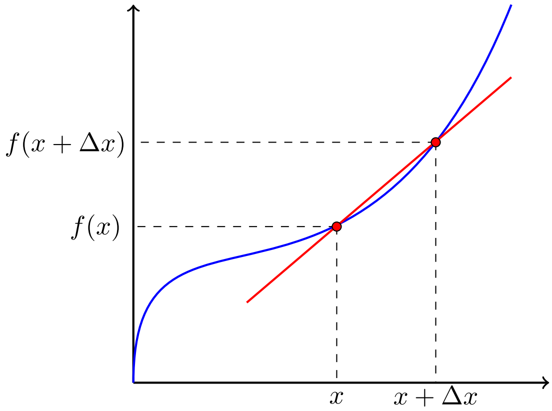
What this means is that when we change our x value by 1 unit, our y value will change by 5. So, if we went from $x = 1$ to $x = 2$ we will have that y goes from 7 to 12 - it increases by *5 times* the amount that x did. For a straight line, the gradient is very easy as it's the same at every point along the line, i.e. no matter where you are on the line, a change in 1 unit in x will result in a change in 5 units in y.

But what happens when we want to find the gradient of a function that isn't straight, say, for example, we wanted to find the gradient of $y = x^2$? Look at the values of this function for x from 1 to 5:

x	1	2	3	4	5
y	1	4	9	16	25

When we change x by 1 unit from 2 to 3, say, we see y increases by 5 from 4 to 9. But if we move by 1 unit from 4 to 5, y increases from 16 to 25 - an increase of 9! So, for non-linear functions, we encounter the problem that the gradient (the rate of change) varies from point to point. This begs the question of how best to define the gradient for a non-linear function such as $y = x^2$?

The answer is somewhat less complicated than might be expected. What we do is, we define the gradient at a point to be the gradient of the *tangent* to the function at that point. A tangent to a graph at a point is a line that "just touches" the function at that point. Another way to think about it is as a straight line through two points *infinitesimally* close together (really, really close together) along the curve of the function.



We now use this definition to calculate the gradient of a function $f(x)$ at a point. Consider the curve of the function $y = f(x)$ and two points along the curve that are very close together: $y = f(x)$ and $y = f(x + \Delta x)$, where Δx is very, very small (i.e. $|\Delta x| \ll 1$). Think about the straight line connecting these two points. What would be it's gradient? Well, gradient is simply rise over run, so we would have

$$m = \frac{\text{“change in } y\text{”}}{\text{“change in } x\text{”}} = \frac{f(x+\Delta x) - f(x)}{(x+\Delta x) - x}$$

Remember above when we said the tangent at a point could be thought of as the straight line through two points infinitesimally close together? Well, now we want to imagine bringing these two points (x and $x + \Delta x$) really, really, *really* close together, so that what we are left with in our expression for m is the gradient of the *tangent* to f at the point x. To do this, we need to let Δx shrink down to zero, and then we would have the gradient of a line through two points infinitesimally close together - i.e. the gradient of the tangent. We denote this 'shrinkage' using the notation $\lim_{\Delta x \to 0}$:

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

This is *exactly* the definition of the *derivative* of the function $f(x)$ at the point x: *the gradient of the tangent to the function at that point.* We denote the derivative of a function $f(x)$ with respect to x by:

$$\frac{\mathrm{d}f}{\mathrm{d}x}$$

pronounced "dee f by dee x". Thus, we have arrived at the *very* important definition of the derivative:

$$\frac{\mathrm{d}f}{\mathrm{d}x} := \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

This $\frac{\mathrm{d}}{\mathrm{d}x}$ operator is something we need to understand. As an operator, it means to take the derivative with respect to x. So, for example, if we wrote $\frac{\mathrm{d}y}{\mathrm{d}x}$, this would mean to take the derivative of y with respect to x. Another piece of notation we will use interchangeably is to write

$$g'(x) \quad \text{to mean} \quad \frac{\mathrm{d}g(x)}{\mathrm{d}x}$$

When we learn about higher-order derivatives, we will use the notation

$$g^{(n)}(x) \quad \text{to mean} \quad \frac{\mathrm{d}^n g(x)}{\mathrm{d}x^n}$$

but don't worry about what that means for the time being.

1.2 Linearity of the derivative operator

We call $\frac{\mathrm{d}}{\mathrm{d}x}$ the derivative *operator*. A very useful property of this operator is that it is *linear* in the following sense:

$$\frac{\mathrm{d}}{\mathrm{d}x}(af(x) + bg(x)) = a\frac{\mathrm{d}f}{\mathrm{d}x} + b\frac{\mathrm{d}g}{\mathrm{d}x}$$

for any constants a, b and differentiable functions f, g. To see this, we simply use the definition:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}(af(x) + bg(x)) &= \lim_{h \to 0} \frac{(af(x+h) + bg(x+h)) - (af(x) + bg(x))}{h} \\
&= \lim_{h \to 0} \frac{a(f(x+h) - f(x)) + b(g(x+h) - g(x))}{h} \\
&= \lim_{h \to 0} \frac{a(f(x+h) - f(x))}{h} + \lim_{h \to 0} \frac{b(g(x+h) - g(x))}{h} \\
&= a \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} + b \lim_{h \to 0} \frac{g(x+h) - g(x)}{h} \\
&= a\frac{\mathrm{d}f}{\mathrm{d}x} + b\frac{\mathrm{d}g}{\mathrm{d}x}
\end{aligned}$$

1.3 Derivatives of Polynomials

An nth degree polynomial is something of the form

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

where $a_0, a_1, \cdots, a_{n-1}$ are arbitrary constants and a_n is a non-zero constant. This looks a little abstract, but all it means is we are looking at something like $3x^2 + 6x - 8$ or $x^4 + x + 1$, or even $x^{1001} - 3$.

Differentiating polynomials turns out to be quite easy. If we plug the expression $f(x) = x^n$ (i.e. we let $f(x)$ be a polynomial of degree n), it can be shown that:

$$\frac{\mathrm{d}}{\mathrm{d}x}x^n = \lim_{h\to 0}\frac{(x+h)^n - x^n}{h} = nx^{n-1}$$

which we have a proof of in the appendices. Let's see how this works in action:

Example 1. $\frac{\mathrm{d}}{\mathrm{d}x}\left(x^4\right)$

Here $n = 4$, so we plug in to find

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(x^4\right) = 4x^3$$

Example 2. $\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{6}x^3\right)$

So $n = 3$ and we plug in:

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{6}x^3\right) = \frac{1}{6} \times 3x^2 = \frac{1}{2}x^2$$

where we use the linearity of the derivative operator to "differentiate past" the constant.

Example 3. $\frac{\mathrm{d}}{\mathrm{d}x}\left(x^3 + 4x^2\right)$

We know differentiation is *linear* (we can do it term-by-term), and so

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(x^3 + 4x^2\right) = \frac{\mathrm{d}}{\mathrm{d}x}\left(x^3\right) + \frac{\mathrm{d}}{\mathrm{d}x}\left(4x^2\right) = 3x^2 + 8x$$

Example 4. $\frac{\mathrm{d}}{\mathrm{d}x}(x)$

We have $n = 1$ and so find

$$\frac{\mathrm{d}}{\mathrm{d}x}(x) = 1$$

Example 5. $\frac{\mathrm{d}}{\mathrm{d}x}(c)$

where c is a constant. We can write $c = cx^0$ (i.e. $n = 0$) and so

$$\frac{\mathrm{d}}{\mathrm{d}x}(c) = \frac{\mathrm{d}}{\mathrm{d}x}\left(cx^0\right) = 0 \times cx^{-1} = 0$$

so the derivative of a constant is 0.

Let's think about that last result. We know that the derivative of a function gives an expression for the gradient. What would a graph of $y = c$ look like? Well, it would just be a horizontal line because $y = c$ everywhere! The gradient of a horizontal line is just 0 because

$$m = \frac{\Delta y}{\Delta x} = \frac{c - c}{x_1 - x_0} = 0$$

and so it makes sense that the derivative of a constant is 0, because the gradient of it's graph is 0!

1.4 Derivatives of Some Standard Functions

Below we list some common derivatives that we need to know:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}(e^x) &= e^x \\ \frac{\mathrm{d}}{\mathrm{d}x}(\log(x)) &= \frac{1}{x} \\ \frac{\mathrm{d}}{\mathrm{d}x}(\sin(x)) &= \cos(x) \\ \frac{\mathrm{d}}{\mathrm{d}x}(\cos(x)) &= -\sin(x) \\ \frac{\mathrm{d}}{\mathrm{d}x}(\tan(x)) &= \sec^2(x) \\ \frac{\mathrm{d}}{\mathrm{d}x}(a^x) &= a^x \ln(a)\end{aligned}$$

If you want to see how these results are derived, please see the appendices.

There are a few nice things to notice here. Firstly, we see that our formula for $\frac{\mathrm{d}}{\mathrm{d}x}(a^x)$ is equivalent to that for the derivative of e^x in the case $a = e$ as $\ln(e) = 1$ and so we arrive at the expected result. Secondly, we notice that the derivative of e^x is just e^x! This is a remarkable result unique to this function.

Another nice property you may have spotted is the cyclical nature of our sine and cosine derivatives. What we mean by this is as follows:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}(\sin(x)) &= \cos(x) \\ \frac{\mathrm{d}}{\mathrm{d}x}(\cos(x)) &= -\sin(x) \\ \frac{\mathrm{d}}{\mathrm{d}x}(-\sin(x)) &= -\frac{\mathrm{d}}{\mathrm{d}x}(\sin(x)) = -\cos(x) \\ \frac{\mathrm{d}}{\mathrm{d}x}(-\cos(x)) &= -\frac{\mathrm{d}}{\mathrm{d}x}(\cos(x)) = \sin(x)\end{aligned}$$

So, we find that taking the derivative of sine or cosine 4 times, will return us to our starting function!

2 Differentiation Theorems

2.1 Chain Rule

Suppose that you're a free diver (for some unknown reason) and you are descending down toward the bottom of the ocean. For the sake of simplicity, let's say that you are descending at a rate of 1 metre every second. As you descend, the temperature around you drops. In fact, it happens to be the case that the temperature of the water around you drops at a rate of 1 degree for every 10 metres you descend.

Given this information, we know how distance descended varies with time, and we know how temperature varies with distance descended, but what if we want to know how the temperature will change with time? This might be a pertinent piece of information if we want to estimate how long you will be able to last under the water.

We can work this out as follows: if we drop 1 metre in 1 second, and for every 10 metres we drop we have a fall of 1 degree, it follows that we will have a temperature drop of 1 degree for every 10 seconds we are underwater.

Let's formalise this slightly. Suppose we let $H(t)$ denote the function denoting the distance descended with time, and $T(H)$ the function denoting temperature variation with distance descended. In this case, we have just shown that the rate of change of $T(t)$ (temperature variation with time) is given by

$$T'(t) = H'(t)T'(H)$$

i.e. we have simply multiplied the two rates of change. It actually turns out this result holds in general, and is called the *chain rule*.

The chain rule gives us a method for finding the derivative of a function that is in fact the composite of other functions. It says that if we have a composite function $h(x)$ such that

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

then

$$h'(x) = g'(x)f'(g(x))$$

This is a *very* useful result and we will be using it repeatedly. It comes in handy when we want to find the derivative of a function that is almost a standard function, but not quite.

Let's take a look at a couple of examples:

Example 6. *Find $\frac{\mathrm{d}h}{\mathrm{d}x}$ in the case that*

$$h(x) = \sin x^2$$

We know how to differentiate $\sin x$, but this is a little different. We compare this to the composition above and see that $f(x) = \sin x$ and $g(x) = x^2$. We should now know that

$$f'(x) = \cos x, \qquad g'(x) = 2x$$

and so we can simply plug this into the formula above and find

$$h'(x) = g'(x)f'(g(x)) = 2x\cos x^2$$

Example 7. *Find $\frac{\mathrm{d}h}{\mathrm{d}x}$ in the case that*

$$h(x) = e^{kx}$$

for some constant function k.

For this example, simply note that $h(x) = f(g(x))$ with $f(x) = e^x$ and $g(x) = kx$. We know

$$f'(x) = e^x, \qquad g(x) = k$$

and so we plug in to get

$$h'(x) = g'(x)f'(g(x)) = ke^{kx}$$

Example 8. *Calculate*

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{1-x^2}\right)$$

In this example, things look a little different. If we stare for long enough, we will find

$$f(x) = x^{-1}$$

and

$$g(x) = 1 - x^2$$

which yields

$$f(g(x)) = \frac{1}{1 - x^2}$$

We can then just use our knowledge of polynomial derivatives to find

$$f'(x) = -x^{-2}, \qquad g'(x) = -2x$$

All we need to do now is plug into the formula:

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{1 - x^2}\right) = g'(x)f'(g(x)) = (-2x)\left(-(1 - x^2)^{-2}\right) = \frac{2x}{(1 - x^2)^2}$$

Example 9. *Calculate*

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{(1 - \cos t)^2}\right)$$

This example may look more intimidating than the others, but it is in fact no more complicated. In this case, we have

$$f(t) = t^{-2}, \qquad g(t) = 1 - \cos t$$

We can then find

$$f'(t) = -2t^{-3}, \qquad g'(t) = \sin t$$

Putting it all together, we find

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{(1 - \cos t)^2}\right) = \frac{-2\sin t}{(1 - \cos t)^3}$$

There are an infinite number of examples we can look at. As with all maths, the best bet is just to keep practising with different combinations of functions until the technique becomes second nature.

It is important to become familiar with writing the chain rule as

$$\frac{\mathrm{d}h}{\mathrm{d}x} = \frac{\mathrm{d}h}{\mathrm{d}y}\frac{\mathrm{d}y}{\mathrm{d}x}$$

for the derivative of a function $h = h(y)$ where $y = y(x)$. This is exactly what we wrote above, just with $y = g(x)$.

As an addition, let's see what happens when we have an n-fold composition, i.e. when

$$h(x) = f_1(f_2(\cdots f_n(x)\cdots))$$

We think of this as $f_1(g_1(x))$ where $g_1(x) = f_2(\cdots f_n(x)\cdots)$. We see that our chain rule gives

$$h'(x) = g_1'(x)f_1'(g_1(x)) = [f_2(\cdots f_n(x)\cdots)]'\, f_1'(f_2(\cdots f_n(x)\cdots))$$

But what is $[f_2(\cdots f_n(x)\cdots)]'$? Well, we think about this again as $f_2(g_2(x))$ where $g_2(x) = f_3(f_4(\cdots f_n(x)\cdots))$. Plugging into our chain rule gives

$$[f_2(\cdots f_n(x)\cdots)]' = g_2'(x)f_2'(g_2(x)) = [f_3(f_4(\cdots f_n(x)\cdots))]'\, f_2'(f_3(f_4(\cdots f_n(x)\cdots))$$

i.e.

$$h'(x) = [f_3(f_4(\cdots f_n(x)\cdots))]'\, f_2'(f_3(f_4(\cdots f_n(x)\cdots))f_1'(f_2(\cdots f_n(x)\cdots))$$

We can continue this procedure iteratively to find

$$h'(x) = f_n'(x)f_{n-1}'(f_n(x))\cdots f_2'(f_3(f_4(\cdots f_n(x)\cdots))f_1'(f_2(\cdots f_n(x)\cdots))$$

Let's have a look at an example of this

Example 10. *Calculate*

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[\sin\left(\cos\left(x^2\right)\right)\right]$$

As a start, let's consider this as a single composition of functions. We will write

$$\sin\left(\cos\left(x^2\right)\right) = f_1(g(x))$$

where

$$f_1(x) = \sin x, \qquad g(x) = \cos\left(x^2\right)$$

In this way, we find

$$\frac{\mathrm{d}}{\mathrm{d}x}[\sin\left(\cos\left(x^2\right)\right)] = g'(x)f_1'(g(x)) = \frac{\mathrm{d}}{\mathrm{d}x}[\cos\left(x^2\right)]\cos\left(\cos\left(x^2\right)\right)$$

Now we need to find the derivative of $g(x)$. To do this, we consider $g(x)$ as $f_2(f_3(x))$, where

$$f_2(x) = \cos x, \qquad f_3(x) = x^2$$

Then we have

$$g'(x) = f_3'(x)f_2'(f_3(x)) = 2x(-\sin\left(x^2\right)) = -2x\sin\left(x^2\right)$$

All we have to do now is combine it all to find

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[\sin\left(\cos\left(x^2\right)\right)\right] = -2x\sin\left(x^2\right)\cos\left(\cos\left(x^2\right)\right)$$

2.2 Product Rule

The product rule gives us a technique for evaluating the derivative of a function that is expressed as the product of other functions. Take the case that we have a product of just two functions, and so let $h(x) = f(x)g(x)$. The derivative of $h(x)$ is then:

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

This is best illustrated by an example.

