Matrix determinant is volume

This has been proven in a video using some visual arguments about vectors.

Matrix inverse formula works

This has been proven in the same video as the determinant proof video using the determinant formula and the fact that it is volume.

Polar coordinate integration

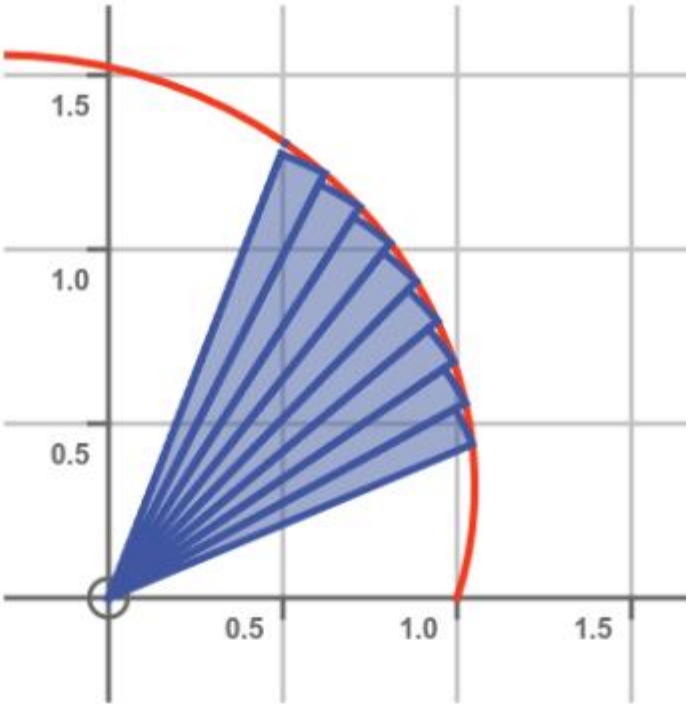


Image: Shows the area of an arc of a polar curve approximated by thin right angled triangles

We see from the diagram above that we are summing the area of right angled triangles which have a base of r, a height of rsin(dθ) which is therefore approximately rdθ and approaches rdθ as dθ goes to 0. So the area of these triangles is $\frac{1}{2}r^2 d\vartheta$. Summing these areas and taking a limit as dθ goes to 0 is kind of what an integral is, so we get the formula $\int \frac{1}{2}r^2 d\vartheta$.

Some power series properties

Definition: Least upper bound / Supremum

This essentially means the smallest number greater than or equal to a set of numbers. This always exists, but it is not necessarily always attained, for example the sequence 0.9, 0.99, 0.999 has a least upper bound of 1 but 1 is not attained in this sequence. It is obvious that any bounded set has a least upper bound, so therefore any increasing sequence which is bounded above converges as it converges to its least upper bound, similarly for decreasing sequences bounded below.

The infimum is the opposite, it is the highest lower bound.

If I have a sequence, then the lim inf of the sequence is defined as the limit of the infima of the sequence as I remove more terms at the start. Take some time to think about this definition and the following corrolaries if you need to, since I'm not sure how confusing this will be for someone who's never done analysis.

As an example, if my sequence is 0, 2, 0.9, 2, 0.99, 2, 0.999, etc then if I remove terms from the start of this sequence the infima will approach 1 so the lim inf is 1. The limit of the sequence does not exist since the terms do not eventually get as close as you want a value. The lim sup, which is the opposite of the lim inf (Essentially the limit of suprema) of this sequence is 2. Since the infimum of a sequence does not decrease when you remove terms, the lim inf can be thought of the supremum of the infima of the tails of the sequence (ie the sequences you get after removing terms). In more precise terms, $\liminf_{n\to\infty} a_n = \sup_n \inf_{k\geq n} a_k$ In fact, the lim inf of a sequence always exists, it just might be infinity if the sequence is unbounded, since the infima of the tails are increasing, and if they are increasing and bounded they will converge to something (as if something is bounded it has a least upper bound), otherwise they will go to infinity.

We will now prove some properties of power series. Specifically, that all power series have a unique radius of convergence R with the property that if |x|<R then the series converges absolutely, and if |x|>R then the series diverges. The proof of this comes from the cambridge tripos notes for analysis I

Lemma. Suppose that $\sum a_n z^n$ converges and $|w| < |z|$, then $\sum a_n w^n$ converges (absolutely).

