Things to be careful of

Do not cancel zeroes, for example 0*3=0*5 does not imply 3=5. If cancelling any factor, justify that it is not zero. If you have ac=bc and cancel the c factor, what you should do is deal with the c=0 case separately, ie ac=bc implies that either a=b OR c=0.

Do not cancel squares and square roots when the square is inside the square root unless you justify that the thing being squared inside the square root is a non-negative real number. For example, $\sqrt{(-3)^2} = \sqrt{9} = 3$ which is not equal to -3. This is because square root is defined the positive square root. If it were a proper inverse of the square function, it would have to take two values (for example $\sqrt{9} = 3, -3$) and then it would not be a function, by definition.

If you square both sides of anything, you must either be sure that it is not the case that one side is non-zero but minus the other side (Since then they could be non-equal but have equal squares, both sides being non-negative reals would suffice), or check after for extraneous solutions, for example,

$$x - 1 = 3$$

(Clearly x=4, but for the sake of example I will square both sides to show what goes wrong)

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

$$x^2 - 2x + 1 = 9$$

$$x^2 - 2x - 8 = 0$$

$$(x-4)(x+2) = 0$$

So the extraneous solution x = -2 arises. It is true when squaring both sides that each step is implied by the previous, but that does not mean that each step is equivalent to the previous.

The reason is essentially that more things can be squared to give 9 than just 3. If a function is one-to-one from the domain to the range, you may apply it to both sides or cancel it from both sides, but otherwise it requires justification.

Just like how squaring both sides can cause you to gain solutions, the opposite is true. For example, x=-3 is a solution of

$$x^2 = 3^2$$

But cancelling the squares would give x=3. When cancelling squares on both sides or cancelling squares on one side or square rooting the other, you should put a $\pm$ on one of the sides to ensure no loss of information.

Also, note that trigonometric functions are one-to-many over the real numbers, so similar problems arise. If y=arcsin(x), it is true that x=sin(y) meaning x=sin(arcsin(x)) when x is from -1 to 1 inclusive. However, if y=sin(x) it is not necessarily the case that x=arcsin(y), because consider what happens if x= π: arcsin is defined to take the value from $-\frac{\pi}{2}$ to $\frac{\pi}{2}$, so arcsin(sin(π))=0 which is a counterexample, kind of like how the square root of a negative number squared is a positive number since square root is defined to output positive values.

I guess my point is you need to carefully justify cancelling functions that you think are inverses of each other or applying or cancelling one-to-many functions to both sides of an equation.

Cosine rule (GCSE)

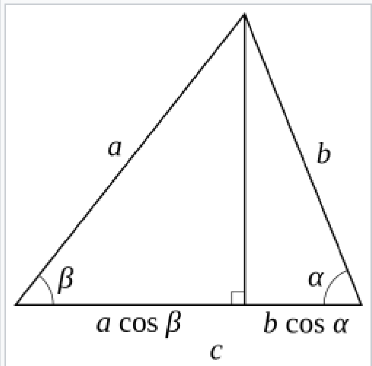


Fig. 5 – An acute triangle with perpendicular

$$c = a\cos\beta + b\cos\alpha.$$

(This is still true if α or β is obtuse, in which case the perpendicular falls outside the triangle.) Multiplying both sides by c yields

$$c^2 = ac\cos\beta + bc\cos\alpha.$$

The same steps work just as well when treating either of the other sides as the base of the triangle:

$$a^2 = ac\cos\beta + ab\cos\gamma,$$
$$b^2 = bc\cos\alpha + ab\cos\gamma.$$

Taking the equation for c^2 and subtracting the equations for b^2 and a^2,

$$c^2 - a^2 - b^2 = \cancel{ac\cos\beta} + \cancel{bc\cos\alpha} - \cancel{ac\cos\beta} - \cancel{bc\cos\alpha} - 2ab\cos\gamma$$
$$c^2 = a^2 + b^2 - 2ab\cos\gamma.$$

This proof includes an image to explain why it works using basic geometric identities.

Prime factorization (GCSE)

Clearly for every number a factorization into primes exists – I can take any number and make a factorization by repeatedly factoring its factors until I can't anymore, which will be when the factors are prime. Now I will prove that this can be done in a unique way. I use the fact that clearly 1 and 2 have a unique factorization (1 factors as no primes and 2 factors as just 2) and I show that all numbers have a unique factorization by showing that if all numbers 2,3,4,…,n have a unique factorization so does n+1, as then 2 having one implies 3 does, 2 and 3 having one implies 4 does, etc. This type of logic is known as induction.

2. Now we show that Property 2 also holds. Suppose that each of the numbers $2, \ldots, n$ has a *unique* prime factorisation. We must show that so does $n+1$. Suppose

$$n + 1 = p_1 \ldots p_s = q_1 \ldots q_t \tag{8}$$

where each of the p_i and q_j is prime. By reordering each side, we may assume that $p_1 \leq p_2 \leq \cdots \leq p_s$ and $q_1 \leq q_2 \leq \cdots q_t$. We consider two cases:

- $p_1 = q_1$, and
- $p_1 \neq q_1$

In the first case, dividing the equality $p_1 \cdots p_s = q_1 \cdots q_t$ by p_1, we deduce that $p_2 \cdots p_s = q_2 \ldots q_t$. As $p_1 > 1$, $p_2 \cdots p_t < n + 1$ and so must have a unique prime factorisation, by our supposition. Therefore $s = t$ and, since we have written the primes in increasing order, $p_2 = q_2, \ldots, p_s = q_s$. Since also $p_1 = q_1$, the two factorisations of $n+1$ coincide, and $n+1$ has a unique prime factorisation.

In the second case, suppose $p_1 < q_1$. Then

$$\begin{aligned} p_1(p_2 \cdots p_s - q_2 \cdots q_t) &= \\ &= p_1 p_2 \cdots p_s - p_1 q_2 \cdots q_t = \\ &= q_1 \cdots q_t - p_1 q_2 \cdots q_t \\ &= (q_1 - p_1) q_2 \cdots q_t \end{aligned} \tag{9}$$

Let $r_1 \ldots r_u$ be a prime factorisation of $q_1 - p_1$; putting this together with the prime factorisation $q_2 \ldots q_t$ gives a prime factorisation of the right hand side of (9) and therefore of its left hand side, $p_1(p_2 \ldots p_s - q_2 \ldots q_t)$. As this number is less than $n+1$, its prime factorisation is unique (up to order), by our inductive assumption. It is clear that p_1 is one of its prime factors; hence p_1 must be either a prime factor of $q_1 - p_1$ or of $q_2 \ldots q_t$. Clearly p_1 is not equal to any of the q_j, because $p_1 < q_1 \leq q_2 \leq \cdots \leq q_t$. So it can't be a prime factor

of $q_2 \ldots q_t$, again by uniqueness of prime factorisations of $q_2 \ldots q_t$, which we are allowed to assume because $q_2 \ldots q_t < n+1$. So it must be a prime factor of $q_1 - p_1$. But this means that p_1 divides q_1. This is absurd: q_1 is a prime and not equal to p_1. This absurdity leads us to conclude that p_1 it cannot be less than q_1.

The same argument, with the roles of the p's and q's reversed, shows that it is also impossible to have $p_1 > q_1$. The proof is complete. $\square$

Sum of polygon angles (convex case) (GCSE)

Technically the proof I will give below for the concave case is sufficient for all cases once we have the well known result for triangles, but I will do this proof for the convex case first since it's nicer.

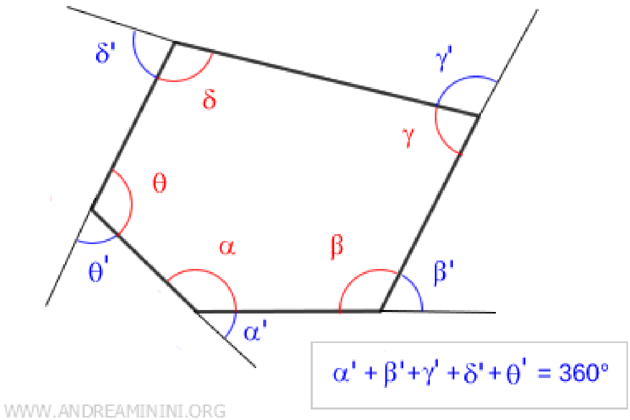


Image: Shows a diagram to demonstrate why internal angles + external angles = 180 * the number of sides, and why the external angles of a convex polygon sum to 360.

We note that the total of the red and blue is clearly 180 degrees times the number of verticies or sides, and that he sum of the blue angles is clearly 360 degrees as you could imagine dragging α' over to β' then those two together over to Y' and so on and then we will see that the total of the blues is clearly 360 degrees. The total of the reds is thus the difference between the total of the red and blue (180n) minus the total of the blue (360) which gives 180n-360 as required.

Sum of polygon angles (concave case) (GCSE)

We use the same principle of induction that we did for the prime factorization proof. We want to show that any n sided polygon can be triangulated into n-2 triangles like in the image below, where each vertex of each triangle is a vertex of the original polygon.

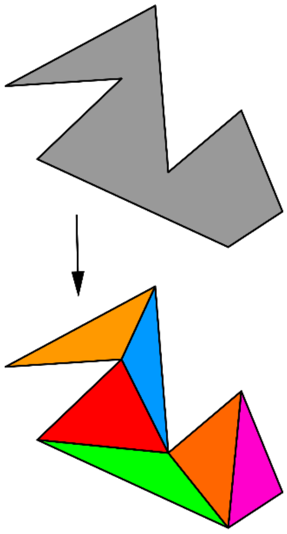


Image: Shows an example of triangulating a polygon.

And then we see that the sum of the interior angles becomes 180(n-2) as the sum of the interior angles of the original polygon must be equal to the sum of the interior angles of the n-2 triangles, which clearly is 180(n-2) by the above convex case as triangles are indeed convex. Now it remains to show that a triangulation exists. The approach will be to show that if a triangulation exists into a number of triangles equal to the number of sides minus 2 for any polygon with 3, 4, 5, … n-1 sides, then it works for n sides as well, as then we can say that well clearly it works for triangles so it must work for four sided shapes, but it works for triangles and four sided shapes so it must work for five sided shapes, and so on until we have it for all numbers of sides.