Example 11.

$$h(x) = x^2 \sin x$$

We see that $h(x) = f(x)g(x)$ where $f(x) = x^2$ and $g(x) = \sin x$. We know the derivatives of both these functions:

$$f'(x) = 2x, \qquad g'(x) = \cos x$$

We then simply plug into our above formula to get:

$$h'(x) = f'(x)g(x) + f(x)g'(x) = 2x\sin x + x^2\cos x = x(2\sin x + x\cos x)$$

Example 12.

$$h(x) = e^x \log x$$

We see $f(x) = e^x$ and $g(x) = \log x$, and so

$$f'(x) = e^x, \qquad g'(x) = \frac{1}{x}$$

Plugging into the formula above:

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

It is handy to note that the product rule can be used for any arbitrary product of functions - it doesn't have to be just two of them. For an n-term product:

$$h(x) = f_1(x)f_2(x)\cdots f_n(x)$$

we have

$$h'(x) = f_1'(x)f_2(x)\cdots f_n(x) + f_1(x)f_2'(x)\cdots f_n(x) + \cdots + f_1(x)f_2(x)\cdots f_n'(x)$$

So, in the case of three functions, we would have

$$h(x) = f_1'(x)f_2(x)f_3(x) + f_1(x)f_2'(x)f_3(x) + f_1(x)f_2(x)f_3(x)$$

and we apply the formula in exactly the same way as we did for a product of three functions.

2.3 Quotient Rule

The quotient rule gives us a way to find the derivative of a function that is the ratio of two other functions. In other words, of a function $h(x)$ such that

$$h(x) = \frac{f(x)}{g(x)}$$

In this case, we can show

$$h'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{[g(x)]^2}$$

and find that this is valid for all x such that $g(x) \neq 0$. This result isn't actually anything new, it's just an application of the product rule. Suppose we rewrite $h(x)$ as

$$h(x) = \frac{f(x)}{g(x)} = f(x)[g(x)]^{-1}$$

Then we can plug this into the product rule to find

$$h'(x) = f'(x)[g(x)]^{-1} + f(x)\left([g(x)]^{-1}\right)'$$

The chain rule says that if we think of $[g(x)]^{-1}$ as $k(g(x))$ where $k(x) = x^{-1}$, then

$$\left([g(x)]^{-1}\right)' = [k(g(x))]' = g'(x)k'(g(x)) = -g'(x)\frac{1}{[g(x)]^2}$$

because $k'(x) = -1/x^2$. Substituting this into the above, we find

$$h'(x) = f'(x)[g(x)]^{-1} - f(x)g'(x)[g(x)]^{-2}$$

which can be rewritten as

$$h'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{[g(x)]^2}$$

I don't like this result very much, and would advise you to use the product rule instead as illustrated above, but it's your choice. As an example, consider

Example 13.

$$h(x) = \frac{\sin x}{\cos x}$$

We have

$$f'(x) = \cos x, \qquad g'(x) = -\sin x$$

and so

$$h'(x) = \frac{\cos x \cos x - \sin x(-\sin x)}{\cos^2 x)} = \frac{\sin^2 x + \cos^2 x}{\cos^2 x} = \frac{1}{\cos^2 x}$$

But notice that $h(x)$ is actually just $\tan x$, and so we have found

$$\frac{\mathrm{d}}{\mathrm{d}x}(\tan x) = \frac{1}{\cos^2 x}$$

3 Stationary Points and Higher Derivatives

3.1 Higher derivatives

So far, we have considered taking derivatives once and seeing what happens. But why must we only take them once? It so happens that we can take derivatives multiple times and use these to deduce different results about a function.

The higher-order derivative that we will be concerning ourselves with is the second-derivative. The second-derivative of a function $f(x)$ is simply the *derivative of its first-derivative*, and we denote it by $f''(x)$ or $\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}$. Explicitly, we have

$$f''(x) = (f'(x))', \qquad \frac{\mathrm{d}^2 f}{\mathrm{d}x^2} = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\mathrm{d}f}{\mathrm{d}x}\right)$$

Example 14. *Calculate $f''(x)$ if $f(x) = x^3 + 2x^2 + 2$*

This is a polynomial function and so is easy to differentiate. We simply differentiate it twice:

$$f''(x) = (f'(x))' = (3x^2 + 4x)' = 6x + 4$$

Example 15. *Calculate*

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}\left(\sqrt{1-\cos x}\right)$$

As a starting point, let's find the first derivative. We need to use the chain rule here, where we say

$$\sqrt{1-\cos x} = f(g(x))$$

with

$$f(x) = \sqrt{x}, \qquad g(x) = 1 - \cos x$$

Then

$$f'(x) = \frac{1}{2\sqrt{x}}, \qquad g'(x) = \sin x$$

and so

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\sqrt{1-\cos x}\right) = \frac{\sin x}{2\sqrt{1-\cos x}}$$

To find the second derivative, we need simply take the derivative of this! This is a more complicated calculation and will need both the quotient and the chain rule. Firstly, we write

$$\frac{\sin x}{2\sqrt{1-\cos x}} = \frac{u(x)}{v(x)}$$

with u and v defined as expected. Then we see

$$u'(x) = \cos x, \qquad v'(x) = \frac{\sin x}{\sqrt{1-\cos x}}$$

The quotient rule then tells us

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{u(x)}{v(x)}\right) = \frac{u'v - v'u}{v^2} = \frac{2\cos x\sqrt{1-\cos x} - \frac{\sin^2 x}{\sqrt{1-\cos x}}}{4(1-\cos x)} = \frac{2\cos x - 2\cos^2 x - \sin^2 x}{4(1-\cos x)^{\frac{3}{2}}} = \frac{-(1-\cos x)^2}{4(1-\cos x)^{\frac{3}{2}}}$$

thus we find

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}\left(\sqrt{1-\cos x}\right) = -\frac{1}{4}\sqrt{1-\cos x}$$

Before we discuss the relevance of second derivatives, we need to first introduce a few other concepts.

3.2 Stationary points

Informally, a stationary point of a one-variable function can be thought of as a point where the function stops increasing or decreasing (and is thus "stationary").

By now, we are familiar with the idea that a function's derivative corresponds to the *rate of change* of the function at a particular point. So, for some continuously differentiable function $f(x)$, if $f'(x) > 0$ then $f(x)$ is increasing, and if $f'(x) < 0$ then it is decreasing.

If we have said that $f'(x) > 0$ corresponds to the function increasing, and that $f'(x) < 0$ to the function decreasing, then this begs the question of what it means for $f'(x) = 0$. It is precisely this question that we will be concerning ourselves with.

The intuitive answer would be that the function is neither increasing nor decreasing, and is thus stationary. In fact, in this case, this intuitive reasoning is correct:

For a function, $f(x)$, if a point $x = x_0$ is such that

$$f'(x_0) = 0$$

then $x = x_0$ is called a *stationary point* of $f(x)$.

One of three things can happen at a stationary point (for a one-variable function):

1. We have a local (or global) maximum
2. We have a local (or global) minimum
3. We have a horizontal point of inflexion

and we will discuss each of these in turn.

3.3 Maxima and minima

A *maximum/minimum* of a function is the largest/smallest value of the function, either within a given range, or over the whole domain.

If we ask for the largest/smallest value within a given range, then we call the value a *local maximum/minimum*.

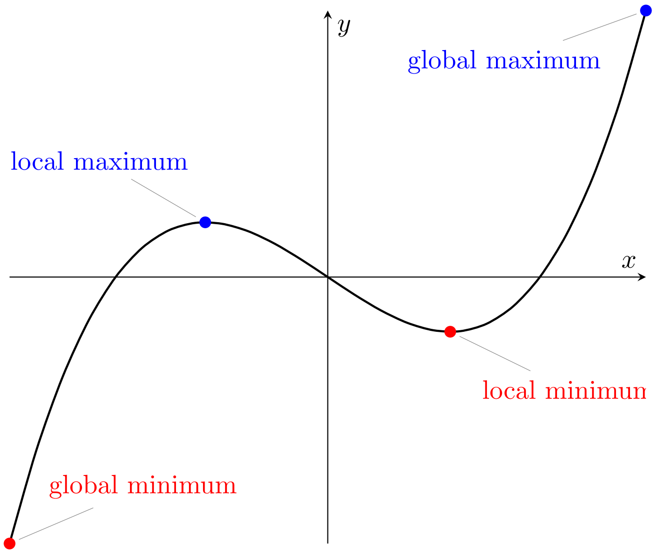
If we ask for the largest/smallest value over the entire domain, then we call this value a *global maximum/minimum*.

For example, take the cubic function below specified over a finite domain. We see here that we have two local *extrema* (one local maximum and one local minimum) and two global extrema (one global maximum and one global minimum). If we were to extend our specified domain, the global extrema would change, but the local extrema would not (unless they were removed from the domain, that is).
For now, we won't concern ourselves with global extrema, but only with local minima and maxima. What does it mean for a function to reach a local maximum/minimum. Well, by definition, it means that the function is at its largest/smallest value.

By this point we are familiar with the derivative of a function corresponding to its rate of change: if the derivative takes a positive value then the function is increasing, and if the derivative is negative, then the function is decreasing. Let's try and relate this to local extrema.

If we move toward a local maximum, by definition we must increase up to it and decrease away from it. This only makes sense, as if we don't (strictly) decrease once we pass it, it can't be the maximum point. So we see that a condition for a local maximum is that we must have a positive derivative to the left of the point (we increase up to it), and a negative gradient to the right of it (we decrease away from it). This gives us the "hill-like" shape that we see in our graph.

In a similar vein, we deduce that at a local minimum, we must have a negative gradient leading up to the point, and a positive gradient moving away from it (giving a "valley").



OK, so we now know what happens either side of the local extrema, but what happens *at* the local extrema. The answer to this is what you might guess: if we move smoothly from positive to negative (or negative to positive), there must be a point at which we pass through 0. This point at which the derivative is zero corresponds to our local extrema:

At a local minimum/maximum of a function $f(x)$, we have that $f'(x) = 0$.

Example 16. *Find the coordinates of any stationary points of the function*

$$y = x^3 - 3x + 1$$

We start by finding the first derivative:

$$\frac{\mathrm{d}y}{\mathrm{d}x} = 3x^2 - 3$$

Now, we know that at a stationary point, the first derivative is 0, so it seems sensible to set $\frac{\mathrm{d}y}{\mathrm{d}x} = 0$ and to solve:

$$\frac{\mathrm{d}y}{\mathrm{d}x} = 0 \quad \Rightarrow \quad 3x^2 - 3 = 0 \quad \Rightarrow \quad x = \pm 1$$

and so we see that when $x = \pm 1$, we have a stationary point. To find the y-coordinate, we simply substitute x back into the original function to obtain:

$$x = 1, \quad y = 1^3 - 3(1) + 1 = -1 \quad \text{and} \quad x = -1, \quad y = (-1)^3 - 3(-1) + 1 = 3$$

So, we now have the coordinates of the stationary points, but how do we classify them? This is where second derivatives come into their own.

In the same way that a first derivative corresponds to the rate of change of a function, the second derivative corresponds to the rate of change of the gradient (i.e. of the first derivative). This is very helpful for us. Consider what happens as we approach a local maximum: It is important to note that if we have a local maximum or minimum, then we have a zero gradient, but if we have a zero gradient then we don't necessarily have a maximum or minimum (it may be our third option mentioned above).

So, we know that at a maximum or minimum, the derivative must take the value 0. But, if we deduce that the derivative is 0 at some point (and we know that it's a maximum or a minimum), how do we determine whether this point is a local maximum or a local minimum.

To answer this, we will consider two methods.

3.4 Categorising stationary points

3.4.1 First derivative test

The first way we can check the type of stationary point is by using the description of a maximum/minimum given above.

We know the behaviour of a function leading up to and beyond a maximum/minimum: namely that it increases/decreases up to it and decreases/increases beyond it. In this way, in order to classify a stationary point, all we need to do is check the value of the gradient at a point to the left and right of the stationary point. Depending on how the sign of the derivative (the sign of the gradient) changes as we move from the left to the right of the stationary point, we will be able to classify the point as a maximum/minimum:

If $x = x_0$ is maximum of the function $f(x)$, we have that $f'(x_0 - \epsilon) > 0$ and $f'(x_0 + \epsilon) < 0$ for some small, positive $\epsilon > 0$.

If $x = x_0$ is minimum of the function $f(x)$, we have that $f'(x_0 - \epsilon) < 0$ and $f'(x_0 + \epsilon) > 0$ for some small, positive $\epsilon > 0$.

This makes intuitive sense, based on the description of maxima and minima given above.

Example 17. *Classify the stationary points of the function*

$$y = x^3 - 3x + 1$$

We know that the coordinates of the stationary points are $(1, -1)$ and $(-1, 3)$. Let's start with $(1, -1)$:

We will consider the point $x = 0$ and $x = 2$ as points to the left and to the right. We find

$$\left.\frac{\mathrm{d}y}{\mathrm{d}x}\right|_{x=0} = 3(0)^2 - 3 < 0$$

and

$$\left.\frac{\mathrm{d}y}{\mathrm{d}x}\right|_{x=2} = 3(2)^2 - 3 > 0$$

and so the gradient changes from negative to positive, and so $(1, -1)$ is a minimum.

Similar analysis of the points $x = -2$ and $x = 0$ for $(-1, 3)$ finds

$$\left.\frac{\mathrm{d}y}{\mathrm{d}x}\right|_{x=-2} = 3(-2)^2 - 3 > 0$$

and

$$\left.\frac{\mathrm{d}y}{\mathrm{d}x}\right|_{x=0} = 3(0)^2 - 3 < 0$$

and so the gradient changes from positive to negative, and $(-1, 3)$ is seen to be a maximum.

3.4.2 Second derivative test

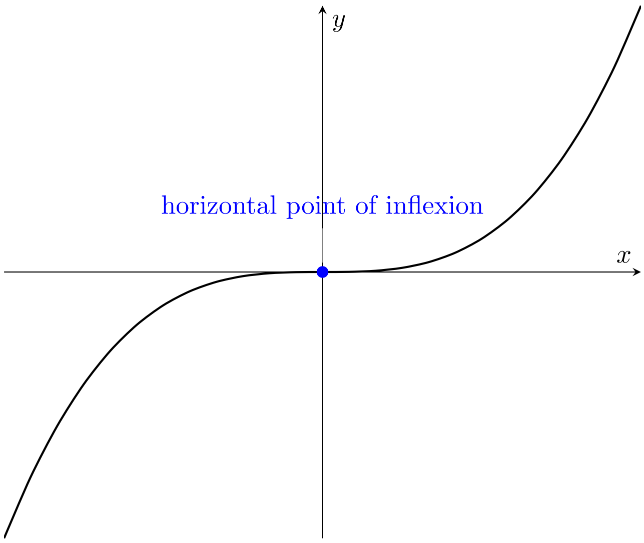
The second way we can classify stationary points is by checking the value of the second derivative at that point.

In the same way that a first derivative corresponds to the *rate of change* of the original function with respect to some variable, the second derivative corresponds to the rate of change of the *derivative* with respect to that variable. Why might this be useful? Well, we know that leading up to a maximum/minimum

3.5 Horizontal point of inflexion

Technically, a point of inflexion is a point at which a function changes from *concave* to *convex* (or vice versa). We won't worry about the definitions of convex and concave, but instead will just consider what happens in terms of a function's shape.

Suppose we have a stationary point for which the derivative is positive (or negative) to the left *and* is positive (or negative) to the right of the point. This will satisfy the condition that our derivative is 0, but clearly the point won't be a maximum or a minimum. If this is the case, we have a *horizontal point of inflexion.* To see this graphically, consider the graph of $y = x^3$:



4 Implicit Differentiation

So far we have learned to deal with functions where y is given *explicitly* as a function of the independent variable x. For example, we would have something like

$$y = (x+2)^2 - \sin(e^x)$$

and find

$$\frac{\mathrm{d}y}{\mathrm{d}x} = 2(x+2) - e^x\cos(e^x)$$

We now want to consider cases when y isn't defined so explicitly. Take, for example, a circle centred at the origin of radius 1:

$$x^2 + y^2 = 1$$

One way we could approach this is to rearrange and find

$$y = \pm\sqrt{1-x^2}$$

and then

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \mp\frac{2x}{\sqrt{1-x^2}}$$

However, it is not always possible to rearrange our expression to find an explicit expression for y. Take, for example, the expression

$$y^3 - x^3 + y - x^2 = 0$$

where $y = y(x)$. In this case, we are unable to rearrange to explicitly obtain $y = f(x)$, and so, as far as we concerned at this point, we are unable to find the derivative of this function. But fear not, for there is something that we can do.