Proof. We know that

$$|a_n w^n| = |a_n z^n| \cdot \left|\frac{w}{z}\right|^n.$$

Since $\sum a_n z^n$ converges, the terms $a_n z^n$ are bounded. So pick C such that

$$|a_n z^n| \leq C$$

for every n. Then

$$0 \leq \sum_{n=0}^{\infty} |a_n w^n| \leq \sum_{n=0}^{\infty} C \left|\frac{w}{z}\right|^n,$$

which converges (geometric series). So by the comparison test, $\sum a_n w^n$ converges absolutely. □

It follows that if $\sum a_n z^n$ does not converge and $|w| > |z|$, then $\sum a_n w^n$ does not converge.

Now let $R = \sup\{|z| : \sum a_n z^n \text{ converges }\}$ (R may be infinite). If $|z| < R$, then we can find z_0 with $|z_0| \in (|z|, R]$ such that $\sum_n^\infty a_n z_0^n$ converges. So by lemma above, $\sum a_n z^n$ converges. If $|z| > R$, then $\sum a_n z^n$ diverges by definition of R.

Definition (Radius of convergence)**.** The *radius of convergence* of a power series $\sum a_n z^n$ is

$$R = \sup\left\{|z| : \sum a_n z^n \text{ converges }\right\}.$$

We will now prove that in fact, $R = \frac{1}{\limsup \sqrt[n]{|a_n|}}$. Here is a proof of that, also from cambridge Analysis I:

Proof. Suppose $|z| < 1/\limsup \sqrt[n]{|a_n|}$. Then $|z| \limsup \sqrt[n]{|a_n|} < 1$. Therefore there exists N and $\varepsilon > 0$ such that

$$\sup_{n \geq N} |z| \sqrt[n]{|a_n|} \leq 1 - \varepsilon$$

by the definition of lim sup. Therefore

$$|a_n z^n| \leq (1 - \varepsilon)^n$$

for every $n \geq N$, which implies (by comparison with geometric series) that $\sum a_n z^n$ converges absolutely.

On the other hand, if $|z| \limsup \sqrt[n]{|a_n|} > 1$, it follows that $|z| \sqrt[n]{|a_n|} \geq 1$ for infinitely many n. Therefore $|a_n z^n| \geq 1$ for infinitely many n. So $\sum a_n z^n$ does not converge. □

These results will be used when we prove a result for when series solutions for differential equations are valid.

We will now prove that sums and products of power series converge inside the common radius of convergence of the power series in question, and that compositions of power series converge on a slightly different radius. We write x as if this were not centered at 0 it would just be $x - x_0$.

For sums: We know from earlier that $\sum a_n x^n$ and $\sum b_n x^n$ are absolutely convergent if |x|<R. Therefore $\sum |a_n + b_n| x^n \leq \sum |a_n| x^n + \sum |b_n| x^n \leq \sum |a_n x^n| + \sum |b_n x^n| < \infty$ (by the triangle inequality, recall the idea that |a+b|≤|a|+|b|).

For products: We note that if we have two power series A and B, then the x^n term of AB has contributions from the x^0 term of A with the x^n term of B, the x^1 term of A with the x^{n-1} term of B, the x^2 term of A with the x^{n-2} term of B, and so on. Therefore, the coefficient of the x^n term can be given by $\sum_{k=0}^{n} a_k c_{n-k}$. So AB can be written as $\sum_{n=0}^{\infty} (\sum_{k=0}^{n} a_k c_{n-k}) x^n$. We need absolute convergence to justify rearranging the terms like this, but luckily we have that, since if |x|<R then $(\sum |a_n x^n|)(\sum |b_n x^n|) < \infty$ since each sum converges in the common radius of convergence of the power series a and b.

For compositions: Suppose we want to find a power series for f(g(x)) where f is $\sum a_n x^n$, g is $\sum b_n x^n$. Set the radius of convergence of f and g to be R_f and R_g respectively. Suppose that $g(x_0) = 0$, because if $g(x_0) = k$ then we can rename f to f(x-k) then g will still be 0 at x_0. Let $g^+(x) = \sum |b_n| x^n$. Then we have a closed disc where $|g^+| < R_f$ (possible since $g^+(0) = 0$ and power series are continuous inside their radius of convergence since they are differentiable so g^+ can be as small as we want if we make the disc small enough, also making sure the disc does not have a radius larger than R_g). Then inside this disc, the sum $\sum a_n g(x)^n$ is absolutely convergent, since the absolute values are given by $\sum |a_n| g^+(|x|)^n < \sum |a_n| R_f^n < \infty$. Therefore, we can safely rearrange the terms as we like, so composite power series converge inside a finite radius. In the case where f is the exponential function, the power series in question converges in the radius of convergence of g, because g is always within R_f so we just need the other condition that x is within R_g. We will use this fact for the differential equations series solutions proof as well.