Now, suppose all polygons with fewer than n sides can be triangulated into the number of sides minus two triangles, and that we have a polygon with n sides. consider the leftmost vertex. Clearly, the angle at that vertex is less than 180 degrees, since otherwise it wouldn't be the leftmost vertex! (Just think about it hard enough to see this). Then consider the two neighbouring vertices and try to draw a line between them. Possibly this line does not intersect part of the polygon, in which case we form a triangle and have successfully split it into a triangle and another polygon with n-1 sides which by our assumption about any polygon with fewer than n sides can be triangulated into n-3 triangles, so our original polygon can be triangulated into n-2 triangles, so done. Otherwise, we connect the two neighbouring vertices to the leftmost one and part of the polygon is inside the triangle formed. In which case, we pick the leftmost vertex inside this triangle (meaning no part of the polygon is in any point in the triangle to the left) and connect it to the leftmost vertex, noting that now this diagonal cannot be obstructed, so we have split the polygon. Suppose it has been split into one with x sides, then the other one has n+2-x sides (As the total number of sides of the two becomes n-2 as the diagonal which we used to cut used to contribute 0 sides and now contributes 2). By our assumption about polygons with fewer than n sides, we have that one part can be triangulated into x-2 triangles and the other part can be triangulated into n-x triangles, so the original polygon can be triangulated into n-2 triangles. So done.

Area of a circle (GCSE)

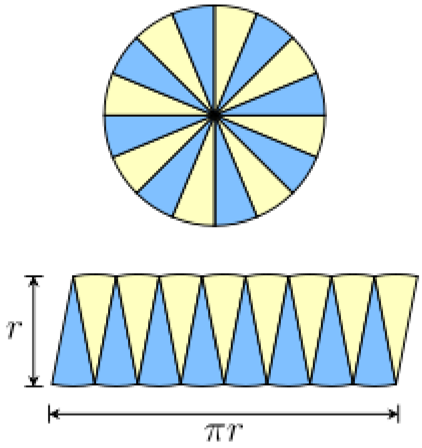


Image: Shows a circle cut into many slices and then below shows the slices rearranged to form a rectangle-looking shape of height r and width πr.

From this diagram it is clear that as we make the slices smaller the area of the figure on the bottom approaches a rectangle which will clearly have an area of πr^2.

Factor theorem

Given a polynomial f(x), x-a divides f(x) **if and only if** a is a root of f(x), ie f(a)=0

First, notice the wording "if and only if". This means that the statements are equivalent, which means one implies the other. Therefore, it is sufficient to show that x-a dividing f(x) implies f(a)=0 and that f(a)=0 implies that x-a divides f(x)

First, suppose x-a divides f(x). This means that there exists a polynomial g(x) such that f(x)=(x-a)g(x). Now evaluate f(a) by substituting a in place of x. This gives f(a)=(a-a)g(a). Since a-a=0, f(a)=0g(a)=0.

Now suppose that f(a)=0. We know that if we went through the long division process, we would get that f(x)=(x-a)g(x)+R for some polynomial g(a) and some remainder R. Since this is true for all values of x, it is true when x=a, so f(a)=(a-a)g(a)+R, but we know that (a-a)g(a)=0 since a-a=0 and that f(a)=0 by our assumption, so 0=0+R, therefore R=0. This means that f(x)=(x-a)g(x) so x-a divides f(x).~

Remainder theorem

If f(x) is a polynomial, and you do long division of f(x) by x-a then get f(x)=(x-a)g(x)+R for some polynomial g(x) so that R is the remainder, then R=f(a). More generally, if you divide f(x) by bx-a then the remainder is f($\frac{a}{b}$).

Proof: First, note that f(x)=(x-a)g(x)+R is true for all values of x, so substituting x=a gives f(a)=(a-a)g(a)+R. Since (a-a)g(a)=0 since a-a=0, f(a)=R.

If we divide f(x) by bx-a we will get f(x)=(bx-a)g(x)+R for some polynomial g. Let x=$\frac{a}{b}$ since this is true for all x, then we get f($\frac{a}{b}$) = (b($\frac{a}{b}$)-a)g($\frac{a}{b}$)+R = (a-a)g($\frac{a}{b}$)+R (since the b's cancel in the b($\frac{a}{b}$) term) = R, since again, (a-a) times anything is always 0.

Equating coefficients

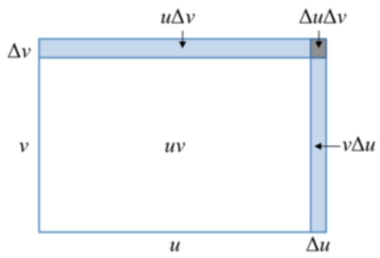
We can do this because if two polynomials with different coefficients were the same, their difference would be a polynomial with non-zero coefficients, which is clearly never 0, since if the x^k term were its smallest non-zero term, the kth derivative of the polynomial would be non-zero at x=0, contradicting the fact that the polynomial is zero. This holds for infinite polynomials too once we have the justification of differentiating those term by term (needed for generalized binomial theroem proof, added as an appendix).

Product rule

$$\frac{d}{dx}f(x)g(x)$$

$$\begin{aligned}
&= \lim_{\Delta x \to 0} \frac{f(x+\Delta x)g(x+\Delta x) - f(x)g(x)}{\Delta x} \\
&= \lim_{\Delta x \to 0} \frac{f(x+\Delta x)g(x+\Delta x) - f(x)g(x+\Delta x) + f(x)g(x+\Delta x) - f(x)g(x)}{\Delta x} \\
&= \lim_{\Delta x \to 0} \frac{[f(x+\Delta x) - f(x)] \cdot g(x+\Delta x) + f(x) \cdot [g(x+\Delta x) - g(x)]}{\Delta x} \\
&= \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x} \cdot \lim_{\Delta x \to 0} g(x+\Delta x) + \lim_{\Delta x \to 0} f(x) \cdot \lim_{\Delta x \to 0} \frac{g(x+\Delta x) - g(x)}{\Delta x} \\
&= f'(x)g(x) + f(x)g'(x).
\end{aligned}$$

The fact that $\lim_{\Delta x \to 0} g(x+\Delta x) = g(x)$ follows from the fact that differentiable functions are continuous.



Geometric illustration of a proof of the product rule[1]

This proof contains an image with a visual proof of the product rule.

In the geometric illustration above u is shorthand for f(x) and v is shorthand for g(x).

Chain rule

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

But this is actually not a proof because $f(x+h) - f(x)$ could be 0.

So, we define a function d(y) to be $\frac{g(y)-g(f(x))}{y-f(x)}$ if y does not equal f(x) and d(y)=g'(f(x)) if y=f(x). This function is continuous at y=f(x) since as y approaches f(x) $\frac{g(y)-g(f(x))}{y-f(x)}$ approaches g'(f(x)) by definition. If g(f(x+h)) does not equal g(f(x)) then we have that $\frac{g(f(x+h))-g(f(x))}{h} = \frac{g(f(x+h))-g(f(x))}{f(x+h)-f(x)} \frac{f(x+h)-f(x)}{h}$ which is the same as saying $\frac{g(f(x+h))-g(f(x))}{h} = d(f(x+h)) \frac{f(x+h)-f(x)}{h}$. Otherwise, we still have that $\frac{g(f(x+h))-g(f(x))}{h} = d(f(x+h)) \frac{f(x+h)-f(x)}{h}$, since f(x+h)=f(x) so we are just saying that 0=0. Therefore, the correct proof is $\frac{dg(f(x))}{dx} = \lim_{h \to 0} d(f(x+h)) \frac{f(x+h)-f(x)}{h}$. But since d(f(x+h)) approaches g'(f(x)) as h goes to 0 by continuity of d, and $\frac{f(x+h)-f(x)}{h}$ approaches f'(x) by definition, the result follows.

Quotient rule

This follows from the product and chain rules.

$\frac{d}{dx}\frac{f(x)}{g(x)} = \frac{d}{dx}\left(f(x) * \frac{1}{g(x)}\right) = f'(x) * \frac{1}{g(x)} + f(x) * \frac{d}{dx}(\frac{1}{g(x)})$

Where I have used the product rule to get the second equality.

We can use the chain rule to find $\frac{d}{dx}\left(\frac{1}{g(x)}\right)$. The chain rule says that for any 2 functions g(x), h(x), the derivative of h(g(x)) with respect to x is given by g'(x)h'(g(x)). In this case, if h(x) is the function $\frac{1}{x}$ then $\frac{1}{g(x)}$=h(g(x)). Now h'(x) = $\frac{d}{dx}\left(\frac{1}{x}\right) = \frac{-1}{x^2}$ by the rule above so

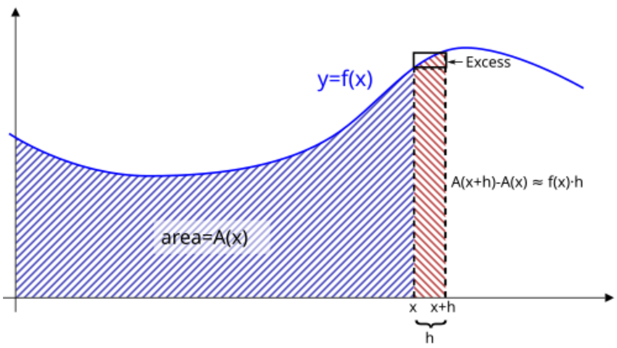
h'(g(x)) = $\frac{-1}{g(x)^2}$. This means that $\frac{d}{dx}\left(\frac{1}{g(x)}\right) = g'(x) * \frac{-1}{g(x)^2}$.

Now finally, $\frac{d}{dx}\frac{f(x)}{g(x)}$=$f'(x) * \frac{1}{g(x)} + f(x) * \ g'(x) * \frac{-1}{g(x)^2}$ = $\frac{g(x)f'(x)-f(x)g'(x)}{g(x)^2}$.