4.1 The chain rule

Consider trying to compute the following derivative:

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(y^2\right)$$

where $y = y(x)$. To evaluate this, we need to recall the *chain rule*. This very helpful rule tells us that if we wish to take a derivative of a function $f(g(x))$ with respect to x, then we simply need to find the derivative of f with respect to g and multiply it by the derivative of g with respect to x. More explicitly,

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(f(g(x))\right) = \frac{\mathrm{d}f}{\mathrm{d}g}\frac{\mathrm{d}g}{\mathrm{d}x}$$

So, let's apply this rule here. Think about y^2 as being $f(y)$, where $f(t) = t^2$. In this way, we can say

$$y^2 = f(y(x))$$

and have something that looks very suited to an application of the chain rule. In this way:

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(y^2\right) = \frac{\mathrm{d}}{\mathrm{d}x}(f(y(x))) = \frac{\mathrm{d}f}{\mathrm{d}y}\frac{\mathrm{d}y}{\mathrm{d}x}$$

But since $f(y) = y^2$, we clearly see $\frac{\mathrm{d}f}{\mathrm{d}y} = 2y$, and so we find

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(y^2\right) = 2y\frac{\mathrm{d}y}{\mathrm{d}x}$$

OK, that wasn't so bad. How about finding

$$\frac{\mathrm{d}}{\mathrm{d}x}(\sqrt{y})$$

Well, we can define $g(t) = \sqrt{t}$, and then we have

$$\sqrt{y} = g(y(x))$$

Then, by the same reasoning as before, we have

$$\frac{\mathrm{d}}{\mathrm{d}x}(\sqrt{y}) = \frac{\mathrm{d}}{\mathrm{d}x}(g(y(x))) = \frac{\mathrm{d}g}{\mathrm{d}y}\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{1}{2\sqrt{y}}\frac{\mathrm{d}y}{\mathrm{d}x}$$

Now, this all seems a little complicated and technical, but if we look closely we see that there is a pattern emerging. Notice how in both the examples above, our results were what we would have obtained if we have differentiated $f(x)$ or $g(x)$ (instead of $f(y)$ or $g(y)$) with respect to x but simply with a $\frac{\mathrm{d}y}{\mathrm{d}x}$ term coming along for the ride. This gives a simple way to calculate derivatives of this form: compute the $\frac{\mathrm{d}}{\mathrm{d}x}$ derivative by "pretending" y is x and then simply add on a $\frac{\mathrm{d}y}{\mathrm{d}x}$ term.

As an example, how could we evaluate

$$\frac{\mathrm{d}}{\mathrm{d}x}(\cos(y))$$

Well, all we need to do is "imagine" that we are computing $\frac{\mathrm{d}}{\mathrm{d}x}(\cos(x))$ (which we know is $-\sin(x)$) and then multiply by $\frac{\mathrm{d}y}{\mathrm{d}x}$:

$$\frac{\mathrm{d}}{\mathrm{d}x}(\cos(y)) = -\sin(y)\frac{\mathrm{d}y}{\mathrm{d}x}$$

Of course, we could be more thorough and write

$$\frac{\mathrm{d}}{\mathrm{d}x}(\cos(y)) = \frac{\mathrm{d}}{\mathrm{d}y}(\cos(y))\frac{\mathrm{d}y}{\mathrm{d}x} = -\sin(y)\frac{\mathrm{d}y}{\mathrm{d}x}$$

but in reality the middle step is unnecessary.

4.2 Some examples

Let's take a look at a complicated example, which we will evaluate in depth

Example 18. *Evaluate the gradient of the tangent to the curve*

$$\sin\left(y^2\right) + e^y - x^2 + x = 0$$

at the point with coordinates $\left(\frac{1+\sqrt{5}}{2}, 0\right)$.

First, we act on both sides with $\frac{\mathrm{d}}{\mathrm{d}x}$:

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\sin\left(y^2\right)\right) + \frac{\mathrm{d}}{\mathrm{d}x}(e^y) - \frac{\mathrm{d}}{\mathrm{d}x}\left(x^2\right) + \frac{\mathrm{d}}{\mathrm{d}x}(x) = 0$$

The terms only involve x are easy to evaluate, and we can rearrange to get

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\sin\left(y^2\right)\right) + \frac{\mathrm{d}}{\mathrm{d}x}(e^y) = 2x - 1$$

For the other terms, we are going to use the method discussed above. For the first derivative, let's think about finding

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\sin\left(x^2\right)\right)$$

This requires the chain rule. We let $u = x^2$ and then $\frac{\mathrm{d}u}{\mathrm{d}x} = 2x$, and so

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\sin\left(x^2\right)\right) = \frac{\mathrm{d}}{\mathrm{d}x}(\sin(u)) = \frac{\mathrm{d}}{\mathrm{d}u}(\sin(u))\frac{\mathrm{d}u}{\mathrm{d}x} = 2x\cos\left(x^2\right)$$

Then, to find

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\sin\left(y^2\right)\right)$$

we just need to multiply by $\frac{\mathrm{d}y}{\mathrm{d}x}$, and so

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\sin\left(y^2\right)\right) = 2y\sin\left(y^2\right)\frac{\mathrm{d}y}{\mathrm{d}x}$$

Similarly, we can find

$$\frac{\mathrm{d}}{\mathrm{d}x}(e^y) = e^y\frac{\mathrm{d}y}{\mathrm{d}x}$$

because

$$\frac{\mathrm{d}}{\mathrm{d}x}(e^x) = e^x$$

So, putting it all together, we find

$$\begin{aligned} 2y\cos\left(y^2\right)\frac{\mathrm{d}y}{\mathrm{d}x} + e^y\frac{\mathrm{d}y}{\mathrm{d}x} &= 2x - 1 \\ \Rightarrow \quad \frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{2x-1}{2y\cos(y^2) + e^y} \end{aligned}$$

Now all we need to do to find the gradient of the tangent is to plug in the coordinate values:

$$\left.\frac{\mathrm{d}y}{\mathrm{d}x}\right|_{\left(\frac{1+\sqrt{5}}{2},0\right)} = \frac{2\left(\frac{1+\sqrt{5}}{2}\right) - 1}{2(0)\cos(0) + e^0} = \sqrt{5}$$

The next example is going to utilise the *product rule*. Before we get into it, consider finding

$$\frac{\mathrm{d}}{\mathrm{d}x}(xy)$$

The product rule says

$$\frac{\mathrm{d}}{\mathrm{d}x}(u(x)v(x)) = \frac{\mathrm{d}u}{\mathrm{d}x}v + u\frac{\mathrm{d}v}{\mathrm{d}x}$$

So, since we have a product here, we can apply this rule. Think of $u = x$ and $v = y$, then

$$\frac{\mathrm{d}}{\mathrm{d}x}(xy) = \frac{\mathrm{d}}{\mathrm{d}x}(x)y + x\frac{\mathrm{d}}{\mathrm{d}x}y = y + x\frac{\mathrm{d}y}{\mathrm{d}x}$$

We didn't learn anything new here, but it's important to remember that we need to apply the product rule in examples like this, and also to include any $\frac{\mathrm{d}y}{\mathrm{d}x}$ terms that we might need.

Example 19. *Find an expression for $\frac{\mathrm{d}y}{\mathrm{d}x}$ in the case that x and y satisfy*

$$\sin y + xy^2 - e^x = y$$

We act with $\frac{\mathrm{d}}{\mathrm{d}x}$ on both sides:

$$\frac{\mathrm{d}}{\mathrm{d}x}(\sin y) + \frac{\mathrm{d}}{\mathrm{d}x}\left(xy^2\right) - \frac{\mathrm{d}}{\mathrm{d}x}(e^x) = \frac{\mathrm{d}}{\mathrm{d}x}(y)$$

so

$$\frac{\mathrm{d}}{\mathrm{d}x}(\sin y) + \frac{\mathrm{d}}{\mathrm{d}x}\left(xy^2\right) - e^x = \frac{\mathrm{d}y}{\mathrm{d}x}$$

Now use the product and chain rule:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}y}(\sin y)\frac{\mathrm{d}y}{\mathrm{d}x} + \left(\frac{\mathrm{d}x}{\mathrm{d}x}y^2 + \frac{\mathrm{d}y^2}{\mathrm{d}y}\frac{\mathrm{d}y}{\mathrm{d}x}\right) &= \frac{\mathrm{d}y}{\mathrm{d}x} \\ \cos y\frac{\mathrm{d}y}{\mathrm{d}x} + \left(y^2 + 2xy\frac{\mathrm{d}y}{\mathrm{d}x}\right) &= \frac{\mathrm{d}y}{\mathrm{d}x} \end{aligned}$$

and we rearrange to find

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{-y^2}{\cos y + 2xy - 1}$$

4.3 Second derivatives

Now that we know how to find stationary points in functions that are defined implicitly, the next step is finding out how to classify these points. As in the case of explicit functions, we use the second derivative to do this.

Finding the second derivative requires no new machinery, but is just a matter of paying attention.

Example 20. *Find* $\frac{\mathrm{d}^2y}{\mathrm{d}x^2}$ *in the case that*

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

We start by taking a first derivative:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}x}\left(y^3 + 2xy + x^3\right) &= 0 \\ \Rightarrow \quad 3y^2\frac{\mathrm{d}y}{\mathrm{d}x} + 2y + 2x\frac{\mathrm{d}y}{\mathrm{d}x} + 3x^2 &= 0 \\ \Rightarrow \quad \frac{\mathrm{d}y}{\mathrm{d}x} &= -\frac{2y + 3x^2}{2x + 3y^2} \end{aligned}$$

To find the second derivative, we simply take derivatives of both sides again, but this time we need to use the quotient rule. We will do this very explicitly.

The quotient rule says:

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{u(x)}{v(x)}\right) = \frac{u'v - v'u}{v^2}$$

So, we define

$$\begin{aligned} u &= 2y + 3x^2 \\ v &= 2x + 3y^2 \end{aligned}$$

and then we find

$$\begin{aligned} \frac{\mathrm{d}u}{\mathrm{d}x} &= 2\frac{\mathrm{d}y}{\mathrm{d}x} + 6x \\ \frac{\mathrm{d}u}{\mathrm{d}x} &= 6y\frac{\mathrm{d}y}{\mathrm{d}x} + 2 \end{aligned}$$

We then make use of our knowledge of $\frac{\mathrm{d}y}{\mathrm{d}x}$ to find

$$\begin{aligned} \frac{\mathrm{d}u}{\mathrm{d}x} &= \frac{18xy^2 + 6x^2 - 4y}{3y^2 + 2x} \\ \frac{\mathrm{d}v}{\mathrm{d}x} &= \frac{4x - 18x^2y - 6y^2}{3y^2 + 2x} \end{aligned}$$

We can then combine all of this to obtain the (ugly) result

$$\frac{\mathrm{d}^2y}{\mathrm{d}x^2} = -\frac{(18xy^2 + 6x^2 - 4y)(2x + 3y^2) - (4x - 18x^2y - 6y^2)(2y + 3x^2)}{(2x + 3y^2)^3}$$

Which simplifies to

$$\frac{\mathrm{d}^2y}{\mathrm{d}x^2} = -\frac{2xy(27x^3 + 54xy + 27y^3 - 8)}{(2x + y^2)^3}$$

So, we see that finding second derivatives can be a pretty tedious process, especially when the functions aren't very friendly. Nonetheless, they can be handled with relative ease, provided care is taken over the algebra.

5 Derivatives of Inverse Trigonometric Functions

Now that we have learned the art of implicit differentiation, we can find the derivatives of inverse trigonometric functions.

Let $y = \arcsin x$, and look to find $\frac{\mathrm{d}y}{\mathrm{d}x}$. As a first step, lets take the sine of both sides to obtain

$$\sin y = x$$

Then we differentiate implicitly to obtain

$$\cos y \frac{\mathrm{d}y}{\mathrm{d}x} = 1$$

and so

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{1}{\cos y}$$

But this isn't exactly very useful. We need to find a way of expressing $\cos y$ in terms of x. To do so, we make use of the fact that $\sin^2 y + \cos^2 y = 1$, and thus

$$\cos y = \sqrt{1 - \sin^2 y} = \sqrt{1 - x^2}$$

Thence, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}x}(\arcsin x) = \frac{1}{\sqrt{1 - x^2}}$$

In a similar way, we can obtain the derivatives of standard inverse trigonometric functions:

- $$\frac{\mathrm{d}}{\mathrm{d}x}(\arcsin x) = \frac{1}{\sqrt{1 - x^2}}$$
- $$\frac{\mathrm{d}}{\mathrm{d}x}(\arccos x) = -\frac{1}{\sqrt{1 - x^2}}$$
- $$\frac{\mathrm{d}}{\mathrm{d}x}(\arctan x) = \frac{1}{1 + x^2}$$

6 Maclaurin Series

6.1 Taylor's theorem

Some functions can be very hard to handle. Take, for example, the function

$$f(x) = \frac{\sin x}{x}$$

How does f behave as we approach zero? More explicitly, what is value of the limit

$$\lim_{x \to 0} \frac{\sin x}{x}$$

With the tools that we currently have at our disposal, we would need to do quite a lot of work to evaluate this limit. As such, we look elsewhere for a method that will help us to accurately describe the behaviour of this function as $x \to 0$.

One possible route would be to try and express the function f in term of simpler functions whose behaviour we understand. If we could find a way to do this, then the limit would become trivial as we would be evaluating the limit of functions that behave in far less complex ways.

To make our task of finding such an expression more achievable, we start by restricting where we want our new form to be valid. It would be very ambitious to ask for an equivalent expression for our function that is valid at every point in our domain. Furthermore, it may not even be possible to find an equivalent representation in the desired way. As such, we will attempt to look for a *local approximation* to our function. That is, we look to express our function in an equivalent form that is accurate up to some error term, but only in the vicinity of some point $x = a$ within the domain of our function.

Now that we've posed the question that we want answered, we now ask what exactly is the best way for us to form such an approximation. We need to choose a functional form that is easy to compute and manipulate.

As far as functions go, polynomials are very easy to deal with: they can be integrated easily, there are simple rules for differentiating them, and we can sketch without much thought (relatively speaking). As such, we would like to find a way to write an arbitrary function as a sum of polynomial terms, that way we could handle it more easily. Let's try and do just this.

Consider the polynomial $p(x) = (x + 1)(x + 2)(x - 3)$. If we expand this, we find

$$p(x) = x^3 - 7x - 6$$

After some playing around, we find that we can write $p(x)$ as

$$p(x) = (x - 1)^3 + 3(x - 1)^2 - 4(x - 1) - 12$$

This can be thought of as an expanded form of p *about* the point $x = 1$. What do we mean by *expanded about*? Well, consider what happens when we substitute in $x = 1 + \epsilon$, where $|\epsilon| \ll 1$ (colloquially, ϵ is very "small"). We find

$$p(1 + \epsilon) = \epsilon^3 + 3\epsilon^2 - 4\epsilon - 12$$

If $|\epsilon| \ll 1$, then $|\epsilon|^2, |\epsilon|^3$ are negligibly small, and so we see

$$p(1 + \epsilon) \approx 4\epsilon - 12$$

So, what we have obtained is a linear approximation to the behaviour of $p(x)$ in a small neighbourhood of $x = 1$. In other words, we get a picture of how $p(x)$ behaves for x-values very close to $x = 1$. In a similar way, we could have written

$$p(x) = (x - 2)^3 + 6(x - 2)^2 + 5(x - 2) - 12$$

which would allow us to obtain similar approximations in a neighbourhood of $x = 2$.

It turns out, that any polynomial, $p(x)$, of degree n can be written as

$$p(x) = \sum_{k=0}^{n} c_k (x - a)^k$$

for any a of our choosing, thus allowing an approximation to the functions behaviour in any neighbourhood of our choosing.

For polynomial functions, this isn't exactly very impressive: it wouldn't have been much work to find $p(1+\epsilon)$ in its original form, and so we haven't really achieved very much. But consider the possibility of being able to obtain such a local approximation for other functions.

We would like to test the hypothesis that any (suitably differentiable) function $f(x)$ can be written in exactly this form.

As a start, let's take some function $f(x)$ and look for some expansion of it in a neighbourhood I containing our point of interest, $x = a$. To be more general, we will assume that only a finite number of the derivatives of f are continuous on our interval, namely that $f, f', f'', \cdots, f^{(n)}$ are continuous on I. We want to expand f about $x = a$ as sum of polynomial terms, so we write:

$$f(x) \approx c_0 + c_1(x-a) + c_2(x-a)^2 + c_3(x-a)^3 + \cdots$$

The question we now ask is how can we determine the values of the c_i? To do that, let's use the information that we know about the function. We know that when $x = a$, the function $f(x)$ takes the value $f(a)$. As such

$$f(a) = c_0 + c_1(a-a) + c_2(a-a)^2 + c_3(a-a)^3 + \cdots = c_0$$

and so we have the value of c_0. We also know that our function is n-times differentiable, and that, at the point $x = a$, the $f'(x)$ takes the value $f'(a)$. So,

$$\begin{aligned} f'(a) &= \frac{\mathrm{d}}{\mathrm{d}x}\left(c_0 + c_1(x-a) + c_2(x-a)^2 + c_3(x-a)^3 + \cdots\right)\Big|_{x=a} \\ &= \left(c_1 + 2c_2(x-a) + 3c_3(x-a)^2 + \cdots\right)\big|_{x=a} \\ &= c_1 \end{aligned}$$

which gives the value of c_1. We can do the same for $f''(a)$:

$$\begin{aligned} f''(a) &= \frac{\mathrm{d}^2}{\mathrm{d}x^2}\left(c_0 + c_1(x-a) + c_2(x-a)^2 + c_3(x-a)^3 + \cdots\right)\Big|_{x=a} \\ &= \left(2c_2 + 3 \times 2c_3(x-a) + \cdots\right)\big|_{x=a} \\ &= 2c_2 \\ \Rightarrow \quad c_2 &= \frac{1}{2!}f''(a) \end{aligned}$$

and $f'''(a)$:

$$\begin{aligned} f'''(a) &= \frac{\mathrm{d}^3}{\mathrm{d}x^3}\left(c_0 + c_1(x-a) + c_2(x-a)^2 + c_3(x-a)^3 + \cdots\right)\Big|_{x=a} \\ &= \left(3 \times 2c_3 + \cdots\right)\big|_{x=a} \\ &= 3 \times 2c_3 \\ \Rightarrow \quad c_3 &= \frac{1}{3!}f''(a) \end{aligned}$$

In fact, we can continue this process up until the nth derivative:

$$c_n = \frac{1}{n!}f^{(n)}(a)$$

After this point, we no longer have information as we don't know if our function is further differentiable.