Integrating factor well-defined-ness

A constant of integration added to the integrating factor can be moved to the constant out front of the exponential so it does not matter which antiderivative we use in this context.

Second order differential equations

Technically this is Second order ordinary linear differential equations with constant coefficients.

To explain why the solutions are of the form they are and why those are the general solutions, I will start by imitating how I think textbooks should introduce it. Currently it feels like you are just following a bunch of rules for the general solution, such as adding x factors for repeated terms, but now I will show where these rules come from. Pretend like you've only seen how to solve first order differential equations using A level techniques, and that you are given the following problem:

Use the substitution $u = \frac{dy}{dx} - 2y$ to solve the differential equation $\frac{d^2y}{dx^2} - 5\frac{dy}{dx} + 6y = 0$.

Lets see how we might do this. We observe that $\frac{du}{dx} = \frac{d^2y}{dx^2} - 2\frac{dy}{dx}$ by differentiating both sides of the given substitution, which is a natural step as we need to somehow get a $\frac{d^2y}{dx^2}$ term. Now we can try subtracting 3u so that that we have the right $\frac{dy}{dx}$ coefficient of -5, specifically we get $\frac{du}{dx} - 3u = \frac{d^2y}{dx^2} - 5\frac{dy}{dx} + 6y$. How lucky indeed, we get the original differential equation! To see what is going on, consider this:

$$\frac{d^2y}{dx^2} - 5\frac{dy}{dx} + 6y$$

$$= \frac{d}{dx}\left(\frac{dy}{dx} - 2y\right) - 3\left(\frac{dy}{dx} - 2y\right)$$

I'm going to do a notational trick to illustrate the point:

$$= \left(\frac{d}{dx} - 3\right)\left(\frac{dy}{dx} - 2y\right)$$

Of course, it doesn't make much sense to say "$\frac{d}{dx} - 3$", but by "expanding" this to the form before, it hopefully makes sense that what we are effectively doing with the substitution is "factoring" the differential equation, ie we use the fact that

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

Now, let's actually solve it. Recall that we got $\frac{du}{dx} - 3u = \frac{d^2y}{dx^2} - 5\frac{dy}{dx} + 6y$ so by the original equation we have that $\frac{du}{dx} - 3u = 0$. The general solution to this is $u = Ae^{3x}$. We now reverse the substitution to get $\frac{dy}{dx} - 2y = Ae^{3x}$. We can solve this using the integrating factor method and we will end up getting that $ye^{-2x} = \int Ae^x dx = Ae^x + B$

(Note, in general, if it was not e^x in the integral but something like e^{2x} the A may become a constant multiple of A, but an arbitrary constant multiplied by something is still an arbitrary constant, so we can just write A. It may be a good idea to make a note of this when you use it, but I don't think this is generally necessary for marks on exams.)

Anyway we have that $y = Ae^{3x} + Be^{2x}$. This approach should shed some light on where the form of the solutions to these equations comes from.

Now, we tackle a problem involving repeated roots. We will use the substitution $u = \frac{dy}{dx} + 4y$ to solve $\frac{d^2y}{dx^2} + 8\frac{dy}{dx} + 16y = 0$. We eventually get that $u = Ae^{-4x}$ and therefore $\frac{dy}{dx} + 4y = Ae^{-4x}$. Now the reason that the repeated roots are difference is rooted in what happens when you now try to do the integrating factor method, which gives $\frac{d}{dx}(ye^{4x}) = A$. What happens is that the exponential terms on the right hand side are cancelled. We get that $ye^{4x} = \int A\,dx = Ax + B$, giving the general solution as $y = (Ax + B)e^{-4x}$. The same thing happens when a term on the right hand side of the differential equation is of the same form as a term in the general solution, but hopefully now it feels less like a rule to follow but something with a clear reason.