Integral well defined-ness

We are given that if F(x) is an antiderivative of f(x) then $\int_a^b f(x)dx$ is defined as F(b)-F(a). We will shortly prove that this is equal to the area under f(x) from a to b. However, first we need to show that this is well defined, since F(x) is not *the* antiderivative of f, rather *an* antiderivative. The family of antiderivatives of f(x) is given by F(x)+c. But, notice that in (F(b)+c)-(F(a)+c) the c's cancel, so as long as we evaluate the difference between an antiderivative when evaluated at b and a, we will get the same value regardless of which antiderivative we use.

Fundamental theorem of calculus



There is another way to *estimate* the area of this same strip. As shown in the accompanying figure, h is multiplied by $f(x)$ to find the area of a rectangle that is approximately the same size as this strip. So:

$$A(x+h) - A(x) \approx f(x) \cdot h$$

Dividing by h on both sides, we get:

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

This estimate becomes a perfect equality when h approaches 0:

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

Image to show why FTC holds intuitively.

Note: A(x) is the area from a starting point, it doesn't matter which starting point we pick, but it should not be negative infinity like my A level textbook does, since otherwise A(x) would not always be well defined!

Now one alternative way of stating the fundamental theorem of calculus is to say that f(x) is the derivative of $\int_a^x f(t)dt$ which is equal to F(x)-F(a) where F(x) is an antiderivative of f(x). Since F(a) is a constant, we have that the derivative of F(x)-F(a) equals f(x).

Existence and uniqueness of e

Specifically, to prove that a number e with the property that the derivative of e^x equals e^x and that e is the unique non-zero real number with this property. We work in the reals for now since for complex numbers exponentiation is more complicated

Let's try to find the derivative of c^x for some arbitrary c real and > 1.

$\lim_{h\to 0}\frac{c^{x+h}-c^x}{h}$ = $c^x \lim_{h\to 0}\frac{c^h-1}{h}$

For now, we assume the limit exists. It is not 0 since $c^x$is not constant but its derivative would be 0 if $\lim_{h\to 0}\frac{c^h-1}{h}$ were 0. Let's call this limit z. Then the derivative of c^x is zc^x. Consider $(c^{\frac{1}{z}})^x$ which equals $c^{\frac{x}{z}}$ by laws of powers. The derivative of this, by the chain rule is $(\frac{x}{z})'(zc^{\frac{x}{z}}) = (\frac{1}{z})(zc^{\frac{x}{z}}) = c^{\frac{x}{z}}$. Therefore $c^{\frac{1}{z}}$ is a number with the property that e has, ie the derivative of $(c^{\frac{1}{z}})^x$ is $(c^{\frac{1}{z}})^x$.

Now suppose f also has the property that e has. Then $\frac{d}{dx}f^x = \frac{d}{dx}\left(e^{\ln(f)}\right)^x = \frac{d}{dx}e^{x\ln(f)}$

$= e^{\text{xln}(f)} \frac{d}{dx} xln(f)$ =$e^{\text{xln}(f)} ln(f) = f^x \ln(f)$.

We see that by our assumption that f has the property that e has that f^x equals its own derivative, we must have ln(f)=1, therefore f=e, so e is unique.

For those of you who want to be particularly rigorous, to show that this limit exists, we will first note that if h is positive then $\lim_{h\to 0} \frac{c^h-1}{h}$ is always positive since the numerator and denominator are both positive as c^h>c^0=1 is implied by h>0 and the fact that clearly c^hincreases as h increases. We will show that $\frac{c^h-1}{h}$ is increasing over the rational numbers. A function that is increasing over the rationals and continuous (which this clearly is) intuitively (and provably) can't possibly be not increasing over the reals since the rationals are dense (ie there are rationals arbitrarily close/as close as we want to any number), and a function that is increasing and has positive outputs for positive inputs cannot possibly not have a limit as its input goes to 0 from the right, since the set of values $\frac{c^h-1}{h}$ for positive h is bounded below by 0, then the fact that it has a highest lower bound is an axiom which is essentially a reverse of the least upper bound property.

Now what we need to actually show is that $\frac{c^a-1}{a} > \frac{c^b-1}{b}$ if a and b are rational with a>b. Suppose k is a common denominator of a and b so a=n/k and b=m/k for some integers m and n. Let x=1/k so a=nx and b=mx. Then we have to prove $\frac{c^{nx}-1}{nx} < \frac{c^{mx}-1}{mx}$. We can expand these as follows:

$$\frac{c^{nx}-1}{nx} = \frac{c^x-1}{x} * \frac{1 + c^x + \cdots c^{x(n-1)}}{n}$$

$$\frac{c^{mx}-1}{mx} = \frac{c^x-1}{x} * \frac{1 + c^x + \cdots c^{x(m-1)}}{m}$$

and it is easy to check that the product of the numerators and denominators match the numerator and denominator of the original expression. We see that the second expression is larger since both are the same constant $\frac{c^x-1}{x}$ multiplied by something. In the first expression, this something is the mean of $1, c^x, \ldots c^{x(n-1)}$ and the second one is the mean of those terms with some terms that are strictly larger than them, ie $c^{xn}, \ldots c^{x(m-1)}$, so it is larger than the first mean, completing the proof of existence of the limit.

Power rule (general case)

The derivative of $x^a$with respect to x is given by $ax^{a-1}$for all a, provided a is a constant, so please do not use this to try to find something like $\frac{d}{dx}(x^x)$

$\frac{d}{dx}(x^a)$ = $\frac{d}{dx}\left((e^{\ln(x)})^a\right) = \frac{d}{dx}\left(e^{a*\ln(x)}\right)$

By the chain rule, we have that the above expression is equal to

$\left(e^{a*\ln(x)}\right)\frac{d}{dx}(a * \ln(x))$

To see this, recall that the chain rule says that for any 2 functions g(x), h(x), the derivative of h(g(x)) with respect to x is given by g'(x)h'(g(x)). Let h(x)=e^x and g(x)=a*ln(x) and the expression above follows. Here we use the fact that a is a constant to find that $\frac{d}{dx}(a * \ln(x)) = \frac{a}{x}$

Therefore,

$\frac{d}{dx}(x^a) = \left(e^{a*\ln(x)}\right) * \frac{a}{x} = \left(e^{\ln(x)}\right)^a * \frac{a}{x} = x^a * \frac{a}{x} = ax^{a-1}$

Log derivative or 1/x antiderivative

One may try to integrate 1/x using the power rule. Here is why that goes wrong.

$\int \frac{1}{x} dx = \int x^{-1} dx = \frac{x^0}{0} + c$

Division by zero. The way I will approach this is by trying to differentiate ln(x) and showing that the derivative is equal to 1/x. Suppose $y = \ln(x)$, then

$$e^y = x$$

$$e^y = \frac{dx}{dy}$$

$$x = \frac{dx}{dy}$$

$$\frac{1}{x} = \frac{dy}{dx}$$

Note: $\frac{dx}{dy}\frac{dy}{dx} = 1$ by the chain rule.

The proper way to do arcsin derivative

Here we assume x is real and $|x| \leq 1$, as that is the domain that arcsin is typically defined.

$$y = \arcsin(x)$$

$$\sin(y) = x$$

Although sin(y)=x is implied by y=arcsin(x), the converse is not true, as discussed

$$\cos(y) = \frac{dx}{dy}$$

Here, many people assert that $\cos(y) = \sqrt{cos^2(y)} = \sqrt{1 - sin^2(y)}$, however the assertion $\cos(y) = \sqrt{cos^2(y)}$ is only true if cos(y) is positive. Luckily it is, since the range of arcsin is by definition $-\frac{\pi}{2}$ to $\frac{\pi}{2}$ where cos is always positive, but usually this step is not done properly. The result for the derivative of arcsin follows, and the derivative of arccos can be derived in the same way, noting that sin is positive and -sin is negative in the range of arccos.

Separation of variables

Suppose we know that $\frac{dy}{dx} = \frac{f(x)}{g(y)}$, we want to justify that $\int f(x)dx = \int g(y)dy$ beyond notational tricks. I'm not sure how rigorous this is but I'm sure it's good enough. Note that here y is a function of x.

$\lim_{h \to 0} \frac{y(x+h)-y(x)}{h} = \frac{f(x)}{g(y)}$ by the definition of $\frac{dy}{dx}$. We have that

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

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

Taking an antiderivative of both sides is allowed as since you add the +c when you find the antiderivative so it's just saying that two equivalent functions have the same family of antiderivatives. We just need to justify cancelling the dx. The idea that you can cancel the dx's will be important for integration by substitution and working with integrals of anything defined parametrically.

First consider the following interpretation of the integral as a sum of rectangles

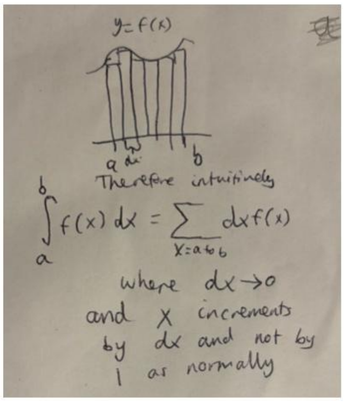
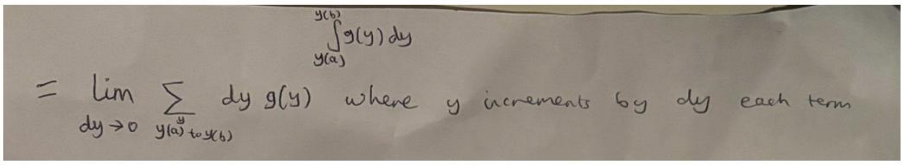


Image to show the interpretation of integration as the sum of rectangles.