As such, we are forced to stop evaluating the c_i when $i = n$. We thus arrive at the result:

$$f(x) \approx f(a) + f'(a)(x-a) + \frac{1}{2!}f''(a)(x-a)^2 + \frac{1}{3!}f'''(a)(x-a)^3 + \cdots + \frac{1}{n!}f^{(n)}(a)(x-a)^n$$

Notice the exchange of an exact equality for an approximation. The reason for this is that we were unable to continue our process of evaluating the c_i indefinitely, as we could only differentiate n times. So, rather than including unknown terms in our expression, we terminate the sum at the nth degree term and accept that there will be some inaccuracy in our approximation. We call this finite sum of terms the *Taylor polynomial* approximation to our function f at $x = a$.

To remedy the lack of accuracy due to the early termination of our sum, we introduce an nth-term error, $R_n(x)$, and so write

$$f(x) = f(a) + f'(a)(x-a) + \frac{1}{2!}f''(a)(x-a)^2 + \frac{1}{3!}f'''(a)(x-a)^3 + \cdots + \frac{1}{n!}f^{(n)}(a)(x-a)^n + R_n(x)$$

where we have returned to an exact equality as we assume this $R_n(x)$ term 'contains all of the error' and so we now have the exact value.

We have thus succeeded in obtaining a polynomial approximation to our function f, up to some error. There are in fact lots of methods known for evaluating this error and bounding it in different ways, but we won't concern ourselves with those.

For the sake of precision, we write Taylor's full theorem below:

Theorem 1 (Taylor's Theorem)**.** *Suppose that f is defined on some open interval I around a and suppose $f^{(n+1)}$ exists on this interval. Then, for each $x \in I$, with $x \neq a$, there is some $c \in I$ such that*

$$f(x) = \sum_{k=0}^{n} \frac{1}{k!}f^{(k)}(a)(x-a)^k + R_n(x)$$

where

$$R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}$$

We have stated the error of our approximation in one of its various forms; but, as stated above, we won't be concerning ourselves with these errors. For certain functions, we find that the error term tends to 0 as we let $n \to \infty$, and thus the approximation becomes exact as we take the sum to infinity. It is with functions of this form that we will be concerning ourselves.

6.2 Taylor polynomials vs. Taylor series

Before moving on to actually calculating Taylor series for functions, there is a subtle difference that it is worth concerning ourselves with. A Taylor polynomial for a function $f(x)$ at a point $x = a$ is, as stated above, a finite sum of the form

$$P_n(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k$$

which provides an approximation for $f(x)$ in some neighbourhood of $x = a$.

It is vital here that we have only a *finite* number of terms in our sum. If it was the case that our sum did not contain a finite number of terms, it would cease to be a polynomial (by definition, a polynomial must have finite degree). In this case, we obtain the *Taylor series* for our function $f(x)$ at the point $x = a$:

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(x-a)^k$$

which may or may not converge. For functions that we will be dealing with, this sum will converge on some domain, and provides an exact representation of our function. In this case, the error term $R_n(x)$ satisfies the property that

$$\lim_{n\to\infty} R_n(x) = 0$$

and thus that there is no error in the approximation.

6.3 Radius of convergence

The keen eyed amongst you may have noticed the constant reference to our approximation being valid only in a neighbourhood of our point $x = a$. This is in fact a subtlety that will be very important and one that we need pay attention to.

Just above, we discussed the possibility that our Taylor series may or may not be convergent. Although, for examination purposes, we will only worry about those series that are always convergent, for the time being let's take a look at those that are not.

For a Taylor series to be convergent, we said that the error term must satisfy:

$$\lim_{n\to\infty} R_n(x) \to 0$$

In this case, there is no error term, and so out series is an exact representation of our function.

What is important is that this error term may only vanish in the limit a certain domain, say

$$\lim_{n\to\infty} R_n(x) = 0$$

only when

$$x \in (a-\delta, a+\delta)$$

for some $\delta > 0$. We call this $\delta > 0$ the *radius of convergence* and the interval $(a-\delta, a+\delta)$ the *interval of convergence*. For certain well-behaved functions, it happens to be that the interval of convergence is $\mathbb{R}$, and so the radius of convergence is ∞. This is actually an amazing property, as it means that our Taylor series provides an exact representation of our function everywhere on the real line: that's quite something!

6.4 Maclaurin series

A Maclaurin series is nothing more than a Taylor series where our centre is the origin, i.e. a Taylor series with $x = a$. In other words, the Maclaurin series expansion of a function $f(x)$ takes the form

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

We've talked in a very general way about these series so far, so let's get down to calculating a few examples.

Example 21. *Find the Maclaurin series for the function* $f(x) = \sin x$

We are going to take it as fact (as we always will in this course) that $\sin x$ is one of those very nice functions whose series converges over the entire real line. This means that we can write

$$\sin x = \sum_{k=0}^{\infty} \frac{1}{n!} \left. \frac{\mathrm{d}^k}{\mathrm{d}x^k}(\sin x)\right|_{x=0}$$

Before we calculate the Maclaurin series in full, let's stop and consider a few Taylor polynomials for this function. Recall that a Taylor polynomial of degree n in a neighbourhood of 0 is a polynomial $P_n(x)$ defined by

$$P_n(x) = \sum_{k=0}^{n} \frac{f^{(k)}(0)}{k!} x^k$$

We will calculate a few of these polynomials and compare them to $\sin x$ to get a feel for the accuracy of their approximations.

We know that all of our derivatives are going to be sine or cosine terms (in alternating fashion) and so will take the values of ± 1 or 0 at $x = 0$. We find

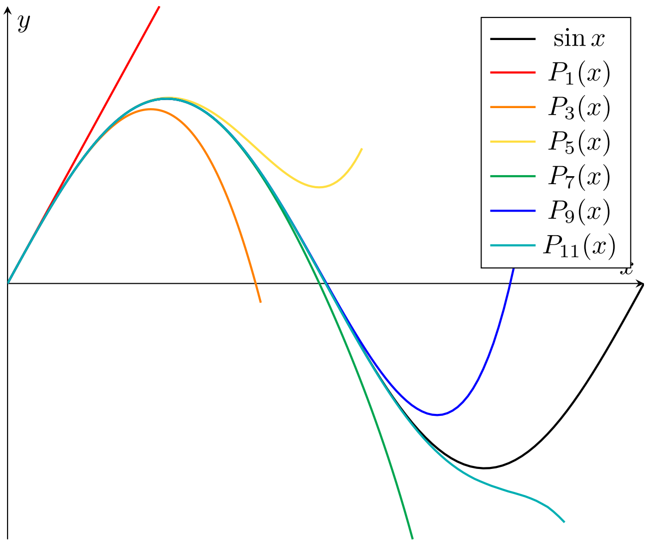
$$\begin{aligned}
\left. \frac{\mathrm{d}}{\mathrm{d}x}(\sin x)\right|_{x=0} &= 1 \\
\left. \frac{\mathrm{d}^2}{\mathrm{d}x^2}(\sin x)\right|_{x=0} &= 0 \\
\left. \frac{\mathrm{d}^3}{\mathrm{d}x^3}(\sin x)\right|_{x=0} &= -1 \\
\left. \frac{\mathrm{d}^4}{\mathrm{d}x^4}(\sin x)\right|_{x=0} &= 0 \\
\left. \frac{\mathrm{d}^5}{\mathrm{d}x^5}(\sin x)\right|_{x=0} &= 1 \\
&\vdots
\end{aligned}$$

As such, we can use our formula to obtain successive approximations to $\sin x$ in the vicinity of $x = 0$:

$$\sin x \approx P_1(x) = x$$
$$\sin x \approx P_3(x) = -\frac{1}{3!}x^3$$
$$\sin x \approx P_5(x) = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5$$
$$\sin x \approx P_7(x) = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \frac{1}{7!}x^7$$

and so on.

Now, to see the increasing accuracy of our approximations to $\sin x$, we plot the polynomial next to $\sin x$ in the range $[0, 2\pi]$:



Notice how, as we add more and more terms, the approximation to $\sin x$ becomes increasingly more accurate. In this way, as we extend the sum to ∞ we can imagine that the approximation becomes exact.

OK, let's now find the full Maclaurin series. We have spotted a pattern in the derivatives. Namely, that

$$f^k(0) = \begin{cases} 0 & k = 2n, \quad n \in \mathbb{N} \\ (-1)^n & k = 2n+1, \quad n \in \mathbb{N} \end{cases}$$

As such, we can write down the full Maclaurin series to be

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

Example 22. *Calculate the Maclaurin series of the function* $f(x) = e^x$

This is the simplest Maclaurin series there is. The reason for this is the fact that e^x is an *eigenfunction* of the derivative operator. This is just a technical way of saying that when we take the derivative of e^x, the function does not change. In this way, we know

$$f^{(n)}(x) = e^x, \quad \text{for all } n$$

and so

$$f^{(n)}(0) = 1, \quad \text{for all } n$$

As such, all we need to do is plug our results into the formula above to obtain:

$$e^x = \sum_{n=0}^{\infty} \frac{1}{n!} x^n$$

Example 23. *Calculate the Maclaurin series of the function* $f(x) = \cos x$

As before, we start by evaluating a few derivatives and try to see if we spot a pattern:

$$\begin{aligned}
f(x) &= \cos x \\
f^{(1)}(x) &= -\sin x \\
f^{(2)}(x) &= -\cos x \\
f^{(3)}(x) &= \sin x \\
f^{(4)}(x) &= \cos x
\end{aligned}$$

We see that the derivatives will be cyclical with a period of 4. We then evaluate the derivatives at the point about which we are expanding (which will always be 0) for a Maclaurin series:

$$\begin{aligned}
f(0) &= 1 \\
f^{(1)}(0) &= 0 \\
f^{(2)}(0) &= -1 \\
f^{(3)}(0) &= 0 \\
f^{(4)}(0) &= 1
\end{aligned}$$

As such, we deduce

$$f^{(n)}(0) = \begin{cases} 0 & n = 2k+1, \quad k \in \mathbb{N} \\ (-1)^k & n = 2k, \quad k \in \mathbb{N} \end{cases}$$

We can then plug this all in to obtain

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

Example 24. *Calculate the Maclaurin seires for the function* $f(x) = \ln(1+x)$

Before reading on, ask yourself why we look to find the Maclaurin series of $\ln(1+x)$ and not of $\ln x$.

We calculate a few derivatives:

$$\begin{aligned}
f(x) &= \ln(1+x) \\
f^{(1)}(x) &= \frac{1}{1+x} \\
f^{(2)}(x) &= \frac{-1}{(1+x)^2} \\
f^{(3)}(x) &= \frac{2}{(1+x)^3} \\
f^{(4)}(x) &= \frac{-2 \times 3}{(1+x)^4} \\
f^{(5)}(x) &= \frac{2 \times 3 \times 4}{(1+x)^5}
\end{aligned}$$

and so

$$\begin{aligned}
f(0) &= 0 \\
f^{(1)}(0) &= 1 \\
f^{(2)}(0) &= -1 \\
f^{(3)}(0) &= 2! \\
f^{(4)}(0) &= -3! \\
f^{(5)}(0) &= 4!
\end{aligned}$$

If you look closely, you will spot the pattern

$$f^{(n)}(0) = (-1)^{n+1}(n-1)!, \quad n \geq 1$$

with $f(0) = 0$. Hence,

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

6.5 Substitution and products to obtain series

The final idea we need to get our head around is how we manipulate series that we already know to obtain series for other functions.

The reason that we can do this is that we consider functions with the entire real line as their radius of convergence, and so making substitutions into series can be done with little care.

As an example, consider the following

Example 25. *Use the Maclaurin series for e^x, and an appropriate substitution, to obtain the Maclaurin series for e^{x^2}*

Defining $f(x) = e^x$, it is pretty obvious that $e^{x^2} = f(x^2)$. Now, in the previous section, we showed that

$$e^x = f(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

Taking it to be true that this series converges over the entire real line, it follows that

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

converges over the entire real line as well. However, we know that $f(x^2) = e^{x^2}$, and so we deduce that the Maclaurin series for e^{x^2} must be

$$e^{x^2} = \sum_{n=0}^{\infty} \frac{x^{2n}}{n!}$$

So we see that we can make use of sensible substitutions to find Maclaurin series for functions where finding them explicitly would require a lot of work.

As another example, consider how we can use the Maclaurin series of two basic functions to find the first few terms of a more complex one.

Example 26. *Find the first 4 terms in the Maclaurin series for the function*

$$e^x \sin\left(x^2\right)$$

From our earlier work, we can write down the Maclaurin series for both $\sin x$ and e^x:

$$\begin{aligned}
\sin x &= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots \\
e^x &= 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots
\end{aligned}$$

We now know that we can make substitutions to find the Maclaurin series for complicated functions, and so we let $x \mapsto x^2$ in the series for $\sin x$ to find

$$\sin\left(x^2\right) = (x^2) - \frac{(x^2)^3}{3!} + \frac{(x^2)^5}{5!} - \frac{(x^2)^7}{7!} + \cdots = x^2 - \frac{x^6}{3!} + \frac{x^{10}}{5!} - \frac{x^{14}}{7!} + \cdots$$

and so to find the Maclaurin series for $e^x \sin\left(x^2\right)$, we can simply say

$$\begin{aligned}
e^x \sin\left(x^2\right) &= \left(1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots\right)\left(x^2 - \frac{x^6}{3!} + \frac{x^{10}}{5!} - \frac{x^{14}}{7!} + \cdots\right) \\
&= x^2 + x^3 + \frac{x^4}{2} + \frac{x^5}{6} + \cdots
\end{aligned}$$

where we only write the first 4 terms, as was required.

7 Introduction to Integration

7.1 Motivation

We can think of integration (in a very colloquial way) as a sort of *super sum.* Integration is a way for us to sum infinitesimally small values and obtain a finite value as an output.

As far as physical motivation is concerned, we have two main areas of concern. Our first is the direct summation of some physical quantity. Take, for example, the problem of finding the mass of a non-uniform rod. Say this rod has a *density* of $\mu(x)$ - meaning that the density is a given function of the length, x, along the rod (defining one end of the rod to be 0 on the x-axis). If we had a uniform rod with mass μ, all we would do is multiply the length ℓ by the density to get the mass m

$$m = \mu\ell$$

But how do we do this when we have our smooth, variable density function $\mu(x)$? The answer to the question will be to *integrate* the function.

The second area of motivation comes in "reversing" differentiation. For example, consider Newton's second law:

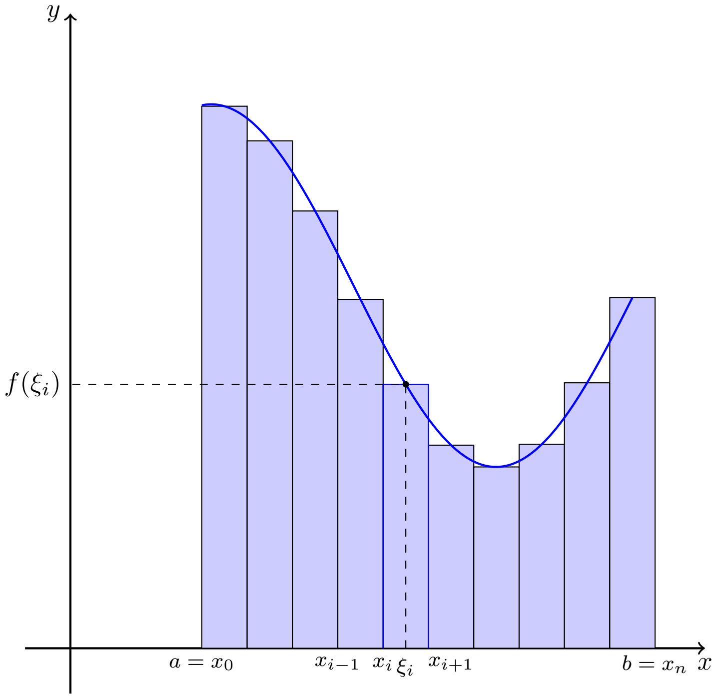
$$F = \frac{\mathrm{d}p}{\mathrm{d}t}$$

where F is force and p is momentum. Given this, how would we find the momentum at some time t? To do this, we would need integrate both sides to essentially "undo" the the derivative and obtain the momentum p.