Now we look at the case of complex roots. The equation $\frac{d^2y}{dx^2} + y = 0$ gives $y = Ae^{ix} + Be^{-ix}$ for reasons the same as above. We will show that this can be converted into a linear combination of sin and cos, and is in fact equivalent to that. We now change our constants to be complex numbers so that

$y = (A + Bi)e^{ix} + (C + Di)e^{-ix}$ where A, B, C, D are arbitrary real numbers. This, by Euler's identity, is equivalent to $y = (A + Bi)\big(\cos(x) + isin(x)\big) + (C + Di)\big(\cos(-x) + isin(-x)\big)$

$$= (A + Bi)\big(\cos(x) + isin(x)\big) + (C + Di)\big(\cos(x) - isin(x)\big)$$

$$= (A + C + Bi + Di)(\cos(x)) + (-B + D + Ai - Ci)\big(sin(x)\big)$$

And we see that these are still arbitrary complex constants. To get $A + C + Bi + Di = X + Yi$ for arbitrary X and Y and $-B + D + Ai - Ci = U + Vi$ for arbitrary U and V we can set

$A = \frac{X+V}{2}, B = \frac{Y-U}{2}, C = \frac{X-V}{2}, D = \frac{Y+U}{2}$

So the general solution is all and only all of the solutions of the form Asin(x)+Bcos(x) for arbitrary complex numbers A and B. The point is we can convert arbitrary coefficients times conjugate complex exponentials into arbitrary coefficients times sines and cosines, usually without going through the calculation above.

This is now enough theory to understand everything going on with these differential equations.

In fact, there is another result, which says that if you find any particular solution to a differential equation, which you can use the ideas above to make an educated guess for without blindly following rules, then that solution plus the complementary function (the solution to the DE with 0 on the right hand side) is the general solution. To see this, if we have a particular solution y_p and a general solution y consider $y - y_p$. If the coefficients in the differential equation are constant, or more generally, arbitrary functions of x, so the differential equation is like this

$$A\frac{d^2y}{dx^2} + B\frac{dy}{dx} + Cy = f(x)$$

Where A, B, C can be constants or functions of x, but we know that y_p satisfies the differential equation so

$$A\frac{d^2y_p}{dx^2} + B\frac{dy_p}{dx} + Cy_p = f(x)$$

But now consider this:

$$A\frac{d^2(y-y_p)}{dx^2}+B\frac{d(y-y_p)}{dx}+C(y-y_p)=\left(A\frac{d^2y}{dx^2}+B\frac{dy}{dx}+Cy\right)-\left(A\frac{d^2y_p}{dx^2}+B\frac{dy_p}{dx}+Cy_p\right)$$
$$=f(x)-f(x)=0$$

Therefore $y-y_p$ is the general complementary function, so we can simply do complementary function plus particular solution.

Differential equation series solutions

Technically, we don't need to prove anything to be able to find series solutions, so the unproven assertion here is the fact that these solutions are valid on an interval. It would be rather silly if they were not. We use the fact that we can differentiate power series inside the radius of convergence.

For the first order case, suppose y'=a(x)y+b(x). This is very easy, as we can find an explicit solution using integrating factors that is therefore given only in terms of integrals of products and compositions of the exponential, a, and b, which all converge within the common radius of convergence of a(x) and b(x) (by the stuff on power series from earlier).

For the second order case, suppose $y''=a_1(x)y'+a_0(x)y+b(x)$

Where $a_1(x)=\sum_{n=0}^{\infty}a_n^{(1)}x^n$, $a_0(x)=\sum_{n=0}^{\infty}a_n^{(0)}x^n$, $b(x)=\sum_{n=0}^{\infty}b_nx^n$. We also assume $y=\sum_{n=0}^{\infty}c_nx^n$, and we will treat the derivative of y as some series which we do not yet know is the true derivative (ie, y':=$\sum_{n=0}^{\infty}(n+1)(c_{n+1}x^n)$). After we prove that the power series for y converges when the actual derivative is replaced with this series derivative, we will know that in fact, our power series y satisfies the original differential equation with the true derivative.