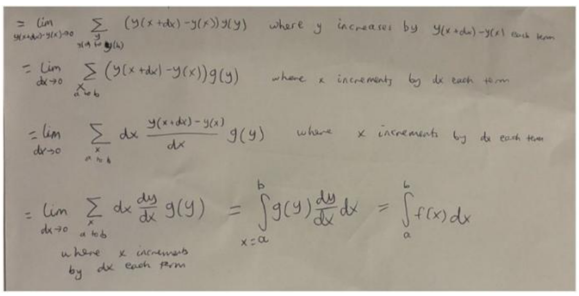
We see that the following are equivalent



$$\int_{y(a)}^{y(b)} g(y)\,dy$$

$$= \lim_{dy \to 0} \sum_{\substack{y \\ y(a) \text{ to } y(b)}} dy\, g(y) \quad \text{where } y \text{ increments by } dy \text{ each term}$$

Images: derivation that under this interpretation of the integral from above, we can work with differentials as expected, as long as we're careful.



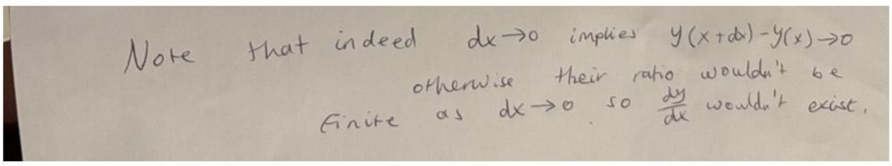
$$= \lim_{y(x+dx)-y(x) \to 0} \sum_{\substack{y \\ y(a) \text{ to } y(b)}} (y(x+dx)-y(x))g(y) \quad \text{where } y \text{ increases by } y(x+dx)-y(x) \text{ each term}$$

$$= \lim_{dx \to 0} \sum_{\substack{x \\ a \text{ to } b}} (y(x+dx)-y(x))g(y) \quad \text{where } x \text{ increments by } dx \text{ each term}$$

$$= \lim_{dx \to 0} \sum_{\substack{x \\ a \text{ to } b}} dx \frac{y(x+dx)-y(x)}{dx} g(y) \quad \text{where } x \text{ increments by } dx \text{ each term}$$

$$= \lim_{dx \to 0} \sum_{\substack{x \\ a \text{ to } b}} dx \frac{dy}{dx} g(y) = \int_{x=a}^{b} g(y)\frac{dy}{dx}dx = \int_a^b f(x)dx$$

where x increments by dx each term



Note that indeed $dx \to 0$ implies $y(x+dx)-y(x) \to 0$ otherwise their ratio wouldn't be finite as $dx \to 0$ so $\frac{dy}{dx}$ wouldn't exist.

So we have that $\int_a^b f(x)dx = \int_{y(a)}^{y(b)} g(y)dy$ for arbitrary a and b. If, say, a=0 then

$\int_0^b f(x)dx = \int_{y(0)}^{y(b)} g(y)dy$ so G(y(b))-G(0)=F(b)-F(0) where G and F are antiderivatives of g and f respectively. We then see that G(y(b)) differs by F(b) by a constant which depends only on which particular solution we consider, which is good enough.

Binomial expansion (positive integers)

We first motivate the definition of $\binom{n}{r}$. What is the number of ways to pick r things from n things? Well, we have n choices for the first thing, n-1 for the second, n-2 for the third, and so on. But, we could pick the same things in a different order, so to take account of that we need to divide by the number of orderings of r things, ie the number of ways we could list those same r things, which is r! (r factorial). This is because I have r choices for the first thing in the list, r-1 choices for the second since I can't use the one that was first in the list, r-2 for the third, etc.

Now it should make sense why these are the binomial coefficients – if I have say (1+x)(1+x)(1+x)…(1+x) with n (1+x)'s then the value of this is equal to the sum of products of one term per bracket, and the x^r coefficient is the number of such products that contain exactly r x's is $\binom{n}{r}$

We also have this identity which proves the additive property of the entries of pascals triangle if they were to be defined by these binomial coefficients.

$$\begin{aligned}\binom{n-1}{k}+\binom{n-1}{k-1} &= \frac{(n-1)!}{k!(n-1-k)!}+\frac{(n-1)!}{(k-1)!(n-k)!}\\ &= (n-1)!\left[\frac{n-k}{k!(n-k)!}+\frac{k}{k!(n-k)!}\right]\\ &= (n-1)!\frac{n}{k!(n-k)!}\\ &= \frac{n!}{k!(n-k)!}\\ &= \binom{n}{k}.\end{aligned}$$

An intuition for this is also, if I am picking k things from n, the number of ways is the sum of the number of ways to do it without picking the first thing, and the number of ways to do it with picking the first thing.

Geometric series valid interval

Before reading on, try using your calculator to calculate the sum of the first few terms of a geometric series with common ratio (i) 0.5 and (ii) 1.5. See how the values change as you add more terms. Can you figure out why we need that for the infinite sum the common ratio has to be between -1 and 1?

The partial sums of a geometric series are given as follows, where the first term is a and the common ratio is r and the sum of the first k terms is

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

Now note that as k goes to infinity, if |r|<1 then $r^k \to 0$ so the expression above actually approaches $\frac{a}{1-r}$ and conversely if it is not the case that |r|<1 then the expression above will blow up to infinity as k increases and not approach a limit. If you're not sure why, consider what happens if you repeatedly multiply by (i) something less than 1 and (ii) something more than 1, and (iii) the fact that if it's something negative you just have an alternating sign and who cares about that.

Integration by substitution

We prove the assertion below which essentially says that integration by substitution works as you expect.

$$\int_{x=u(a)}^{u(b)} f(u(x))\frac{du}{dx}dx = \int_{x=u(a)}^{u(b)} f(u(x))du = \int_{u=a}^{u=b} f(u(x))du$$

The justification is the same as what I did for the differential equations, where the steps here are justified the same way as certain steps in that picture of the paper I took because I was too lazy to type it out at 3 in the morning. (Hint: u here is like y there, f here is like g there).

Integration by parts

Product rule for differentiation says (u(x)v(x))'=u(x)v'(x)+u'(x)v(x) for functions u and v.

Now integrate both sides, or I like to call it finding the family of antiderivatives of both sides, giving

$$u(x)v(x) + c = \int u(x)v'(x)dx + \int v(x)u'(x)dx$$

Rearranging gives the integration by parts formula.

Volumes of revolution

Here you can see from the lazy quick sketch diagram below that the volume of revolution will approach as the width of these cylinders gets small the sum of the volumes of the cylinders which are each equal to $\pi y^2 dx$ and using the idea of integration as a sum from the separation of variables section we see that summing these volumes and taking a limit as dx → 0 is the same as the integral typically used which is $\int \pi y^2 dx$.

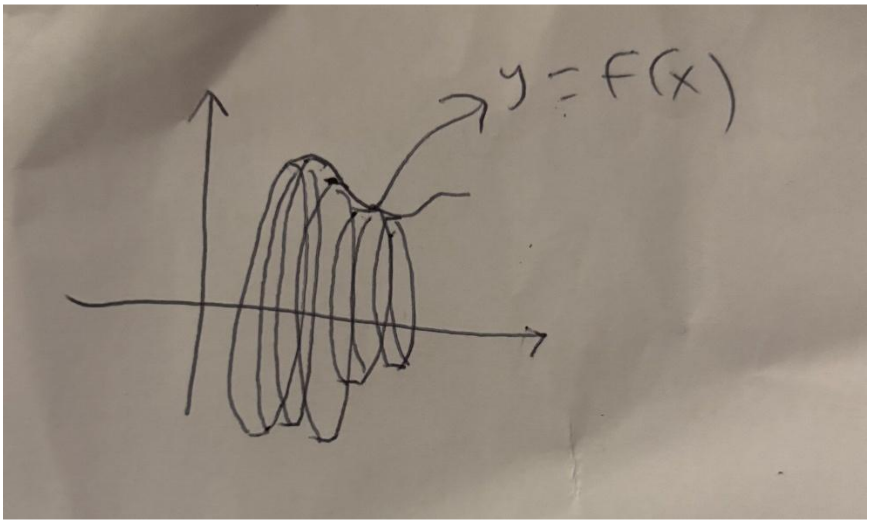


Image: Bad diagram of concentric thin cylinders to illustrate why the formula works.

Volume of cone (GCSE but proven using A level stuff)

Intuitively, the volume should scale with the height times the square of the radius. The volume of a cone with radius r and height h is equal to the volume bounded by the line segment connecting (0, r) to (h, 0) after rotating it 2π radians about the x axis, as that volume is literally the cone on its side. The equation for this line segment is $y = r - \left(\frac{r}{h}\right)x$. We thus need to evaluate $\int_0^h \pi \left(r - \left(\frac{r}{h}\right)x\right)^2 dx$.

$$= \pi \int_0^h r^2 - 2r\left(\frac{r}{h}\right)x + \left(\frac{r}{h}\right)^2 x^2 dx.$$

$$= \pi r^2 \int_0^h 1 - \left(\frac{2}{h}\right)x + \left(\frac{1}{h^2}\right)x^2 dx.$$

$$= \pi r^2 \left[h - \left(\frac{2}{h}\right)\left(\frac{1}{2}h^2\right) + \left(\frac{1}{h^2}\right)\left(\frac{1}{3}h^3\right)\right] - \left[0 - \left(\frac{2}{h}\right)0 + \left(\frac{1}{h^2}\right)0\right]$$

$$= \frac{1}{3}\pi r^2 h$$

Volume of sphere (GCSE but proven using A level stuff)

Intuitively, the volume should scale with the cube of the radius. We find the volume of a semicircle rotated 2π radians about the x axis, which is a sphere. A semicircle with radius r is given by $y = \sqrt{r^2 - x^2}$ (Note: This is a semicircle and not a circle because square root is defined as just the positive square root). We then find the volume of the sphere using the volumes of revolution trick which gives the volume of a sphere with radius r as follows

$$\pi \int_{-r}^{r} (r^2 - x^2)dx$$

$$= \pi \left[\left[r^2(r) - \frac{r^3}{3}\right] - \left[r^2(-r) - \frac{(-r)^3}{3}\right]\right]$$

$$= \pi r^3 \left[\left[1 - \frac{1}{3}\right] - \left[-1 - \frac{(-1)^3}{3}\right]\right]$$

$$= \frac{4}{3}\pi r^3$$

Surface area of sphere (GCSE but proven using A level stuff)

Intuitively the surface area should depend on r^2. As r changes what is the rate at which the volume of a sphere changes, ie what is $\frac{dV}{dr}$? Visualise this scenario in your head and realize that the rate of change of volume at any instant should be the surface area, as the corresponding change in volume when we change the radius by dr is, informally, like a thingy which thickness dr, area of the surface area, and volume dV. We find that $\frac{dV}{dr}$ is $4\pi r^2$ using the power rule for differentiation.