7.2 Derivation of the Definite Integral

We concern ourselves with two types of integral: the *definite integral* and the *indefinite* integral. We first consider the definite integral.

Consider a function, $f(x)$, whose graph we display below. Suppose we want to find the area underneath this graph between two x-values: $x = a$ and $x = b$. To do this, we first split the interval $[a, b]$ into n separate parts, with each being of width $\Delta x = \frac{b-a}{n}$. We thus obtain a sequence of x values, $\{x_0, x_2, \cdots, x_n\}$, with $x_i - x_{i-1} = \Delta x$, and $x_0 = a$, $x_n = b$.



Now that we have the width of the rectangle, we need to find its height. This is actually quite straightforward: since we have drawn the rectangles such that they are a good estimate of the area enclosed under the curve, we see that the ith rectangle intersects the curve $y = f(x)$ at a point $f(\xi_i)$, with $\xi_i \in [x_i, x_{i+1}]$.

A good estimate for the area would then be

$$A = \sum_{i=0}^{n} f(\xi_i)\Delta x$$

Intuitively, as the width of the rectangles is decreased, we would expect the approximation of our sum to the area enclosed under the curve to become more and more exact. We thus take the limit as $n \to \infty$ and obtain the *definite integral* of $f(x)$ between $x = a$ and $x = b$:

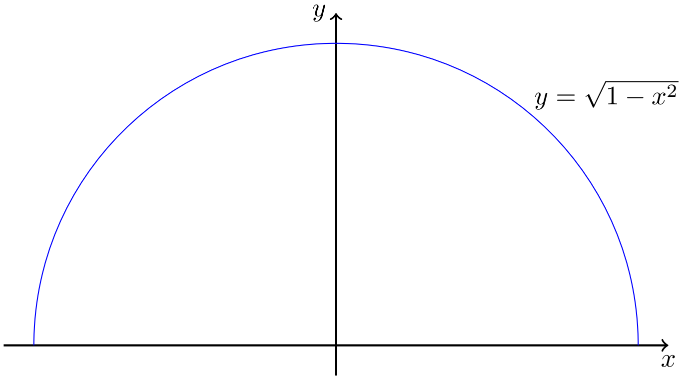
$$\int_a^b f(x)\mathrm{d}x = \lim_{n\to\infty} \sum_{i=0}^{n} f(\xi_i)\Delta x, \quad \Delta x = \frac{b-a}{n}$$

But what does this actually mean? Can we calculate this quantity? The answer, as you should have guessed, is yes.

One way we could assign a numerical value to this integral would be to calculate the are underneath the graph in question. For example, consider the integral

$$\int_{-1}^{1} \sqrt{1-x^2}\mathrm{d}x$$

What does this curve look like? We draw this graph below:



All we have here is a semi-circle and we know how to find the enclosed area. As such, it follows simply that

$$\int_{-1}^{1} \sqrt{1-x^2}\mathrm{d}x = \frac{1}{2}\pi$$

But this method falls apart quickly when we have a more complicated curve for which there's no standard formula for finding its area. It is at this point that we need to develop a method for assigning numerical values to our definite integrals that does not require the ability to directly calculate the area.

To do this, we need to first consider our other type of integral: the *indefinite integral*, or *antiderivative.*

7.3 Indefinite integration

Where we think of the definite integral of a function f between two points as the area contained beneath the graph of f between the specified boundary, we think of the indefinite integral of f as giving us the function F such that $F' = f$, i.e. a function that, when differentiated, gives us f.

For example, consider

$$F(x) = \int x^n \mathrm{d}x$$

This tells us that $F(x)$ is a function that differentiates to give us x^n. But what would such a function look like? Well, when we differentiate a polynomial term, we take 1 away from the power. As such, a sensible guess would be

$$F(x) = x^{n+1}$$

Just to check, let's differentiate $F(x)$:

$$\frac{\mathrm{d}F(x)}{\mathrm{d}x} = \frac{\mathrm{d}}{\mathrm{d}x}\left(x^{n+1}\right) = (n+1)x^n$$

So we see that we are out by a factor of $n+1$. To remedy this, we simply divide by this and say

$$F(x) = \frac{1}{n+1}x^{n+1}$$

But we still aren't quite there yet. What would the derivative of $\frac{1}{n+1}x^{n+1} + 5$ be? Since the derivative of a constant is 0, we see

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{n+1}x^{n+1} + 5\right) = x^n$$

In fact, this is evidently true for any constant that we add on. Thus, the true *antiderivative* of x^n is

$$\frac{1}{n+1}x^{n+1} + C$$

where C is some constant. We write

$$\int x^n \mathrm{d}x = \frac{1}{n+1}x^{n+1} + C$$

A similar line of reasoning can be used to obtain the antiderivative of most standard functions. For example,

$$\int \sin x \mathrm{d}x = -\cos x + C$$

since

$$\frac{\mathrm{d}}{\mathrm{d}x}(\cos x + C) = -\sin x$$

OK, so we have now formulated how to find an antiderivative, and we now need to be able to relate this to a definite integral in order to complete our goal. Our final step in completing the definite integral requires an exceptionally important theorem. In fact, this theorem is so important that its called *fundamental*.

7.4 The Fundamental Theorem of Calculus

We state the Fundamental Theorem of Calculus as follows:

Theorem 2. *Let f be a continuous, real-valued function on the closed interval $[a, b]$. Let F be the function defined for all $x \in [a, b]$ by*

$$F(x) = \int_a^x f(t)\mathrm{d}t$$

Then $F(x)$ is uniformly continuous on the closed interval $[a, b]$ and differentiable on the open interval (a, b). In particular,

$$F'(x) = f(x)$$

A decent proof of this theorem requires familiarity with $\epsilon - \delta$ style proofs and so isn't worth noting down, but it is included in the appendices for the sake of completeness.

This isn't the form of the FTC that we are concerned with. What we want to use is a very important corollary of the FTC (again, the proof is included in the appendices):

Corollary 1. *If f is continuously differentiable on $[a, b]$, then*

$$\int_a^b f'(t)\mathrm{d}t = f(b) - f(a)$$

7.5 Calculating definite integrals

We now (finally) have the tools we need to assign a value to a definite integral. We see from the FTC that the definite integral of a function $f(x)$ between $x = a$ and $x = b$ is simply the difference antiderivative of $f(x)$ at the two endpoints, a and b. This may seem "obvious" because that is what you have been told is the value of a definite integral, but consider where this actually came from.

We were able to define the expression

$$\int_a^b f(x)\mathrm{d}x$$

by considering the area beneath a graph of $f(x)$ between the relevant end points. Once we obtained this definition, however, we had no means of assigning a numerical value to it. To do this, we had to first consider an antiderivative - which we defined simply as being an operation that "reverses" the action of the derivative operator. The FTC then gave us a means of taking this idea of an antiderivative and applying it to a definite integral. It tells us that, in fact,

$$\int_a^b f(x)\mathrm{d}x$$

is the difference between the function values of $g(x)$ at both end points, where $g(x)$ is such that $g'(x) = f(x)$.

So, for example, to evaluate

$$\int_{\pi/2}^{\pi} \sin x \mathrm{d}x$$

we need to note

$$\int \sin x \mathrm{d}x = -\cos x + C$$

and so

$$\int_{\pi/2}^{\pi} \sin x \mathrm{d}x = (-\cos(\pi) + C) - \left(-\cos\left(\frac{\pi}{2}\right) + C\right) = -\cos(\pi) + \cos\left(\frac{\pi}{2}\right) = 1$$

In particular, notice how the constant of integration, C, has "disappeared" from the final result. This is self-evident when you consider the stages in deriving the definite integral above as the constant appears in both the term we are subtracting and the term we are subtracting from and hence "cancels".

8 Integration Theorems

8.1 Basic results

Now that we have defined our integral and explored, to some extent, how we can think about the definite and indefinite integral, we now establish some properties of the integral.

Lemma 1.

$$\int_a^b f(x)\mathrm{d}x = -\int_b^a f(x)\mathrm{d}x$$

Proof. We have

$$\begin{aligned}
\int_a^b f(x)\mathrm{d}x &= \lim_{n\to\infty}\sum_{i=0}^{n} f(\xi_i)\left(\frac{b-a}{n}\right) \\
&= \lim_{n\to\infty}\sum_{i=0}^{n} f(\xi_i)\left(\frac{-(a-b)}{n}\right) \\
&= \lim_{n\to\infty}\left(-\sum_{i=0}^{n} f(\xi_i)\left(\frac{a-b}{n}\right)\right) \\
&= -\lim_{n\to\infty}\sum_{i=0}^{n} f(\xi_i)\left(\frac{b-a}{n}\right) \\
&= -\int_b^a f(x)\mathrm{d}x
\end{aligned}$$

□

Lemma 2. *For any constant c, we have*

$$\int_a^b cf(x)\mathrm{d}x = c\int_a^b f(x)\mathrm{d}x$$

Proof. We have

$$\begin{aligned}
\int_a^b cf(x)\mathrm{d}x &= \lim_{n\to\infty}\sum_{i=0}^{n} cf(\xi_i)\left(\frac{a-b}{n}\right) \\
&= c\left(\lim_{n\to\infty}\sum_{i=0}^{n} f(\xi_i)\left(\frac{a-b}{n}\right)\right) \\
&= c\int_a^b f(x)\mathrm{d}x
\end{aligned}$$

□

Lemma 3.

$$\int_a^b \left(f(x)\pm g(x)\right)\mathrm{d}x = \int_a^b f(x)\mathrm{d}x \pm \int_a^b g(x)\mathrm{d}x$$

Proof. We have

$$\begin{aligned}
\int_a^b \left(f(x)\pm g(x)\right)\mathrm{d}x &= \lim_{n\to\infty}\sum_{i=0}^{n}\left(f(\xi_i)\pm g(\xi_i)\right)\left(\frac{a-b}{n}\right) \\
&= \lim_{n\to\infty}\sum_{i=0}^{n}\left(f(\xi_i)\left(\frac{a-b}{n}\right)\pm g(\xi_i)\left(\frac{a-b}{n}\right)\right) \\
&= \lim_{n\to\infty}\left[\sum_{i=0}^{n} f(\xi_i)\left(\frac{a-b}{n}\right)\pm\sum_{i=0}^{n} g(\xi_i)\left(\frac{a-b}{n}\right)\right] \\
&= \lim_{n\to\infty}\sum_{i=0}^{n} f(\xi_i)\left(\frac{a-b}{n}\right)\pm\lim_{n\to\infty}\sum_{i=0}^{n} g(\xi_i)\left(\frac{a-b}{n}\right) \\
&= \int_a^b f(x)\mathrm{d}x \pm \int_a^b f(x)\mathrm{d}x
\end{aligned}$$

□

These three results are very useful. For example, say we wanted to calculate:

$$\int_0^1 \left(x^3 + 3x^2\right) \mathrm{d}x$$

We can use our results to split the integral as:

$$\int_0^1 x^3 \mathrm{d}x + 3\int_0^1 x^2 \mathrm{d}x$$

and then calculate each of these integrals.

9 Integration by substitution

9.1 The theorem

Theorem 3 (Integration by Substitution)**.** *Let $f : [a, b] \to \mathbb{R}$ be continuous. Let $g : [u, v] \to \mathbb{R}$ be continuously differentiable, and suppose $g(u) = a$, $g(v) = b$, and f is defined everywhere on $g([u, v])$ (and is continuous). Then*

$$\int_a^b f(x)\mathrm{d}x = \int_u^v f(g(t))g'(t)\mathrm{d}t$$

Proof. By the FTC, f has an antiderivative F defined on $g([u, v])$. Then

$$\begin{aligned}\int_u^v f(g(t))g'(t)\mathrm{d}t &= \int_u^v F'(g(t))g'(t)\mathrm{d}t \\ &= \int_u^v (F \circ g)'(t)\mathrm{d}t \\ &= F \circ g(v) - F \circ g(u) \\ &= F(b) - F(a) \\ &= \int_a^b f(x)\mathrm{d}x\end{aligned}$$

□

9.2 Some simple examples

An integral is best handled with a substitution when it contains a function and its derivative. The tricky bit can be spotting when this is the case, but through sufficient practice this will become easier and easier.

Example 27.

$$\int (2x - 1)e^{x^2-1}\mathrm{d}x$$

To handle this, we need to look at it carefully. If we stare at this integral long enough, then hopefully we will see that $\frac{\mathrm{d}}{\mathrm{d}x}\left(x^2 - x\right) = 2x-1$. This is shows that the question is primed and ready for a sensible substituion.

With this in mind, let's set $u = x^2 - 1$. It then follows

$$\frac{\mathrm{d}u}{\mathrm{d}x} = 2x - 1$$

Thence

$$\int (2x - 1)e^{x^2-x}\mathrm{d}x = \int (2x - 1)e^{x^2-x}\frac{\mathrm{d}x}{\mathrm{d}u}\mathrm{d}u = \int (2x - 1)e^u\frac{1}{2x - 1}\mathrm{d}u = \int e^u\mathrm{d}u = e^u + C = e^{x^2-x} + C$$

Example 28.

$$\int \frac{x}{\sqrt{1 - 2x^2}}\mathrm{d}x$$

We notice that

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

and so the numerator of our expression is a scalar multiple of the term inside the root at the bottom. As such, it makes sense to let $u = 1 - 2x^2$, and then

$$\frac{\mathrm{d}u}{\mathrm{d}x} = -4x$$

From which we obtain

$$\int \frac{x}{\sqrt{1 - 2x^2}}\mathrm{d}x = \int \frac{x}{\sqrt{1 - 2x^2}}\frac{\mathrm{d}x}{\mathrm{d}u}\mathrm{d}u = \int \frac{x}{\sqrt{u}}\frac{1}{-4x}\mathrm{d}u = -\frac{1}{4}\int u^{-1/2}\mathrm{d}u = -\frac{1}{2}\sqrt{1 - 2x^2} + C$$

Example 29.

$$\int \sin x \cos^4 x \mathrm{d}x$$

Again, we are looking for a combination of a function and a (scalar multiple of) its derivative. Notice here that we have trigonometric terms, and so this is a prime example for a substitution. We let $u = \cos x$ and then

$$\int \sin x \cos^4 x \mathrm{d}x = \int \sin x u^4 \frac{\mathrm{d}x}{\mathrm{d}u} \mathrm{d}u = \int \sin x u^4 \frac{1}{-\sin x} \mathrm{d}u = -\int u^4 \mathrm{d}u = \frac{1}{5} \cos^5 x + C$$

Example 30.

$$\int \frac{1}{x \ln x} \mathrm{d}x$$

This isn't as obvious a candidate for substitution as the others have been, but if we look hard enough, we will see that in fact it is. Notice that if we let $u = \ln x$

$$\frac{\mathrm{d}u}{\mathrm{d}x} = \frac{1}{x}$$

Then

$$\int \frac{1}{x \ln x} \mathrm{d}x = \int \frac{1}{xu} \frac{\mathrm{d}x}{\mathrm{d}u} \mathrm{d}u = \int \frac{1}{xu} x \mathrm{d}u = \ln |u| + C = \ln |\ln x| + C$$

Now let's look at some examples of definite integrals that we handle with substitution. The ideas of when to use a substitution don't change, but this time we have an extra step in that we need to consider our limits. Let's look at some examples.

Example 31.

$$\int_{-\pi}^{\pi/2} \cos x \cos(\sin x) \mathrm{d}x$$

We have a combination of trigonometric functions, and we find that $u = \sin x$ is the best substitution. We then have

$$\frac{\mathrm{d}u}{\mathrm{d}x} = \cos x$$

Now we need to think about the limits of the integral. Our lower bound is $-\pi$ and our upper bound is $\pi/2$:

$$\begin{aligned} x = -\pi \quad &\Rightarrow \quad u = \sin(-\pi) = 0 \\ x = \pi/2 \quad &\Rightarrow \quad u = \sin(\pi/2) = 1 \end{aligned}$$

Hence,

$$\int_{-\pi}^{\pi/2} \cos x \cos(\sin x) \mathrm{d}x = \int_0^1 \cos x \cos u \frac{\mathrm{d}x}{\mathrm{d}u} \mathrm{d}u = \int_0^1 \cos u \mathrm{d}u = [\sin u]_0^1 = \sin 1$$

Example 32.

$$\int_1^2 \frac{1}{w^2} e^{2/w} \mathrm{d}w$$

Some thinking will show you that the sensible substitution is to let $u = 2/w$, then

$$\frac{\mathrm{d}u}{\mathrm{d}w} = -\frac{2}{w^2}$$

Now we consider the limits:

$$\begin{aligned} w = 1 \quad &\Rightarrow \quad u = 2 \\ w = 2 \quad &\Rightarrow \quad u = 1 \end{aligned}$$

Then

$$\int_1^2 \frac{1}{w^2} e^{2/w} \mathrm{d}w = \int_2^1 \frac{1}{w^2} e^u \frac{\mathrm{d}w}{\mathrm{d}u} \mathrm{d}u = \int_2^1 \frac{1}{w^2} e^u \frac{-w^2}{2} \mathrm{d}u = -\frac{1}{2} \int_2^1 e^u \mathrm{d}u = \frac{1}{2} \int_1^2 e^u \mathrm{d}u = \frac{e}{2}(e-1)$$

9.3 Trigonometric substitutions

Here we look at some examples of where our integrals are based solved using a trigonometric substitution.

The idea of a trigonometric substitution rests upon the realisation that our integral takes the form of some variant of the fundamental identity

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

Once we identify the form it most resembles, we can make the substitution most appropriate to it and (hopefully) drastically simplify our integral!

Example 33.

$$\int \sqrt{1-x^2}\mathrm{d}x$$

We use this example to illustrate the relevant ideas. The premise is to use the identity

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

So, if we make the substitution $x = \sin\theta$, we obtain

$$\int \sqrt{1-x^2}\mathrm{d}x = \int \sqrt{1-\sin^2\theta}\cos\theta\mathrm{d}\theta$$

Then we use the identity highlighted above to rewrite this as

$$\int \cos^2\theta\mathrm{d}\theta$$

and then

$$\int \cos^2\theta\mathrm{d}\theta = \frac{1}{2}\int (\cos 2\theta + 1)\,\mathrm{d}\theta = \frac{1}{2}\theta + \frac{1}{4}\sin 2\theta + C$$

and we have the result! However, our original integral is given in terms of the variable x, and so we must transform back from θ to x. The θ term is simple since we just invert our substitution to obtain $\theta = \arcsin x$. The $\sin 2\theta$ term, however, is a little harder to handle. To do this, we notice

$$\sin 2\theta = 2\sin\theta\cos\theta = 2\sin\theta\sqrt{1-\sin^2\theta} = 2x\sqrt{1-x^2}$$

and so

$$\int \sqrt{1-x^2}\mathrm{d}x = \frac{1}{2}\left(\arcsin x + x\sqrt{1-x^2}\right) + C$$

Look back to the example of the area under a semicircle that we discussed back in section 6.2. Here we used a geometric argument to establish:

$$\int_{-1}^{1} \sqrt{1-x^2}\mathrm{d}x = \frac{1}{2}\pi$$

Let's use the method we have just discussed and see if we arrive at the same result:

$$\int_{-1}^{1} \sqrt{1-x^2}\mathrm{d}x = \left[\frac{1}{2}\left(\arcsin x + x\sqrt{1-x^2}\right)\right]_{-1}^{1} = \frac{1}{2}(\arcsin(1) - \arcsin(-1)) = \frac{1}{2}\left(\frac{\pi}{2} - \left(-\frac{\pi}{2}\right)\right) = \frac{1}{2}\pi$$

and so we are in agreement!

Example 34.

$$\int \sqrt{9-16x^2}\mathrm{d}x$$

This involves a term that mind lend its hand to $\sin^2\theta + \cos^2\theta = 1$, but we run into trouble with the numbers floating around inside the square root. How can we deal with this? A clever idea would be to factor out the 3^2 and notice that we can then write:

$$\int \sqrt{9-16x^2}\mathrm{d}x = \int 3\sqrt{1-\left(\frac{4x}{3}\right)^2}\mathrm{d}x = 3\int \sqrt{1-\left(\frac{4x}{3}\right)^2}\mathrm{d}x$$

Now we make the substitution

$$\frac{4x}{3} = \sin\theta$$

and so

$$\int \sqrt{9-16x^2}\mathrm{d}x = 3\int \sqrt{1-\sin^2\theta}\left(\frac{3}{4}\cos\theta\right)\mathrm{d}\theta = \frac{9}{4}\int \cos^2\theta\mathrm{d}\theta = \frac{9}{8}\left(\theta + \frac{1}{2}\sin 2\theta\right) + C$$

and we invert to find

$$\int \sqrt{9-16x^2}\mathrm{d}x = \frac{9}{8}\left(\arcsin\left(\frac{4x}{3}\right) + \frac{4x}{3}\sqrt{1-\left(\frac{4x}{3}\right)^2}\right) + C$$

If we can make substitutions of the form $x = a\sin\theta$, we may be asking ourselves if there are occasions for us to use other trigonometric substitutions to simplify integrals. The answer is (of course) that there are indeed integrals suited different trig substitutions.

Example 35.

$$\int \frac{x^3}{\sqrt{1+9x^2}}\mathrm{d}x$$

In our previous examples, we have looked to utilise the identity $\sin^2\theta + \cos^2\theta = 1$. But this doesn't seem to be the form of the integral we have here. To choose an appropriate substitution, it would be sensible to consider other trigonometric identities, for example

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

Now this looks a lot like what we have int the denominator above. Let's try the substitution $3x = \tan\theta$ $(9x^2 = (3x)^2)$. Then

$$\frac{\mathrm{d}x}{\mathrm{d}\theta} = \frac{1}{3}\sec^2\theta$$

and

$$\int \frac{x^3}{\sqrt{1+x^2}}\mathrm{d}x = \frac{1}{81}\int \frac{\tan^3\theta}{\sqrt{1+\tan^2\theta}}\sec^2\theta\mathrm{d}\theta = \frac{1}{81}\int \frac{\tan^3\theta\sec^2\theta}{\sec\theta}\mathrm{d}\theta = \frac{1}{81}\int \tan^3\theta\sec\theta\mathrm{d}\theta$$

So we see that we have simplified our integral quite a bit. How can we handle this integral now? A sensible idea would be to write $\tan^2\theta = \sec^2\theta - 1$. Then

$$\frac{1}{81}\int \tan^3\theta\sec\theta\mathrm{d}\theta = \frac{1}{81}\int \left(\tan\theta\sec^3\theta - \tan\theta\sec\theta\right)\mathrm{d}\theta = \frac{1}{81}\int \sec^3\theta\tan\theta\mathrm{d}\theta - \frac{1}{81}\int \sec\theta\tan\theta\mathrm{d}\theta$$

To evaluate these integrals, we need to recall the fact that $\frac{\mathrm{d}}{\mathrm{d}\theta}\sec\theta = \sec\theta\tan\theta$. Then we can write

$$= \frac{1}{81}\int \frac{1}{3}\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\sec^3\theta\right)\mathrm{d}\theta - \frac{1}{81}\int \frac{\mathrm{d}}{\mathrm{d}\theta}(\sec\theta)\mathrm{d}\theta$$

The FTC then says

$$= \frac{1}{243}\sec^3\theta - \frac{1}{81}\sec\theta + C_1$$

Now we just need to rewrite in terms of our original variable:

$$\sec\theta = \sqrt{1+\tan^2\theta} = \sqrt{1+9x^2}$$

Putting this all together, we find

$$\begin{aligned}
\int \frac{x^3}{\sqrt{1+9x^2}}\mathrm{d}x &= \frac{1}{243}\left((1+9x^2)^{3/2}\right) - \frac{1}{81}\left(\sqrt{1+9x^2}\right) + C \\
&= \frac{1}{243}\sqrt{1+9x^2}\left(1+9x^2-\frac{1}{3}\right) + C \\
&= \frac{1}{243}\sqrt{1+9x^2}\left(9x^2+1-3\right) + C \\
&= \frac{1}{243}\sqrt{1+9x^2}\left(9x^2-2\right) + C
\end{aligned}$$

Example 36.

$$\int \frac{x}{\sqrt{2x^2-4x-7}}\mathrm{d}x$$

This doesn't immediately look like something that would lend itself to a trig substitution, but in fact it does. To see this, we need to first complete the square on our denominator:

$$2x^2-4x-7 = 2(x-1)^2-9$$

and so our integral becomes

$$\int \frac{x}{\sqrt{2(x-1)^2-9}}$$

Now this looks like a more familiar form. However, despite looking suspiciously like a trigonometric integral, it doesn't look as if a sine or tan substitution will work. It turns out that the best substitution makes use of $\sec^2\theta - 1 = \tan^2\theta$ and so we let

$$(x-1) = \frac{3}{\sqrt{2}}\sec\theta$$

Then

$$2(x-1)^2-9 = 2\left(\frac{9}{2}\sec^2\theta\right) - 9 = 9\sec^2\theta - 9 = 9\tan^2\theta$$

Noting that

$$\frac{\mathrm{d}x}{\mathrm{d}\theta} = \frac{3}{\sqrt{2}}\sec\theta\tan\theta$$

We obtain

$$\begin{aligned}\int \frac{x}{\sqrt{2x^2-4x-7}}\mathrm{d}x &= \int \frac{\left(\frac{3}{\sqrt{2}}\sec\theta+1\right)\frac{3}{\sqrt{2}}\sec\theta\tan\theta}{3\tan\theta}\mathrm{d}\theta \\ &= \frac{3}{2}\int\sec^2\theta\mathrm{d}\theta + \frac{\sqrt{2}}{2}\int\sec\theta\mathrm{d}\theta\end{aligned}$$

To integrate $\sec\theta$, we can either use a common trick of multiplying by $\frac{\sec\theta+\tan\theta}{\sec\theta+\tan\theta}$, or we can use our brains a little. Let's push forward with the latter. To do so, let's rewrite our integral:

$$\int\sec\theta\mathrm{d}\theta = \int\frac{1}{\cos\theta}\mathrm{d}\theta = \int\frac{\cos\theta}{\cos^2\theta}\mathrm{d}\theta = \int\frac{\cos\theta}{1-\sin^2\theta}\mathrm{d}\theta$$

To evaluate this, we make the substitution $y = \sin\theta$ to yield

$$= \int\frac{1}{1-y^2}\mathrm{d}y = \frac{1}{2}\int\left(\frac{1}{y+1} - \frac{1}{y-1}\right)\mathrm{d}y = \frac{1}{2}\left(\ln|y+1| - \ln|y-1|\right) + C_1 = \frac{1}{2}\ln\left|\frac{\sin\theta+1}{\sin\theta-1}\right| + C_1$$

Thus, we obtain

$$\frac{3}{2}\int\sec^2\theta\mathrm{d}\theta + \frac{\sqrt{2}}{2}\int\sec\theta\mathrm{d}\theta = \frac{3}{2}\tan\theta + \frac{\sqrt{2}}{4}\ln\left|\frac{\sin\theta+1}{\sin\theta-1}\right| + C$$

Since

$$x-1 = \frac{3}{\sqrt{2}}\sec\theta \quad\Rightarrow\quad \cos\theta = \frac{3}{\sqrt{2}(x-1)} \quad\Rightarrow\quad \sin\theta = \frac{\sqrt{2x^2-4x-7}}{\sqrt{2}(x-1)} \quad\Rightarrow\quad \tan\theta = \frac{\sqrt{2x^2-4x-7}}{3}$$

we can plug it all in and (after a little bit of algebra) obtain:

$$\int\frac{x}{\sqrt{2x^2-4x-7}}\mathrm{d}x = \frac{\sqrt{2}}{2}\ln\left|\frac{\sqrt{2}(x-1)}{3} + \frac{\sqrt{2x^2-4x-7}}{3}\right| + \frac{\sqrt{2x^2-4x-7}}{2} + C$$

So, we have considered a fair few examples here of integrals that can be greatly simplified with a sensible trigonometric substitution. To pick the best substitution, we had to consider which identity the integrand most reminded us of.

We looked at integrals involving a substitution of $\sin\theta$, $\tan\theta$ and $\sec\theta$. It is worth pointing out that, should we feel so inclined, we can make a cosine substitution instead of a sine substitution in any of our integrals above. The reason for this is that we have symmetry between sine and cosine in the identity

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

and so there is no difference between letting $x = \sin\theta$ and $x = \cos\theta$.

To illustrate this, let's take our first example and compute it again using a cosine substitution:

Example 37.

$$\int \sqrt{1-x^2}\mathrm{d}x$$

This times, let $x = \cos\theta$ and obtain

$$\int \sqrt{1-x^2}\mathrm{d}x = \int \sqrt{1-\cos^2\theta}(-\sin\theta)\mathrm{d}\theta = -\int \sin^2\theta\mathrm{d}\theta$$

If we then make the substitution

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

we arrive at

$$= \frac{1}{2}\int (\cos 2\theta - 1)\,\mathrm{d}\theta = \frac{1}{4}\sin 2\theta - \frac{1}{2}\theta + C$$

Now we change back to the variable x:

$$\frac{1}{4}\sin 2\theta - \frac{1}{4}\theta + C = \frac{1}{2}x\sqrt{1-x^2} - \frac{1}{2}\arccos x + C$$

This is close to, but not the same as what we obtained with the sine substitution. When we let $x = \sin\theta$, we ended up with

$$\int \sqrt{1-x^2}\mathrm{d}x = \frac{1}{2}x\sqrt{1-x^2} + \frac{1}{2}\arcsin x + C$$

However, all we need to notice is that if we let $\sin\theta = x$, we have

$$x = \sin\theta = \cos\left(\frac{\pi}{2} - \theta\right) \quad \Rightarrow \quad \arccos x = \frac{\pi}{2} - \theta = \frac{\pi}{2} - \arcsin x$$

and so

$$\arcsin x + \arccos x = \frac{\pi}{2}$$

Thus we can say

$$\frac{1}{2}x\sqrt{1-x^2} - \frac{1}{2}\arccos x = \frac{1}{2}x\sqrt{1-x^2} + \frac{1}{2}\arcsin x + \underbrace{C'}_{=\frac{\pi}{4}+C}$$

which is identical to the previous result up to an arbitrary constant. So we see that we have the freedom to choose a sine or cosine substitution.

Let's summarise our choices:

Form	**Looks like**	**Substitution**
$\sqrt{a^2-b^2x^2}$	$1-\sin^2\theta = \cos^2\theta$	$x = \frac{a}{b}\sin\theta$ or $x = \frac{a}{b}\cos\theta$
$\sqrt{a^2+b^2x^2}$	$\tan^2\theta + 1 = \sec^2\theta$	$x = \frac{a}{b}\tan\theta$
$\sqrt{b^2x^2-a^2}$	$\sec^2\theta - 1 = \tan^2\theta$	$x = \frac{a}{b}\sec\theta$

10 Integration by parts

10.1 The theorem

Theorem 4 (Integration by Parts)**.** *Let $u, v : [a, b] \to \mathbb{R}$ be integrable such that everything below exists. Then*

$$\int_a^b u(x)v'(x)\mathrm{d}x = u(b)v(b) - u(a)v(a) - \int_a^b u'(x)v(x)\mathrm{d}x$$

Proof. Notice that

$$u'(x)v(x) + u(x)v'(x) = (u(x)v(x))'$$

Thus

$$\int_a^b (u'(x)v(x) + u(x)v'(x))\,\mathrm{d}x = \int_a^b (u(x)v(x))'\mathrm{d}x$$

By the FTC, we have

$$\int_a^b (u(x)v(x))'\mathrm{d}x = u(b)v(b) - u(a)v(a)$$

and a rearrangement gives us the desired result. □

10.2 Some examples

The hardest part of integrating by parts is selecting the correct u and v'. It is probable that we will get this wrong from time to time, but that's not a problem. We will know that we've made the right choice if we end up with an integral we can compute!

As a rule of thumb, we want to choose u to be the function that simplifies after differentiation. In a similar vein, we pick v' to be the function that changes the least when integrated. A general chain of preference for v' would be e^x, then $\sin x$ or $\cos x$, then anything else. If you have either an exponential or those trig functions, they will probably be the sensible choice for v'.

Example 38.

$$\int 3x\sin(1-x)\mathrm{d}x$$

Here we have an integrand containing the product of two functions, neither of which is the derivative of the other. This should be an indicator that integration by parts might be a sensible approach.

We have a sine term, and so we will choose $v' = \sin(1-x)$. We should be comforted by the subsequent choice of $u = 3x$ as this polynomial term will differentiate into a more accessible constant term. Let's crack on:

$$\begin{aligned} u &= 3x & v' &= \sin(1-x) \\ u' &= 3 & v &= \cos(1-x) \end{aligned}$$

So,

$$\int 3x\sin(1-x)\mathrm{d}x = 3x\cos(1-x) - \int 3\cos(1-x)\mathrm{d}x$$

We can then integrate this last term to obtain:

$$\int 3x\sin(1-x)\mathrm{d}x = 3x\cos(1-x) + 3\sin(1-x) + C$$

Example 39.

$$\int_1^2 xe^{2x}\mathrm{d}x$$

In this example, we have limits on our integral and so need to take these into account.

We have an exponential term, and so we choose $v' = e^{2x}$ (again, notice the convenient choice of $u = x$). Then

$$u = x \qquad v' = e^{2x}$$
$$u' = 1 \qquad v = \frac{1}{2}e^{2x}$$

so,

$$\int_1^2 xe^{2x}\mathrm{d}x = \left[\frac{1}{2}xe^{2x}\right]_1^2 - \frac{1}{2}\int e^{2x}\mathrm{d}x$$

The last integral is easy to evaluate:

$$\int_1^2 xe^{2x}\mathrm{d}x = \left[\frac{1}{2}xe^{2x} - \frac{1}{4}e^{2x}\right]_1^2 = e^4 - \frac{1}{4}e^4 - \frac{1}{2}e^2 + \frac{1}{4}e^2 = \frac{3}{4}e^4 - \frac{1}{4}e^2$$

Example 40.

$$\int x^2 \sin(2x)\mathrm{d}x$$

This is a prime candidate for integration by parts. We choose $v' = \sin(2x)$ and $u = x^2$. We have

$$u = x^2 \qquad v' = \sin(2x)$$
$$u' = 2x \qquad v = -\frac{1}{2}\cos(2x)$$

Then

$$\int x^2 \sin(2x)\mathrm{d}x = -\frac{1}{2}x^2\cos(2x) + \int x\cos(2x)\mathrm{d}x$$

Now, take a look at the integral we have ended up with on the RHS. This contains the product of two terms and looks itself to be a prime candidate for integration by parts! This is not due to any error we have made: sometimes, we need to integrate by parts more than once.