Iterative formulas/Numerical methods

I have a video on this, where I demonstrate that under certain conditions with the function and its derivatives, iterative formulas and newton raphson work.

Trigonometry addition formulae

See the relevant A level further maths chapter in the textbook for an introduction to complex numbers. Now we consider a very important geometric intuition about multiplying complex numbers, which is often the first chapter of further maths. Some textbooks give a geometric proof, but this proof justifies it for all angles, not just angles between 0 and 90 degrees, and is much more beautiful.

In order to understand this, you should either try to visualize or draw this until it makes sense to you. 3blue1brown's lockdown math explains all these concepts beautifully, but I cannot guarantee that those will exist since I am not affiliated with 3blue1brown nor youtube.

When you multiply 2 complex numbers a*b, what you can think of is that you stretch and rotate the entire argand diagram such that the number 1 lines up with the number a, then go to the number b on that new argand diagram. Note that i(a+bi)=ai-b by algebra. Draw a+bi and ai-b on an argand diagram and try to convince yourself that they are always perpendicular. Now the reason my trick from earlier works is becauseif you have a(x+yi) this is ax+ayi. Because moving on path 1 then path 2 on an argand diagram is the same as adding the numbers given by path 1 and path 2, we have that if you move |a|x units in the direction parallel to a then |a|y units in the direction perpendicular to a. It should hopefully be clear that this gets you to a(x+yi) in the normal grid and x+yi in the grid stretched and rotated (specifically, stretched by a factor of |a| and rotated by arg(a) radians).

Now consider the function cos(x)+isin(x) with x real. This is the function you get from walking x radians around the unit circle anticlockwise, since on an argand diagram if I have the function you get from walking x radians around the unit circle then indeed you can see that cos(x) will be the real part and sin(x) will be the imaginary part. You can draw a diagram to convince yourself of this.

Now consider (cos(x)+isin(x))(cos(y)+isin(y)) with x,y real. Using the multiplication trick from earlier, what you should do is rotate the argand diagram x radians so that the point 1 on your rotated grid corresponds to the point cos(x)+isin(x) on your original argand diagram. Then you can see that the point cos(y)+isin(y) on your rotated diagram is gotten to by walking form the original point 1 x radians anticlockwise then y radians anticlockwise, so you walk x+y radians anticlockwise. This is significant because you then end up at the point cos(x+y)+isin(x+y), which intuitively establishes the following identity:

(cos(x)+isin(x))(cos(y)+isin(y)) = cos(x+y)+isin(x+y)

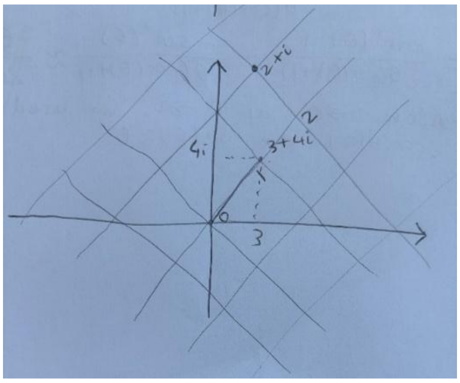
Notice how this behavior is similar to exponentials. We will expand on this idea in the exponentials and logarithms video.

Now expand the left hand side, noting that $i^2 = -1$ by definition, then we have

cos(x)cos(y)-sin(x)sin(y)+i(cos(x)sin(y)+sin(x)cos(y)=cos(x+y)+sin(x+y)

Equating the real and imaginary parts to eachother gives the desired result.

Below is an illustration for both the multiplication idea with the example (3+4i)(2+i) and in the case of multiplying on the unit circle. (Grids not accurately drawn)



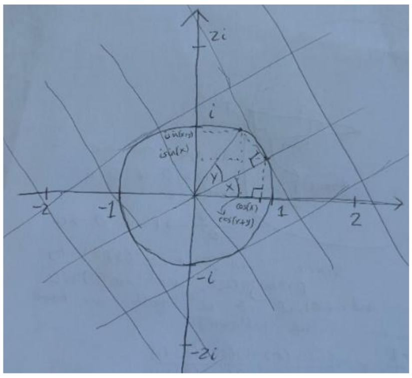


Image: Shows rotated and stretched grids on an argand diagram to visually show my point, with unit circles and stuff included in the diagrams.

As an extra challenge for capable/interested people, can you use these ideas to find that $\frac{1}{2}\sqrt{2}(1 \pm i)$ are two square roots of i and that two other cube roots of 1 that are not just 1 are given by $-\frac{1}{2} \pm \frac{1}{2}\sqrt{3}i$ without directly computing the squares or cubes but rather using the ideas above and known sines and cosines of certain angles? This is just an exercise and these ideas will be covered later in further maths and are covered in the previously mentioned lockdown math youtube videos anyway.

Sine and cosine derivatives

You can use the addition rules proved above in the definition of the derivative, but then you have to prove the following results to get the full result

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

Below is an illustration/geometric intuition for both of these results, with some algebra used.

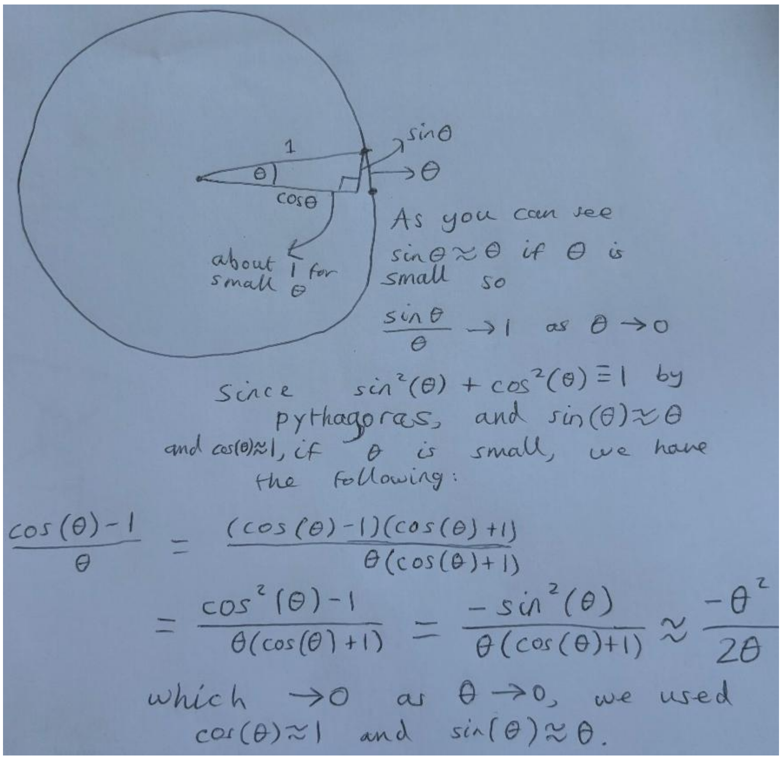


Image: Geometric proof of these limits.

Note that these last few proofs may be longer and harder to follow than the previous ones. For partial fractions, you may be satisfied with just knowing that you can always check by hand that your partial fractions indeed equal the original fraction, but I will prove why it’s done the way it’s done.

Generalized binomial theorem

The generalized binomial theorem is a special case of a concept called a Taylor series. A Taylor series is essentially when we have a function and we use the following method to try to fit an infinite polynomial to the function f(x): Let the constant term of the polynomial be f(0) so the functions and our polynomial at least match in terms of their value at x=0. Now we let the coefficient of x in our polynomial be f'(0) so our polynomial and its slope at x=0 match that of the original function, so if we zoom into the graph enough that f looks like roughly a straight line (since we are assuming f is "smooth" using the intuitive informal definition) our polynomial will be a decent approximation for f. Now we let the term in x^2 of our polynomial be $\frac{1}{2}f''(0)$ so that the second derivatives at x=0 match. We keep doing this until we get an infinite polynomial, and the hope is that this polynomial well approximates the original function as we add more terms, and that in the limit the infinite polynomial we get equals the original function. The infinite polynomial we get is called the Taylor series of the function. This is true in a lot of cases, and we prove it for the binomial theorem $(1+x)^n$ as that is used at A level, as the binomial expansion is actually the Taylor series of $(1+x)^n$ for general n.