When integrating by parts for the second time, it is important that we make the same choice for the u and v', else we will find ourselves simply reversing what we did the first time. As such, we now relabel

$$u = x \qquad v' = \cos(2x)$$
$$u' = 1 \qquad v = \frac{1}{2}\sin(2x)$$

Then

$$\int x^2\sin(2x) = -\frac{1}{2}x^2\cos(2x) + \frac{1}{2}x\sin(2x) - \frac{1}{2}\int \sin(2x)\mathrm{d}x = -\frac{1}{2}x^2\cos(2x) + \frac{1}{2}x\sin(2x) + \frac{1}{4}\cos(2x) + C$$

So we see that sometimes we need to integrate by parts more than once to

11 Some Classic Integration Examples

11.1 Integral of $\ln x$

What function differentiates to gives $\ln x$? More exactly, what is

$$\int \ln x \mathrm{d}x$$

To answer this, we need to be clever. Notice

$$\int \ln x \mathrm{d}x = \int 1 \times \ln x \mathrm{d}x$$

and so we will integrate by parts:

$$u = \ln x$$
$$v' = 1$$

Then

$$u' = \frac{1}{x}$$
$$v = x$$

So

$$\int \ln x \mathrm{d}x = x \ln x - \int \frac{1}{x} x \mathrm{d}x = x \ln x - x + C$$

11.2 A loopy example

Consider

$$\int e^x \sin x \mathrm{d}x$$

How would we approach this? A sensible approach would be to integrate by parts. When we integrate by parts, it is best to set u to be the function that changes most when differentiated: in this case we let $u = \sin x$. Then

$$\int e^x \sin x \mathrm{d}x = e^x \sin x - \int e^x \cos x \mathrm{d}x$$

But now we end up with another integral we can't directly solve. Notice this is of a very similar form to the integral above, and so we also approach this integral with integration by parts.
It is important that we are consistent with our choice of u, and so let $u = \cos x$. If instead we had changed our minds and let $u = e^x$, we would simply reverse what we have just done - try it!
Now,

$$\int e^x \cos x \mathrm{d}x = e^x \cos x + \int e^x \sin x \mathrm{d}x$$

So,

$$\int e^x \sin x \mathrm{d}x = e^x \sin x - e^x \cos x - \int e^x \sin x \mathrm{d}x$$

Take a close look at this equation. Notice that our integral occurs on both sides of the equation! We can thus add it to both sides to obtain

$$2 \int e^x \sin x \mathrm{d}x = e^x(\sin x - \cos x)$$

and hence

$$\int e^x \sin x \mathrm{d}x = \frac{1}{2} e^x(\sin x - \cos x) + C$$

12 First-Order Ordinary Differential Equations

12.1 Introduction

Simply put, a *differential equation* is an equation involving a function and its derivative. In this course, we concern ourselves with *ordinary differential equations (ODEs)*, which are differential equations that contain no partial derivatives.

The *order* of a differential equation is the highest derivative that the equation involves. For example, we may have an equation that involves $\frac{\mathrm{d}^2y}{\mathrm{d}x^2}$ and no higher derivatives. An ODE such as this would be called second-order.

In this course, we will only concern ourselves with first-order ODEs.

There are countless examples of when we might encounter an ODE in both pure and applied branches of mathematics. They're so widely applicable, that we can conjure up a physical example without much thought:

In a radioactive sample, isotope A decays at a rate proportional to the number a of remaining nuclei in the sample. Another way of writing this, is to say

$$\frac{\mathrm{d}a}{\mathrm{d}t} \propto -a$$

We can then write this as

$$\frac{\mathrm{d}a}{\mathrm{d}t} = -ka$$

for some positive constant k. This is an explicit example of a first-order ODE.

The most general first-order ODE takes the form

$$\frac{\mathrm{d}y}{\mathrm{d}x} = f(x, y)$$

where $y \equiv y(x)$. We further classify first-order ODEs according to the form of $f(x, y)$.

12.2 The exponential function

As you will know by now, the exponential function, e^x, can be defined as

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

However, we can give an entirely equivalent way of defining the exponential function in terms of its derivative:

The exponential function, e^x, is the unique function f satisfying $f'(x) = f(x)$ and $f(0) = 1$.

This definition starts to illuminate the importance of the exponential function. One reason why the applications of the exponential function are so far reaching, is that it is an *eigenfunction* of the differential operator. Explicitly, this says that e^x is a function whose functional form is unchanged by the operator. We do in fact know this already as we are aware that taking a derivative of this function equates to no more than scaling by a constant:

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(e^{mx}\right) = me^{mx}$$

It is precisely this property that we will be utilising endlessly in our study of differential equations.

Consider the differential equation

$$3\frac{\mathrm{d}y}{\mathrm{d}x} - 2y = 0$$

How might we go about solving this? What we do is we utilise the properties of the exponential function by writing $y = e^{mx}$ for some $m \in \mathbb{R}$. Upon doing this, we substitute back into the ODE to obtain

$$3me^{mx} - 2e^{mx} = 0$$

or

$$e^{mx}(3m - 2) = 0$$

Now, since we have $e^{mx} > 0$ for all $x \in \mathbb{R}$, it follows that this equality is only satisfied if

$$3m - 2 = 0 \quad \Rightarrow \quad m = \frac{2}{3}$$

and so $y = e^{\frac{2}{3}x}$ is one solution to our ODE. We can easily verify this:

$$3\frac{\mathrm{d}}{\mathrm{d}x}\left(e^{\frac{2}{3}x}\right) - 2e^{\frac{2}{3}x} = e^{\frac{2}{3}x}\left(3\left(\frac{2}{3}\right) - 2\right) = 0$$

Notice, however, that we were careful to say that this is only one solution and not *the* solution. The reason for this is that we have made no specification of a *boundary condition* on our function y. We will discuss this in more depth shortly.

12.3 Seperable differential equations

A first-order ODE is called *separable* if it can be written in the form

$$F(y)\frac{\mathrm{d}y}{\mathrm{d}x} = G(x)$$

These are (by a long way) the nicest form we can hope a differential equation to be posed in. From this form, we can simply integrate both sides:

$$\int F(y)\frac{\mathrm{d}y}{\mathrm{d}x}\mathrm{d}x = \int G(x)\mathrm{d}x$$

Since $y \equiv y(x)$, we can make the substitution $u = y(x)$, and then

$$\mathrm{d}u = \frac{\mathrm{d}y}{\mathrm{d}x}\mathrm{d}x$$

so we find

$$\int F(u)\mathrm{d}u = \int G(x)\mathrm{d}x$$

which, depending on the functions F and G, will be very easy to solve!

We will look first at the most basic example we can think of:

Example 41. *Find y as an explicit function of x in the case*

$$\frac{\mathrm{d}y}{\mathrm{d}x} = xy$$

We can see here that upon dividing by x we obtain

$$\frac{1}{y}\frac{\mathrm{d}y}{\mathrm{d}x} = x$$

and so, upon comparing to the standard form above, we can see that $F(y) = \frac{1}{y}$ and $G(x) = x$, and so our ODE is separable.

We then integrate both sides with respect to x:

$$\int \frac{1}{y}\frac{\mathrm{d}y}{\mathrm{d}x}\mathrm{d}x = \int x\mathrm{d}x$$
$$\Rightarrow \quad \int \frac{1}{y}\mathrm{d}y = \int x\mathrm{d}x$$

From which we deduce

$$\ln y = \frac{1}{2}x^2 + C$$

for some $C \in \mathbb{R}$. Finally, raise e to the power of both sides to get

$$y = Ae^{\frac{1}{2}x^2}$$

where $A \in \mathbb{R}$.

Example 42. *Solve the following ODE for y*

$$x + \frac{\mathrm{d}y}{\mathrm{d}x} \sec x \sin y = 0$$

We look to show that this ODE is separable, and so we rearrange as

$$\sin y \frac{\mathrm{d}y}{\mathrm{d}x} = -x \cos x$$

and so $F(y) = \sin y$ and $G(x) = -x \cos x$. Now that we know it is separable, we know how to solve it. We proceed as before by integrating both sides with respect to x:

$$\int \sin y \frac{\mathrm{d}y}{\mathrm{d}x} \mathrm{d}x = \int \sin y \mathrm{d}y = -\int x \cos x \mathrm{d}x$$

The left-hand size integrates easily to $-\cos y + c_1$ and the right-hand size can be integrated by parts to give $-(x \sin x + \cos x) + c_2$. Equating gives

$$-\cos y + c_1 = -(x \sin x + \cos x) + c_2$$

and so we find

$$y = \arccos\left(x \sin x + \cos x + C\right)$$

12.4 Integrating factor

Consider a first order ODE of the form

$$\frac{\mathrm{d}y}{\mathrm{d}x} + p(x)y = q(x)$$

For an ODE of this form, we have a very clever technique at our disposal for solving it!

To solve an ODE of this form, we use something called an *integrating factor* to transform it into a total derivative. Let's see how this works.

Take the ODE above, but instead multiply through by the term

$$e^{\int^x p(t)\mathrm{d}t}$$

which we call an *integrating factor*. If we do this, we have:

$$e^{\int^x p(t)\mathrm{d}t} \frac{\mathrm{d}y}{\mathrm{d}x} + e^{\int^x p(t)\mathrm{d}t} p(x)y = e^{\int^x p(t)\mathrm{d}t} q(x)$$

Why did we do this? Consider evaluating the following:

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(e^{\int^x p(t)\mathrm{d}t} y\right)$$

By the product and chain rule, we find

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(e^{\int^x p(t)\mathrm{d}t} y\right) = \frac{\mathrm{d}}{\mathrm{d}x}\left(e^{\int^x p(t)\mathrm{d}t} y\right) y + \left(e^{\int^x p(t)\mathrm{d}t} y\right) \frac{\mathrm{d}y}{\mathrm{d}x}$$

From the fundamental theorem of caclulus, we know

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(e^{\int^x p(t)\mathrm{d}t} y\right) = \left(e^{\int^x p(t)\mathrm{d}t} y\right) p(x)$$

and thus

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(e^{\int^x p(t)\mathrm{d}t} y\right) = e^{\int^x p(t)\mathrm{d}t}\left(p(x)y + \frac{\mathrm{d}y}{\mathrm{d}x}\right)$$

Now, look back at our ODE multiplied by the integrating factor. The LHS is simply

$$e^{\int^x p(t)\mathrm{d}t} \frac{\mathrm{d}y}{\mathrm{d}x} + e^{\int^x p(t)\mathrm{d}t} p(x)y$$

which we can factor as

$$e^{\int^x p(t)\mathrm{d}t}\left(\frac{\mathrm{d}y}{\mathrm{d}x} + p(x)y\right)$$

But we notice that this is precisely equal to the expression we derived for the derivative of the product of our integrating factor and y. Explicitly, our ODE now becomes

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(e^{\int^x p(t)\mathrm{d}t}y\right) = e^{\int^x p(t)\mathrm{d}t}q(x)$$

In this form, we can solve our ODE with little effort. All we have to do is integrate both sides with respect to x and we will have an explicit expression for y!

Example 43. *Solve the following equation for y*

$$\frac{\mathrm{d}y}{\mathrm{d}x} + y = e^{-x}; \quad y(0) = 1$$

Upon comparing to the standard form

$$\frac{\mathrm{d}y}{\mathrm{d}x} + p(x)y = q(x)$$

we see that this ODE is suited perfectly for solution via an integrating factor, with $p(x) = 1$ and $q(x) = e^{-x}$.

Our integrating factor is then

$$e^{\int^x \mathrm{d}t} = e^x$$

and so, by the theory outlined above, we multiply and simplify to obtain

$$\begin{aligned} e^x\frac{\mathrm{d}y}{\mathrm{d}x} + e^x y &= 1 \\ \Rightarrow \quad \frac{\mathrm{d}}{\mathrm{d}x}\left(e^x y\right) &= 1 \end{aligned}$$

We then integrate both sides with respect to x:

$$\int \frac{\mathrm{d}}{\mathrm{d}x}(e^x y)\mathrm{d}x = \int \mathrm{d}x$$

and so

$$e^x y = x + C \quad \Rightarrow \quad y = xe^{-x} + Ce^{-x}$$

Using the initial condition $y(0) = 1$ gives

$$1 = C$$

and so

$$y = e^{-x}(x+1)$$

Example 44. *Solve the following for y*

$$x\frac{\mathrm{d}y}{\mathrm{d}x} - 2y = x^2$$

This isn't immediately an equation obviously suited for solution via an integrating factor. However, if we divide through by x, we see

$$\frac{\mathrm{d}y}{\mathrm{d}x} + \frac{-2}{x}y = x$$

and so $p(x) = -2/x$ and $q(x) = x$, and we can use the technique of an integrating factor. The integrating factor is

$$e^{\int^x \frac{-2}{t}\mathrm{d}t} = e^{-2\ln x} = \frac{1}{x^2}$$

As before, we multiply through by this factor and obtain

$$\frac{1}{x^2}\frac{\mathrm{d}y}{\mathrm{d}x} - \frac{2}{x^3}y = 1$$

which becomes

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{y}{x^2}\right) = \frac{1}{x}$$

Integrating and rearranging yields

$$y = x^2\ln x + Cx^2$$

12.5 Homogeneous differential equations

A first-order differential equation

$$\frac{\mathrm{d}y}{\mathrm{d}x} = f(x, y)$$

is said to be *homogeneous* if the function $f(x, y)$ satisfies the condition

$$f(tx, ty) = tf(x, y)$$

for all t. An entirely equivalent way of expressing an equation of this type is

$$\frac{\mathrm{d}y}{\mathrm{d}x} = f\left(\frac{y}{x}\right)$$

To see this equivalence, we simply let $t = \frac{1}{x}$ and thus obtain

$$f(x, y) = f(tx, ty) = f\left(1, \frac{y}{x}\right) = f\left(\frac{y}{x}\right)$$

To solve an equation such as this, we make a *substitution.* More explicitly, we write

$$v(x) = \frac{y(x)}{x}$$

where we make explicit the dependence of y on x. From here, we can rewrite the substitution as $y(x) = xv(x)$ and use the product rule to find

$$\frac{\mathrm{d}y}{\mathrm{d}x} = x\frac{\mathrm{d}v}{\mathrm{d}x} + v(x)$$

We can now rewrite our original ODE as

$$x\frac{\mathrm{d}v}{\mathrm{d}x} + v(x) = f(v(x))$$

Now we notice that this is in face a separable ODE. We can thus repose it in the form

$$\frac{\mathrm{d}v}{\mathrm{d}x} = \frac{f(v(x)) - v(x)}{x} \quad \Rightarrow \quad \int \frac{1}{f(v(x)) - v(x)}\mathrm{d}v(x) = \int \frac{1}{x}\mathrm{d}x$$

which, if f is a sufficiently nice function, will be relatively easy to solve!

Appendices

A Derivative of e^x

We begin by plugging $f(x) = e^x$ into our definition of a derivative by first principles:

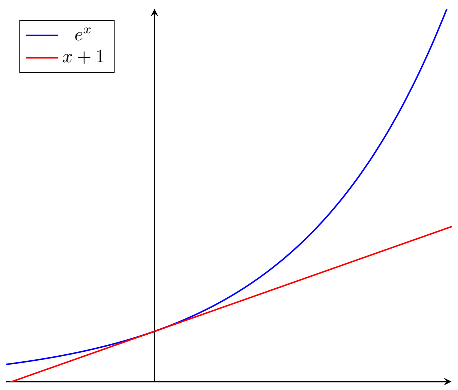
$$\begin{aligned}\lim_{h\to 0}\frac{e^{x+h}-e^x}{h} &= \lim_{h\to 0}\frac{e^x(e^h-1)}{h}\\ &= e^x\lim_{h\to 0}\frac{e^h-1}{h}\end{aligned}$$

Now we need to evaluate this limit. There are a few ways to do this, but we want to find a way to do so without calculus. A result we need is the definition of e^x:

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

Bernoulli's inequality (easily proved via induction - do this as an exercise) says that, for $x > -1$, we have

$$\left(1+\frac{x}{n}\right)^n \geq 1+x$$



for $n \in \mathbb{N}$ (i.e. $n = 0, 1, 2, 3, \cdots$). We can see this graphically in the sketch to the left.