We prove that $(1+x)^n$ actually equals its Taylor series when |x|<1. Consider this: It can be shown using standard A level techniques that the unique solution to the differential equation

$$\frac{dy}{dx} = \frac{ny}{1+x}$$

that satisfies y=1 when x=0 is $y = (1+x)^n$ (you should be able to show this), and that if

$$y = 1 + nx + \frac{n(n-1)}{2!}x^2 + \frac{n(n-1)(n-2)}{3!}x^3 + \cdots$$

Then the differential equation above is satisfied (I will walk through this part)

Goal: To show that this satisfies the differential equation above with the initial conditions so it must equal $(1+x)^n$. Trying to differentiate this power series will give

$$\frac{dy}{dx} = n + n(n-1)x + \frac{n(n-1)(n-2)}{2!}x^2 + \frac{n(n-1)(n-2)(n-3)}{3!}x^3 + \cdots$$

$$x\frac{dy}{dx} = nx + n(n-1)x^2 + \frac{n(n-1)(n-2)}{2!}x^3 + \frac{n(n-1)(n-2)(n-3)}{3!}x^4 + \cdots$$

$$(1+x)\frac{dy}{dx} = n + nx + n(n-1)x + n(n-1)x^2 + \frac{n(n-1)(n-2)}{2!}x^2 + \frac{n(n-1)(n-2)}{2!}x^3 + \frac{n(n-1)(n-2)(n-3)}{3!}x^3 + \frac{n(n-1)(n-2)(n-3)}{3!}x^4 + \frac{n(n-1)(n-2)(n-3)(n-4)}{4!}x^4 + \cdots$$

This looks complicated, but don't worry, it simplifies nicely.

$$(1+x)\frac{dy}{dx} = n + nx\big(1+(n-1)\big) + n(n-1)x^2\left(1+\frac{n-2}{2}\right) + \frac{n(n-1)(n-2)}{2!}x^3\left(1+\frac{n-3}{3}\right) + \frac{n(n-1)(n-2)(n-3)}{3!}x^4\left(1+\frac{n-4}{4}\right) + \cdots$$

$$= n + nx(n) + n(n-1)x^2\left(\frac{n}{2}\right) + \frac{n(n-1)(n-2)}{2!}x^3\left(\frac{n}{3}\right) + \frac{n(n-1)(n-2)(n-3)}{3! * 4}x^4\left(\frac{n}{4}\right) + \cdots$$

$$= n(1 + nx + \frac{n(n-1)}{2!}x^2 + \frac{n(n-1)(n-2)}{3!}x^3 + \cdots$$

So we satisfy

$$(1+x)\frac{dy}{dx} = ny$$

which is a rearrangement of the original differential equation. However, we're not quite done, since although the derivative of the sum is the sum of the derivative for finite sums, I have just used it for an infinite sum and I have not yet justified it. In general in maths, it requires justification to extrapolate anything finite to anything infinite. We also haven't addressed why this expansion is only valid if |x|<1.

Before reading on, try using your calculator to calculate the sum of the binomial series for $(1+x)^{-0.5}$ at (i) x=0.5 and (ii) x=2. See how the values change as you add more terms until the term in x^5. Can you guess why x has to be between -1 and 1?

Note: Since calculating the coefficients is not the point of this exercise, I'll give that the series is

$(1+x)^{-0.5} = 1 - \frac{1}{2}x + \frac{3}{8}x^2 - \frac{5}{16}x^3 + \frac{35}{128}x^4 - \frac{63}{256}x^5 + \cdots$

Consider the ratio between, say, the x^{99} coefficient and the x^{100}coefficient. In order to get from one coefficient to the other, we need to divide by 100x and multiply by n-99. So, since n is fixed, the ratio which is $\frac{100x}{n-99}$ will approach -x as "100" and "99" get larger and larger. The takeaway should be that only when x is between -1 and 1 does the ratio approach between -1 and 1 meaning the terms get smaller in size so the sum doesn't go off to infinity. For positive integer exponents the series is valid everywhere sine the terms eventually are all 0.

The formal justification for why we can differentiate an infinite power series and why it indeed converges with a rigorous proof and not a hand waving argument when x is between -1 and 1will be left as an appendix, since it is technical and not part of the "core" of this argument. These last few paragraphs relate to an area of maths called analysis.

For other taylor series which are typically only used in further maths, here is the differential equation used to prove their correctness, note that the full justification is also given in the appendix.

e^x: $\frac{dy}{dx} = y$ and y=1 when x=0 (Always converges since the ratio of consecutive terms always -> 0)

$\ln(1+x)$: $\frac{dy}{dx} = \frac{1}{1+x}$ and y=0 when x=0, or from differentiating the series for $(1+x)^{-1}$ (Also valid when |x|<1)

$\sin(x)$ and $\cos(x)$: Comes from Euler's identity $e^{ix} = cos(x) + isin(x)$ which we will prove in the video in the section on exponentials and logarithms, then equating the real and imaginary parts. (Always converges since the ratio of consecutive terms always -> 0). Alternatively, you can use the differential equation $\frac{d^2y}{dx^2} + y = 0$ with the initial conditions satisfied by sin and cos when x=0. We learn how to solve this differential equation above and differential equations like it in A level further maths and its corresponding document like this one. It is ok for us to not develop this theory here because it is only in further maths that students are actually expected to know the series for sin and cos.

Partial fractions

This one is tricky, in the exam if you're satisfied just knowing you can check manually that your answer works then great, but here we will justify why we use the method that we do.

The precise statement we want to prove is as follows:

If we have a fraction like

$$\frac{P(x)}{(x-a_1)^{m_1}(x-a_2)^{m_2}\dots(x-a_n)^{m_n}}$$

Where P(x) is a polynomial

Then it can be written in the following form

$Q(x)+\frac{b_1}{x-a_1}+\frac{b_2}{(x-a_1)^2}+\cdots+\frac{b_{m_1}}{(x-a_1)^{m_1}}$ +(n-1 more analogous sums for the rest of the roots a_k)

Where Q(x) is a polynomial.

Note that this is an extremely general statement, in A levels n and the m's will not get bigger than 2 or 3.

Definition: the degree (often called deg) of a polynomial is defined as the highest power of x that appears.

Note that if we divide P(x) by the denominator of our original expression, we will get a polynomial Q(x) (which will be 0 if P(x) has a lower degree than the denominator) and a remainder divided by the denominator of the original expression, where this remainder will also have a degree lower than the degree of the denominator (Otherwise we could keep doing long division on it until the degree is lower). Therefore we can assume that P(x) has a degree lower than the denominator and then prove that we get the not Q(x) terms.

In order to do this proof, we will need to use a collorary of Bezout's identity, essentially the assertion that if $P_1(x)$ and $P_2(x)$ are polynomials with no roots in common (which by the factor theorem is equivalent to having a greatest common divisor of 1) and that they are both of the form $C(x-a_1)^{m_1}(x-a_2)^{m_2}\dots(x-a_n)^{m_n}$ (ie a product of linear factors) then there exist polynomials A(x) and B(x) such that A$P_1(x)$+B$P_2(x)$=1. It turns out that the assertion that the polynomial is a product of linear factors is true for all polynomials and this is called the fundamental theorem of algebra, however this is not needed for our purposes, so although there are nice proofs of this fact, we will do them in a later document since this is long enough, so we will add the condition that $P_1(x)$ and $P_2(x)$ are products of linear factors so this way there are no implicit assumptions, and it will be enough for our proof. We also do not have uniqueness of polynomial factorization fully proven, but the factor theorem means no common roots is equivalent to no common factors.

First, we will do a proof assuming the fact above, then we will prove it. We define that the Euclidian division of a polynomial P by a non-zero polynomial T is defined as P=TQ+R where R is a remainder polynomial with deg(R)<deg(T). You can see that surely this must always exist if you've ever done polynomial long division before and you think about it hard enough.

Now I will define F as the numerator of the original expression which by assumption has degree lower than the denominator of the original expression which I will call G. I will then define $G_1=(x-a_1)^{m_1}$

and $G_2 = (x - a_2)^{m_2}(x - a_3)^{m_3} \dots (x - a_n)^{m_n}$. Since the a's are distinct, G_1 and G_2 have no common roots, so Bezout's identity applies. Let CG_1+DG_2=1 then we have the following:

Let $DF = G_1Q + F_1$ with $\deg F_1 < \deg G_1$ be the Euclidean division of DF by G_1. Setting $F_2 = CF + QG_2$, one gets

$$\begin{aligned}\frac{F}{G} &= \frac{F(CG_1 + DG_2)}{G_1G_2} = \frac{DF}{G_1} + \frac{CF}{G_2}\\ &= \frac{F_1 + G_1Q}{G_1} + \frac{F_2 - G_2Q}{G_2}\\ &= \frac{F_1}{G_1} + Q + \frac{F_2}{G_2} - Q\\ &= \frac{F_1}{G_1} + \frac{F_2}{G_2}.\end{aligned}$$

It remains to show that $\deg F_2 < \deg G_2$. By reducing the last sum of fractions to a common denominator, one gets $F = F_2G_1 + F_1G_2$, and thus

$$\begin{aligned}\deg F_2 &= \deg(F - F_1G_2) - \deg G_1 \leq \max(\deg F, \deg(F_1G_2)) - \deg G_1\\ &< \max(\deg G, \deg(G_1G_2)) - \deg G_1 = \deg G_2\end{aligned}$$

The last line follows from the fact that clearly the degree of the product of polynomials is the sum of the degrees so therefore deg($F - F_1G_2$) = deg(F_2)+deg(G_1), since we have $F = F_1G_2 + F_2G_1$

so $F - F_1G_2$=F_2G_1.

Then the next step is because the degree of the difference of two polynomials can't be more than the degree of both of them, then the next step follows because by earlier assertions it follows that the degree of both things in the max() are decreasing in the step given, and in fact now both equal to deg(G). Since $G = G_1G_2$ by definition, we have deg(G)=deg(G_1)+deg(G_2) so the last step follows.

Applying this again to the fraction $\frac{F_2}{G_2}$ and then iteratively doing this n times we get the following:

$$\frac{P(x)}{(x - a_1)^{m_1}(x - a_2)^{m_2} \dots (x - a_n)^{m_n}}$$

$=Q(x) + \frac{T_1(x)}{(x-a_1)^{m_1}} + \frac{T_2(x)}{(x-a_2)^{m_2}} + \cdots + \frac{T_n(x)}{(x-a_n)^{m_n}}$

Where each T polynomial has degree less than its denominator. It remains to show that each of these can be broken up into constants divided by powers of the respective linear factors. I will show how this is done for the first fraction and the method for the rest of them is analogous. If m_1=1 then we're done, so suppose m_1>1 so $T_1(x)$ has degree at most $m_1 - 1$.