If we take the limit as $n \to \infty$, we see the LHS becomes simply e^x and so it follows

$$e^x \geq 1+x$$

for $x > -1$. If we now let $x \to -x$ (we are going to be concerned with the limit as the exponent shrinks to 0, so when we take the negative we are still 'above' -1 and thus the inequality is valid), we deduce

$$e^{-x} \geq 1-x \quad\Rightarrow\quad e^x \leq \frac{1}{1-x}$$

and thus obtain the bounds

$$1+x \leq e^x \leq \frac{1}{1-x}$$

Now we subtract 1 from both sides to get

$$x \leq e^x - 1 \leq \frac{x}{1-x} \quad\Rightarrow\quad 1 \leq \frac{e^x-1}{x} \leq \frac{1}{1-x}$$

From which the sandwich theorem implies

$$\lim_{x\to 0}\frac{e^x-1}{x} = 1$$

So, finally, if we plug this into the original first principles formula:

$$\frac{\mathrm{d}e^x}{\mathrm{d}x} = e^x$$

B Derivative of $\log_a x$

First we use the log base-change formula:

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

to write

$$\log_a x = \frac{\ln x}{\ln a}$$

Now we derive the derivative of $\ln x$:

This follows from a combination of the above and implicit differentiation. We let $y = \ln x$, then $x = e^y$. If we act on both sides of this equation with the operator $\frac{\mathrm{d}}{\mathrm{d}x}$, we obtain

$$1 = \frac{\mathrm{d}}{\mathrm{d}x}(e^y) = e^y \frac{\mathrm{d}y}{\mathrm{d}x} \quad \Rightarrow \quad \frac{\mathrm{d}y}{\mathrm{d}x} = e^{-y} = \frac{1}{x}$$

and so

$$\frac{\mathrm{d}}{\mathrm{d}x}(\ln x) = \frac{1}{x}$$

Now that we have this, we can write

$$\frac{\mathrm{d}}{\mathrm{d}x}(\log_a x) = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\ln x}{\ln a}\right) = \frac{1}{\ln a}\frac{\mathrm{d}}{\mathrm{d}x}(\ln x) = \frac{1}{x \ln a}$$

C Derivative of a^x

This is again a matter of clever manipulation:

$$y = a^x \quad \Rightarrow \quad x = \log_a y$$

Now we act on both sides with $\frac{\mathrm{d}}{\mathrm{d}x}$ to get

$$1 = \frac{\mathrm{d}y}{\mathrm{d}x}\frac{1}{y \ln a}$$

and so

$$\frac{\mathrm{d}y}{\mathrm{d}x} = y \ln a = a^x \ln a$$

D Derivative of $\sin x$

We start with the usual formula for a derivative from first principles:

$$\frac{\mathrm{d}}{\mathrm{d}x}(\sin x) = \lim_{\Delta x \to 0} \frac{\sin(x + \Delta x) - \sin x}{\Delta x}$$

Now expand using the trigonometric addition formulae and simplify,

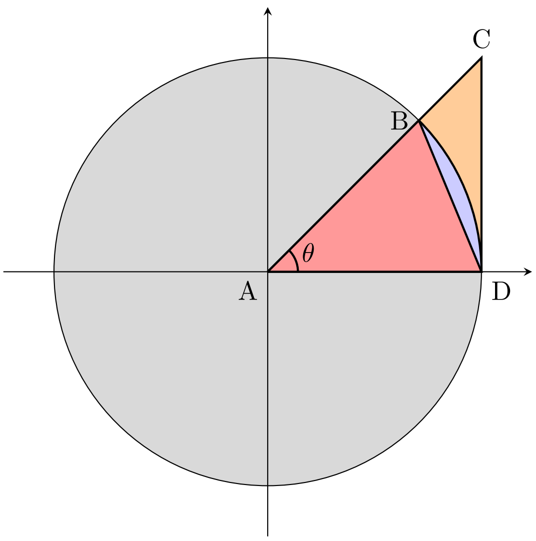
$$\begin{aligned}\lim_{\Delta x \to 0} \frac{\sin(x + \Delta x) - \sin x}{\Delta x} &= \lim_{\Delta x \to 0} \frac{\sin x \cos \Delta x + \sin \Delta x \cos x - \sin x}{\Delta x} \\ &= \sin x \left(\lim_{\Delta x \to 0} \frac{\cos \Delta x - 1}{\Delta x} \right) + \cos x \left(\lim_{\Delta x \to 0} \frac{\sin \Delta x}{\Delta x} \right)\end{aligned}$$

First we look to deal with the limit

$$\lim_{\Delta x \to 0} \frac{\sin \Delta x}{\Delta x}$$

and we do so via a geometric argument. Consider the diagram of the unit circle below. Since this is the unit circle, the area of the red triangle is easy enough to calculate:

$$\frac{1}{2} ab \sin = \frac{1}{2} \times 1 \times 1 \times \sin\theta = \frac{1}{2}\sin\theta$$



The area of the circle *sector* ABD is also easy to find:

$$\frac{\theta}{2\pi} \times \pi \times 1^2 = \frac{1}{2}\theta$$

What is the length of CD? Well, considering the triangle ACD, we see that CD $= \tan\theta$ and so, using $1/2 \times$ base $\times$ height, we find that the area of ACD is

$$\frac{1}{2} \times 1 \times \tan\theta = \frac{1}{2}\tan\theta$$

Now we use simple observations to bound these terms. From the diagram, it's simple to note that the area of triangle ABD is less than or equal to the sector ABD which is less than or equal to ACD, i.e.

$$\frac{1}{2}\sin\theta \leq \frac{1}{2}\theta \leq \frac{1}{2}\tan\theta = \frac{1}{2}\frac{\sin\theta}{\cos\theta}$$

WLOG we will deal with the case that we are in the upper-right quadrant and so $\sin\theta > 0$. The other cases are simply handled by placing absolute value signs around the terms. We multiply these inequalities through by 2 and take the reciprocal (which reverses the signs) to give

$$\begin{aligned} &\frac{\cos\theta}{\sin\theta} \leq \frac{1}{\theta} \leq \frac{1}{\sin\theta} \\ \Rightarrow \quad &\cos\theta \leq \frac{\sin\theta}{\theta} \leq 1 \end{aligned}$$

Now we take the limit as $\theta \to 0$ to deduce

$$\lim_{\theta \to 0} \frac{\sin\theta}{\theta} = 1$$

by the squeeze theorem.

Now for the second limit

$$\lim_{\theta \to 0} \frac{1 - \cos\theta}{\theta}$$

Multiply top and bottom by $(1 + \cos\theta)$ to get

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

and by our earlier result, we see

$$\lim_{\theta \to 0} \frac{1 - \cos\theta}{\theta} = 0$$

So, finally, combining this all together, we get

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}(\sin x) &= \sin x \underbrace{\left(\lim_{\Delta x \to 0} \frac{\cos \Delta x - 1}{\Delta x}\right)}_{=0} + \cos x \underbrace{\left(\lim_{\Delta x \to 0} \frac{\sin \Delta x}{\Delta x}\right)}_{=1} \\
&= \cos x
\end{aligned}$$

E Derivative of $\cos x$

We take a very similar approach here to that above:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}(\cos x) &= \lim_{h \to 0} \frac{\cos(x+h) - \cos x}{h} \\
&= \lim_{h \to 0} \frac{cosx \cos h - \sin x \sin h - \cos x}{h} \\
&= \cos x \lim_{h \to 0} \frac{\cos h - 1}{h} - \sin x \lim_{h \to 0} \frac{\sin h}{h} \\
&= -\sin x
\end{aligned}$$

by the arguments outlined above.

F Derivative of $\tan x$

This derivative follows simply from the derivatives of $\sin x$, $\cos x$ and the quotient rule:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}(\tan x) = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\sin x}{\cos x}\right) &= \frac{\frac{\mathrm{d}}{\mathrm{d}x}(\sin x)\cos x - \sin x \frac{\mathrm{d}}{\mathrm{d}x}(\cos x)}{\cos^2 x} \\
&= \frac{\cos^2 x - \sin x(-\sin x)}{\cos^2 x} \\
&= \frac{\sin^2 x + \cos^2 x}{\cos^2 x} \\
&= \sec^2 x
\end{aligned}$$

G Derivative of x^n

Starting in the usual way:

$$\frac{\mathrm{d}}{\mathrm{d}x}(x^n) = \lim_{\Delta x \to 0} \frac{(x + \Delta x)^n - x^n}{\Delta x}$$

We then binomially expand the numerator and simplify to get the required result,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}(x^n) &= \lim_{\Delta x \to 0} \frac{x^n + nx^{n-1}\Delta x + \frac{n(n-1)}{2!}x^{n-2}\Delta x^2 + \cdots + \Delta x^n - x^n}{\Delta x} \\
&= \lim_{\Delta x \to 0} \frac{nx^{n-1}\Delta x + \frac{n(n-1)}{2!}x^{n-2}\Delta x^2 + \cdots \Delta x^n}{\Delta x} \\
&= \lim_{\Delta x \to 0} \left(nx^{n-1} + \frac{n(n-1)}{2!}x^{n-2}\Delta x + \cdots + \Delta x^{n-1}\right) \\
&= nx^{n-1}
\end{aligned}$$

H The Chain Rule

Take two functions, $f(y)$ and $g(x)$. What would the derivative of $h(x) = f(g(x))$ be? To answer this question, we need to define two functions. By definition, the derivative of $f(y)$ is

$$f'(y) = \lim_{\Delta y \to 0} \frac{f(y + \Delta y) - f(y)}{\Delta y}$$

We then define a function $v(y)$

$$v := \frac{f(y + \Delta y) - f(y)}{\Delta y} - f'(y)$$

and note that

$$\lim_{\Delta y \to 0} v = 0$$

by definition. Similarly, we define

$$w := \frac{g(x + \Delta x) - g(x)}{\Delta x} - g'(x)$$

Now, notice that we can rearrange these expressions to get expressions for $f(y + \Delta y)$ or $g(x + \Delta x)$:

$$f(y + \Delta y) = f(y) + \Delta y(v + f'(y))$$

and a similar expression for $g(x + \Delta x)$. Taking $y = g(x)$ and $\Delta y = \Delta x(v + f'(y))$, we deduce

$$\begin{aligned} f(g(x + \Delta x)) &= f\left(g(x) + \Delta x(w + g'(x))\right) \\ &= f(g(x)) + \Delta x(w + g'(x))(v + f'(g(x))) \end{aligned}$$

Note how $\Delta y = \Delta a(v + f'(y))$ says $\Delta y \to 0$ as $\Delta x \to 0$, which in turn says $v \to 0$ as $\Delta x \to 0$.

By definition, the derivative of $f(g(x))$ is

$$\frac{\mathrm{d}}{\mathrm{d}x}(f(g(x))) = \lim_{\Delta x \to 0} \frac{f(g(x + \Delta x)) - f(g(x))}{\Delta x}$$

Now we substitute in what we derived above:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}x}(f(g(x))) &= \lim_{\Delta x \to 0} \frac{f(g(x)) + \Delta x(w + g'(x))(v + f'(g(x))) - f(g(x))}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{\Delta x(w + g'(x))(v + f'(g(x)))}{\Delta x} \\ &= \lim_{\Delta x \to 0} (w + g'(x))(v + f'g(x)) \end{aligned}$$

Using that $w \to 0$ as $\Delta x \to 0$ and we showed that we have $v \to 0$ as $\Delta x \to 0$, we arrive at the result:

$$\frac{\mathrm{d}}{\mathrm{d}x}(f(g(x))) = g'(x)f'(g(x))$$

I The Product Rule

Recall the definitions we made above in the section on proving the chain rule:

$$v := \frac{f(y + \Delta y) - f(y)}{\Delta y} - f'(y)$$

and

$$w := \frac{g(x + \Delta x) - g(x)}{\Delta x} - g'(x)$$

We want to find the derivative of $h(x) := f(x)g(x)$ and start with our first principles formula

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

Now plug in our formulae:

$$\begin{aligned} h'(x) &= \lim_{\Delta x \to 0} \frac{(f(x) + \Delta x(v + f'(x)))(g(x) + \Delta x(w + g'(x))) - f(x)g(x)}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{\Delta x f(x)(w + g'(x)) + \Delta x g(x)(v + f'(x)) + (\Delta x)^2(v + f'(x))(w + g'(x))}{\Delta x} \\ &= \lim_{\Delta x \to 0} (f(x)w + f(x)g'(x) + g(x)v + g(x)f'(x)) \end{aligned}$$

where, in moving to the last line, we lost all terms with a Δx still multiplying it. Recalling that $w, v \to 0$ as $\Delta x \to 0$, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}x}(f(x)g(x)) = f'(x)g(x) + f(x)g'(x)$$

J Proof of the Fundamental Theorem of Calculus

I'm going to divert slightly here into some more complicated mathematics. None of this is strictly relevant to your studies and is by no means examinable. Nonetheless, it may be of interest to some of you and, as such, it is included for completeness.

Theorem 5 (Rolle's Theorem)**.** *Let f be a continuous function on a closed interval $[a, b]$ and differentiable on (a, b). Suppose that $f(a) = f(b)$. Then there exists $x \in (a, b)$ such that $f'(x) = 0$.*

This is intuitively obvious: if you move up and down, and finally return to the same point, then you must have changed your direction at some point. In mathematical language: there must be a stationary point somewhere within the interval.

Proof. If f is constant then we are done.

If f is not constant, then there exists some u in our interval such that $f(u) \neq f(a)$. Without loss of generality, assume $f(u) > f(a)$. Since f is continuous, it has a maximum, and as $f(u) > f(a) = f(b)$, the maximum does not occur at $x = a$ or $x = b$.

Suppose that the maximum is attained at $x \in (a, b)$, then for any $h \neq 0$, we have

$$\frac{f(x+h) - f(x)}{h} \begin{cases} \leq 0 & h > 0 \\ \geq 0 & h < 0 \end{cases}$$

Taking the limit as $h \to 0$, we see that $f'(x) \geq 0$ and $f'(x) \leq 0$, and so $f'(x) = 0$. □

Corollary 2 (Mean Value Theorem)**.** *Let f be continuous on $[a, b]$ and differentiable on (a, b). Then there exists $x \in (a, b)$ such that*

$$f'(x) = \frac{f(b) - f(a)}{b - a}$$

Notice that this is the gradient of the straight line joining $f(a)$ and $f(b)$. So, what we are actually saying is that there exists some x in our interval such that the derivative at that point is equal to the tangent joining the two function values at the end points.

Proof. Let

$$g(x) = f(x) - \frac{f(b) - f(a)}{b - a} x$$

Then

$$g(b) - g(a) = f(b) - f(a) - \frac{f(b) - f(a)}{b - a}(b - a) = 0$$

Thus, by Rolle's theorem, there is some $x \in (a, b)$ such that $g'(x) = 0$. But

$$g'(x) = f'(x) = \frac{f(b) - f(a)}{b - a}$$

and the result follows. □

We state the Fundamental Theorem of Calculus as follows:

Theorem 6. *Let f be a continuous, real-valued function on the closed interval $[a, b]$. Let F be the function defined for all $x \in [a, b]$ by*

$$F(x) = \int_a^x f(t) \mathrm{d}t$$

Then $F(x)$ is uniformly continuous on the closed interval $[a, b]$ and differentiable on the open interval (a, b). In particular,

$$F'(x) = f(x)$$

A decent proof of this theorem requires familiarity with $\epsilon - \delta$ style proofs and so isn't worth noting down, but we put it here for the sake of completeness:

Proof.
$$\frac{F(x+h)-F(x)}{h} = \frac{1}{h}\int_x^{x+h} f(t)\mathrm{d}t$$
Let $\epsilon > 0$. Since f is continuous at x, there exists a $\delta > 0$ such that $|y-x| < \delta \Rightarrow |f(y)-f(x)| < \epsilon$. If $|h| < \delta$, then
$$\begin{aligned}\left|\frac{1}{h}\int_x^{x+h} f(t)\mathrm{d}t - f(x)\right| &= \left|\frac{1}{h}\int_x^{x+h}(f(t)-f(x))\mathrm{d}t\right| \\ &\leq \frac{1}{|h|}\left|\int_x^{x+h}|f(t)-f(x)|\mathrm{d}t\right| \\ &\leq \frac{\epsilon|h|}{|h|} \\ &= \epsilon\end{aligned}$$

□

There is a useful corollary of the FTC:

Corollary 3. *If f is continuously differentiable on $[a,b]$, then*
$$\int_a^b f'(t)\mathrm{d}t = f(b)-f(a)$$

Proof. We assume that the FTC holds. Define
$$g(x) = \int_a^x f'(t)\mathrm{d}t$$
Then, by the FTC, we have
$$g'(x) = f'(x) = \frac{\mathrm{d}}{\mathrm{d}x}(f(x)-f(a))$$
Since $g'(x)-f'(x)=0$, $g(x)-f(x)$ must be a constant function by the Mean Value Theorem. This says that
$$g(x) = f(x)+C$$
for some $C \in \mathbb{R}$. We find
$$g(a) = \int_a^a f'(t)\mathrm{d}t = 0$$
and so it follows $C = -f(a)$. It follows
$$g(x) = f(x)-f(a)$$
for every $x \in [a,b]$. In particular, this holds for $x=b$ and so
$$g(b) = \int_a^b f'(t)\mathrm{d}t = f(b)-f(a)$$

□