We can write $T_1(x) = a_0 + a_1x^1 + a_2x^2 + \cdots a_{m_1-1}x^{m_1-1}$ where each a can be 0 or anything else. Now just watch this

$$\frac{T_1(x)}{(x - a_1)^{m_1}}$$

$$= \frac{a_0 + a_1x^1 + a_2x^2 + \cdots a_{m_1-1}x^{m_1-1}}{(x - a_1)^{m_1}}$$

$$= \frac{a_0 + a_1x^1 + a_2x^2 + \cdots a_{m_1-1}x^{m_1-1} - a_{m_1-1}(x - a_1)^{m_1-1}}{(x - a_1)^{m_1}} + \frac{a_{m_1-1}(x - a_1)^{m_1-1}}{(x - a_1)^{m_1}}$$

Notice that the degree of the first term is at most $m_1 - 2$ because the coefficient in x^{m_1-1} is being cancelled. Let's suppose the new numerator of this first term is $b_0 + b_1x^1 + b_2x^2 + \cdots b_{m_1-2}x^{m_1-2}$

$$= \frac{b_0 + b_1x^1 + b_2x^2 + \cdots b_{m_1-2}x^{m_1-2}}{(x-a_1)^{m_1}} + \frac{a_{m_1-1}}{x-a_1}$$

Now do a similar trick:

$$= \frac{b_0 + b_1x^1 + b_2x^2 + \cdots b_{m_1-2}x^{m_1-2} - b_{m_1-2}(x-a_1)^{m_1-2}}{(x-a_1)^{m_1}} + \frac{b_{m_1-2}(x-a_1)^{m_1-2}}{(x-a_1)^{m_1}} + \frac{a_{m_1-1}}{x-a_1}$$

Now the degree of the first term decreases again, and we get a $\frac{b_{m_1-2}}{(x-a_1)^2}$ term after cancelling $(x-a_1)^{m_1-2}$ from the numerator and denominator of our second term, and we can keep decreasing the degree and getting a constant over the next power m_1 times until we get the desired result.

Proof of part of Bezout's identity for the partial fractions proof

Ok so we have two polynomials which are products of linear factors and have no roots in common. Call the one with larger degree P_1 and the one with smaller degree P_2. If their degrees are the same it does not matter what we call them. Now do Euclidean division: $P_1 = Q_1P_2 + R_1$ where deg(R_1)<deg(P_2). Now we work with polynomials R_1 and P_2 and do the same thing, except now the smallest degree of the two is lower. We get $P_2 = Q_2R_1 + R_2$ and then do the same with R_1 and $R_2$and again the smaller of these (R_2) has a smaller degree than the smallest degree of the previous two we worked with. Keep doing this until the remainder is a constant. Here is an example where for two polynomials with no common root we explicitly find polynomials such that 1 equals a polynomial times P_1 plus a polynomial times P_2:

$$P_1 = (x-1)^2(x-2), P_2 = (x+1)(x+3)$$

$$P_1 = x^3 - 4x^2 + 5x - 2, P_2 = x^2 + 4x + 3$$

$$x^3 - 4x^2 + 5x - 2 = (x-8)(x^2+4x+3) + (34x+22)$$

$$x^2 + 4x + 3 = \left(\frac{1}{34}x + \frac{57}{578}\right)(34x+22) + \frac{240}{289}$$

Now we are done, since we can express $\frac{240}{289}$ as $34x + 22$ times a polynomial plus $x^2 + 4x + 3$ times a polynomial (by rearranging the bottom equation) but then we can express $34x + 22$ as $x^2 + 4x + 3$ times a polynomial plus $x^3 - 4x^2 + 5x - 2$ times another polynomial (by rearranging the second equation from the bottom). This means that our original target, $\frac{240}{289}$ equals $x^2 + 4x + 3$ times a polynomial plus $x^3 - 4x^2 + 5x - 2$ times another polynomial. Simply multiply these polynomials by $\frac{240}{289}$ then we're done, in fact we get:

$$\frac{240}{289} = x^2 + 4x + 3 - \left(\frac{1}{34}x + \frac{57}{578}\right)(34x+22)$$

$$1 = \frac{289}{240}(x^2+4x+3) - \frac{289}{240}\left(\frac{1}{34}x + \frac{57}{578}\right)(34x+22)$$

$$1 = \frac{289}{240}(x^2+4x+3) - \frac{289}{240}\left(\frac{1}{34}x + \frac{57}{578}\right)\left(x^3 - 4x^2 + 5x - 2 - (x-8)(x^2+4x+3)\right)$$

$$1 = \frac{289}{240}(x^2+4x+3) - \frac{289}{240}\left(\frac{1}{34}x + \frac{57}{578}\right)(x^3-4x^2+5x-2) + \frac{289}{240}\left(\frac{1}{34}x + \frac{57}{578}\right)(x-8)(x^2+4x+3)$$

$$1 = -\left(\frac{17}{480}x + \frac{57}{480}\right)(x^3-4x^2+5x-2) + \left(\frac{17}{480}x^2 - \frac{79}{480}x + \frac{61}{240}\right)(x^2+4x+3)$$

$$1 = -\left(\frac{17}{480}x + \frac{57}{480}\right)\left((x-1)^2(x-2)\right) + \left(\frac{17}{480}x^2 - \frac{79}{480}x + \frac{61}{240}\right)\left((x+1)(x+3)\right)$$

Note that it remains to show that we always end up with a constant like $\frac{240}{289}$ that is not zero, otherwise we can't divide by it like in the example above!

Any two pairs of polynomials we work with will not share any common factors, since lets say for the sake of example that R_3 and R_4 shared a common factor, then R_2 which is a quotient polynomial times R_3 plus R_4 will have that factor, then R_1 which is a quotient polynomial times R_2 plus R_3 must also have that factor, and we keep going until realizing that by this reasoning our original P's must have had that original factor. If R_n was a polynomial of degree at least 1 but then R_{n+1} was suddenly 0 and not a non-zero constant, then that would mean that R_n would divide R_{n-1} with zero remainder as we would get $R_n = QR_{n-1} + 0$, meaning R_n and R_{n-1} share a common factor of R_{n-1} which is a contradiction, as our original P's were assumed to have no common factors.

More on exponentials and logarithms

This is explained in a video (which is quite involved), but consider the following motivating question to see why further discussion on this is needed:

If we have $\frac{dy}{dx} = y$

Then the standard separation of variables method gives $\int \frac{1}{y} dy = \int 1\, dx$

So ln(y)=x+c

So $y = e^x e^c$

But then we say that $y = Ae^x$ for any A, but if we are working in the real numbers e^c is always positive. It turns out that $y = -e^x$ does in fact satisfy the differential equation, but it is not at all obvious how it isn't absurd to say A=-1 when A was derived from the exponential function which does not appear to take negative values. After all, e^x is clearly always positive from the graph! But, things are not always as they seem.

The video derives from basic exponent laws what it means to take a positive real number to a non-integer real power, and then defines exponentiation with complex exponents and extend the definition of powers to the complex numbers. I define the complex logarithm and the problems that come from it being multi-valued or having a branch cut. I explain which exponent and logarithm rules work in the complex numbers and why. I then discuss a bit about the integral of 1/x and give a better explanation for the ln(|x|) standard answer. I do a brief aside on contour integration (Don't worry if you don't understand this, I explain it again in the technical results document. Hopefully you understand it then.) since it will be used in a later proof, then finally go back and answer the motivating differential equations question above.

Appendix: Binomial theorem technical details (If everything else wasn't basically an analysis course in disguise, this definitely is)

First, we will tackle the problem of showing that the binomial series always converges if |x|<1.

We have established that the ratio of consecutive terms in the binomial series approaches -x, formally the ratio is given by $\frac{Ax}{n-(A-1)}$ between the x^A and x^{A-1} terms. The difference between this and -x is $\frac{Ax}{n-(A-1)} + x = \frac{Ax+nx-(A-1)x}{n-(A-1)} = \frac{x(n+1)}{n+1-A}$ which gets as small as we want if we make A large enough. Formally, we can see that for any ε, no matter how small, if A is at least $n + 1 + \frac{x(n+1)}{\varepsilon}$ then the ratio between consecutive terms will always be within ε of -x. This is what it formally means to say that the ratios will converge to -x in the limit. Now we show that if in a series the ratio of the terms approaches in this sense a value with absolute value less than 1 (which we have shown that the binomial series does when |x|<1) it converges, and in fact it is absolutely convergent, meaning that the sum of the absolute values of the terms converges, which will be important later. Note that absolute convergence implies convergence as informally by the triangle inequality the "tails" of the original sum are less big than the "tails" of the sum of absolute values.

First, we see that $\sum_{n=1}^{\infty} |a_n|$ converges if $\frac{|a_{n+1}|}{|a_n|}$ is always less than some number c which is less than 1 because $\sum_{n=1}^{\infty} c^{n-1}|a_1|$ converges by geometric series and the sequence $\sum_{n=1}^{\infty} |a_n|$ is a sequence of positive numbers where since $c^{n-1}|a_1| \geq |a_n|$ by definition we have that $\sum_{n=1}^{\infty} |a_n|$ is a limit of increasing positive numbers bounded above by $\sum_{n=1}^{\infty} c^{n-1}|a_1|$ so the values $\sum_{n=1}^{m} |a_n|$ for different m's must have a least upper bound (This is obvious and often a first principle/axiom) so that is what it approaches as m goes to infinity, ie $\sum_{n=1}^{\infty} |a_n|$ converges.

Now we see that $\sum_{n=1}^{\infty} |a_n|$ converges if $\frac{|a_{n+1}|}{|a_n|}$ is eventually (ie whenever n>k for some k) always less than some number c which is less than 1. This is because we split the sum into the first k terms (which converges as it is a finite sum) and the rest of the terms (which converges by the fact above). Now we see that $\sum_{n=1}^{\infty} |a_n|$ converges if $\frac{|a_{n+1}|}{|a_n|}$ approaches a number less than 1. This is because, for example, lets say it approached 0.5, then by the definition of approaching it will always be eventually between 0.25 and 0.75 and we apply the thing above.

Also, if |x|>1 the terms eventually continue to grow in magnitude as their ratio eventually becomes >1 for all terms after a certain point so the binomial series can't possibly converge.

Ok so now we know that the binomial series converges absolutely if x is less than 1. Now I will explain why absolute convergence is important.

When we justify differentiating a power series we will be working with sums. We note that intuitively, we know that if we rearrange the order of the terms in the sum, the sum will not change. However, this is not always true for infinite sums. For example:

Consider the infinite series

$$1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\frac{1}{5}-\frac{1}{6}+\frac{1}{7}-\frac{1}{8}+\frac{1}{9}-\frac{1}{10}+\cdots.$$

Let us designate the given series by S. Then

(a) $S = 1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\frac{1}{5}-\frac{1}{6}+\frac{1}{7}-\frac{1}{8}+\frac{1}{9}-\frac{1}{10}+\cdots.$

(b) $2S = 2-1+\frac{2}{3}-\frac{2}{4}+\frac{2}{5}-\frac{2}{6}+\frac{2}{7}-\frac{2}{8}+\frac{2}{9}-\frac{2}{10}+\cdots$

$\quad = 1+\frac{2}{3}-\frac{1}{2}+\frac{2}{5}-\frac{1}{3}+\frac{2}{7}-\frac{1}{4}+\frac{2}{9}-\frac{1}{5}+\cdots$

[First multiply S by 2 and then simplify]

(c) $2S = 1 - \frac{1}{2} + \left(\frac{2}{3} - \frac{1}{3}\right) - \frac{1}{4} + \left(\frac{2}{5} - \frac{1}{5}\right) + \cdots$ [Rearranging]

(d) $2S = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} + \cdots$ [After simplification]

(e) $2S = S$ [Using (a)]

We conclude that **1 = 2!**

Fun fact: it turns out that S=ln(2)

However, if you take the absolute value of all the terms in the sum, then the sum is well behaved and will equal the same value no matter how you order the terms. The proof of this will use a formal definition of convergence, and the triangle inequality which says that

$$|a_1 + a_2 + a_3 + \cdots + a_n| \leq |a_1| + |a_2| + |a_3| + \cdots + |a_n|$$

The reason for this is (using complex numbers ideas) because the first expression is like walking a distance of $|a_1|$ in some direction in the complex plane, then a distance of $|a_2|$ in some other direction, then a distance of $|a_3|$ in some other direction, and so on, then recording your distance from the origin at the end. The right hand expression is like walking all of those distances in the same direction and then recording how far you've walked, which intuitively will always be a distance at least as long. This will be used in the proof below (Note: The triangle inequality is true for infinite sums, as the sum to n of $|a_1 + a_2 + a_3 + \cdots + a_n|$ is less than or equal to the sum to n of $|a_1| + |a_2| + |a_3| + \cdots + |a_n|$ for all n so in the limit as n goes to infinity (ie the infinite sum) the latter expression cannot possibly be less!) Now here is the proof:

Note: For a sum to equal a value what we mean is that all finite partial sums eventually will always be as close as we want to that value.

Define $\sum_{n=1}^{\infty} a_n = S$, and let this be a convergent sum.

Suppose $\sum a_n$ is absolutely convergent so that $\sum_{n=1}^{\infty} |a_n| = L$, then there exists an M for any ε no matter how small such that $\sum_{n=M+1}^{\infty} |a_n| < \varepsilon$. Let b_n be a re-ordering of the terms of a_n and let $\sum_{n=1}^{N} b_n = S_N$, we need to show that as $N \to \infty$ $S_N \to S$, ie $|S_N - S| < \varepsilon$ for any ε no matter how small provided N is large enough, ie if, say, $\varepsilon = 0.0001$ then we need to show that if we pick N large enough, S_x will be within 0.0001 of S for all $x \geq N$. For any $\varepsilon > 0$ pick M so $\sum_{n=M+1}^{\infty} |a_n| < \varepsilon$.

$n=M+1 \quad < \varepsilon.$

Let C_0 be large enough such that in $b_1, b_2, b_3 \cdots b_c$ all terms $a_1, a_2, \cdots a_m$ appear. Then

$$S_{c_0} = \sum_{n=1}^{c_0} b_n = \sum_{n=1}^{M} a_m + \sum_{\substack{n>M \\ \text{and } a_n \\ \text{is in the first} \\ c_0\ b\text{'s somewhere}}} a_n$$

and ~~a similar expression~~ this is true if we keep making C_0 larger so we have for any $C > C_0$

$$|S_c - S| = \left|S_c - \sum_{n=1}^{\infty} a_n\right| = \left|\sum_{n=1}^{M} a_n + \sum_{\substack{n>M \\ \text{and } a_n \in b_{1,\dots c}}} a_n - \sum_{n=1}^{\infty} a_n\right|$$

$$= \left|\cancel{\sum_{n=1}^{\infty} a_n} - \sum_{\substack{n>M \\ \text{and} \\ a_n \notin b_{1,\dots c}}} a_n - \cancel{\sum_{n=1}^{\infty} a_n}\right| = \left|\sum_{\substack{n>M \\ \text{and} \\ a_n \notin b_{1,\dots c}}} a_n\right| \leq \sum_{\substack{n>M \\ \text{and} \\ a_n \notin b_{1,\dots c}}} |a_n| \leq \sum_{n>M} |a_n| < \varepsilon.$$

Images: Handwritten proof absolute convergence implies stuff plays nice.

Ok now that that's done here is a proof we can differentiate infinite power series in the usual way under certain conditions. We call the radius of convergence of a power series R where R is the largest value such that for all x such that |x|<R the power series converges when evaluated at x. We know that for the binomial series R=1, and we will show that differentiation of power series works when x is within the radius of convergence. Here are some things we will need:

$$b^n - a^n - n(b-a)a^{n-1} = (b-a)^2(b^{n-2} + 2ab^{n-3} + 3a^2b^{n-4} + \cdots + (n-1)a^{n-2}).$$

Proof. If $b = a$, we are done. Otherwise,

$$\frac{b^n - a^n}{b - a} = b^{n-1} + ab^{n-2} + a^2b^{n-3} + \cdots + a^{n-1}.$$

Differentiate both sides with respect to a. Then

$$\frac{-na^{n-1}(b-a) + b^n - a^n}{(b-a)^2} = b^{n-2} + 2ab^{n-3} + \cdots + (n-1)a^{n-2}.$$

Rearranging gives the result.

This implies that

$$(z+h)^n - z^n - nhz^{n-1} = h^2((z+h)^{n-2} + 2z(z+h)^{n-3} + \cdots + (n-1)z^{n-2}),$$

which is actually the form we need.

Lemma. Let $a_n z^n$ have radius of convergence R, and let $|z| < R$. Then $\sum n a_n z^{n-1}$ converges (absolutely).

Proof. Pick r such that $|z| < r < R$. Then $\sum |a_n| r^n$ converges, so the terms $|a_n| r^n$ are bounded above by, say, C. Now

$$\sum n|a_n z^{n-1}| = \sum n|a_n| r^{n-1} \left(\frac{|z|}{r}\right)^{n-1} \leq \frac{C}{r} \sum n \left(\frac{|z|}{r}\right)^{n-1}$$

The series $\sum n \left(\frac{|z|}{r}\right)^{n-1}$ converges, by the ratio test. So $\sum n|a_n z^{n-1}|$ converges, by the comparison test. □

Note: the comparison test is the idea that a sum of positive terms bounded by a convergent sum of positive terms must be convergent, because the first sum's partial sums (sums up to n) are bounded above so they have a least upper bound which they must converge to.

Corollary. Under the same conditions,

$$\sum_{n=2}^{\infty} \binom{n}{2} a_n z^{n-2}$$

converges absolutely.

Proof. Apply Lemma above again and divide by 2. $\square$

Note: in the image below, something being o(h) means that something divided by h approaches 0 as h approaches 0, in other words it gets small "much faster than h", and by rearranging, you can see that the thing it says we want to be o(h) being o(h) means the difference between the actual derivative in question and the derivative we want is indeed zero.

Theorem. Let $\sum a_n z^n$ be a power series with radius of convergence R. For $|z| < R$, let

$$f(z) = \sum_{n=0}^{\infty} a_n z^n \text{ and } g(z) = \sum_{n=1}^{\infty} n a_n z^{n-1}.$$

Then f is differentiable with derivative g.

Proof. We want $f(z+h) - f(z) - hg(z)$ to be $o(h)$. We have

$$f(z+h) - f(z) - hg(z) = \sum_{n=2}^{\infty} a_n((z+h)^n - z^n - hnz^n).$$

We started summing from $n = 2$ since the $n = 0$ and $n = 1$ terms are 0. Using our first lemma, we are left with

$$h^2 \sum_{n=2}^{\infty} a_n\left((z+h)^{n-2} + 2z(z+h)^{n-3} + \cdots + (n-1)z^{n-2}\right)$$

We want the huge infinite series to be bounded, and then the whole thing is a bounded thing times h^2, which is definitely $o(h)$.

Pick r such that $|z| < r < R$. If h is small enough that $|z+h| \leq r$, then the last infinite series is bounded above (in modulus) by

$$\sum_{n=2}^{\infty} |a_n|(r^{n-2} + 2r^{n-2} + \cdots + (n-1)r^{n-2}) = \sum_{n=2}^{\infty} |a_n| \binom{n}{2} r^{n-2},$$

which is bounded. So done. $\square$

Interestingly, by taking the antiderivative of the binomial series for $(1+x)^{-1}$ we get a taylor series for ln(1+x) valid when |x|<1, and it turns out that this series converges when x=1, and while the full justification will not be given here (see Abel's theorem for power series), this is an intuition for the "Fun fact: S=ln(2)" fun fact above. It really is all related!

Note: Differentiation of power series multiplies the ratio between consecutive terms by n/n-1 and shifts the terms by 1, so if the ratio approached something less than 1 or more than 1 that property will still hold after differentiation, therefore the radius of convergence of the differentiated or antidifferentiated power series is the same as the radius of convergence of the original power series (This argument works for the "usual" functions where it is the case that the ratio between consecutive terms approaches something, but the statement is true more generally even in some cases where this argument does not work, but you do not need to know this for A level).