We therefore get the following relation by equating the x^n coefficient of both sides of our differential equation.

$$(n+2)(n+1)c_{n+2}=\sum_{p=0}^{n}(n-p+1)a_p^{(1)}c_{n-p+1}+\sum_{p=0}^{n}a_p^{(1)}c_{n-p}+b_n$$

Now set $A_r^{(1)}:=\sum_{p=0}^{\infty}|a_p^{(1)}||x|^p$, $A_r^{(0)}:=\sum_{p=0}^{\infty}|a_p^{(0)}||x|^p$, $B_r:=\sup_n|B_n||x|^n$. Note that since |x|<R, where R is the common radius of convergence of $a_1(x), a_0(x), b(x)$, these are all finite. We will assume that |x|<R for the rest of this proof. Set $M_n:=\max_{0\le k\le n}|c_k||x|^k$. Using the triangle inequality on the equation above that relates the coefficients, we have that

$$(n+2)(n+1)|c_{n+2}|\le\sum_{p=0}^{n}(n-p+1)\left|a_p^{(1)}\right||c_{n-p+1}|+\sum_{p=0}^{n}\left|a_p^{(0)}\right||c_{n-p}|+|b_n|$$

Multiplying both sides by $|x|^{n+2}$ gives

$$(n+2)(n+1)|c_{n+2}||x|^{n+2}\le\sum_{p=0}^{n}(n-p+1)\left|a_p^{(1)}\right||c_{n-p+1}||x|^{n+2}+\sum_{p=0}^{n}\left|a_p^{(0)}\right||c_{n-p}||x|^{n+2}+|b_n||x|^{n+2}$$

$$(n+2)(n+1)|c_{n+2}||x|^{n+2}\le\sum_{p=0}^{n}(n-p+1)|a_p^{(1)}||c_{n-p+1}||x|^{n-p+1}|x|^{p+1}+\sum_{p=0}^{n}|a_p^{(0)}||c_{n-p}||x|^{n-p}|x|^{p+2}+|x|^2B_r$$

By the definition of B_r so this inequality still holds.

$$(n+2)(n+1)|c_{n+2}||x|^{n+2} \le (n+1)\sum_{p=0}^{n}\left|a_p^{(1)}\right||c_{n-p+1}||x|^{n-p+1}|x|^{p+1} + \sum_{p=0}^{n}\left|a_p^{(0)}\right||c_{n-p}||x|^{n-p}|x|^{p+2} + |x|^2 B_r$$

By the definition of M and A,

$$(n+2)(n+1)|c_{n+2}||x|^{n+2} \le (n+1)M_{n+1}\sum_{p=0}^{n}\left|a_p^{(1)}\right||x|^{p+1} + M_n\sum_{p=0}^{n}\left|a_p^{(0)}\right||x|^{p+2} + |x|^2 B_r$$

$$(n+2)(n+1)|c_{n+2}||x|^{n+2} \le (n+1)M_{n+1}A_r^{(1)} + M_n A_r^{(0)} + |x|^2 B_r$$

$$|c_{n+2}||x|^{n+2} \le \frac{M_{n+1}A_r^{(1)}}{n+2} + \frac{M_n A_r^{(0)}}{(n+2)(n+1)} + \frac{|x|^2 B_r}{(n+2)(n+1)}$$

But by the definitions, $M_{n+2} - M_{n+1} \le |c_{n+2}||x|^{n+2}$ since the amount M_{n+2} can increase by from M_{n+1} is no more than the new term in our list of terms we are finding a maximum from, since they are all positive. Therefore,

$$M_{n+2} \le M_{n+1} + \frac{M_{n+1}A_r^{(1)}}{n+2} + \frac{M_n A_r^{(0)}}{(n+2)(n+1)} + \frac{|x|^2 B_r}{(n+2)(n+1)}$$

Now set $S_n := M_n + C$ where $C := \frac{|x|^2 B_r}{A_r^{(1)}(n+1) + A_r^{(0)}}$, so we get

$$S_{n+2} \le S_{n+1} + \frac{S_{n+1}A_r^{(1)}}{n+2} + \frac{S_n A_r^{(0)}}{(n+2)(n+1)} + \frac{|x|^2 B_r}{(n+2)(n+1)} - \frac{CA_r^{(1)}}{n+2} - \frac{CA_r^{(0)}}{(n+1)(n+2)}$$

Therefore we have

$$S_{n+2} \le S_{n+1}\left(1 + \frac{A_r^{(1)}}{n+2}\right) + \frac{S_n A_r^{(0)}}{(n+2)(n+1)}$$

Since the three terms on the right of the previous equation cancel. Since M, and therefore S, is non decreasing, we get

$S_{n+2} \le S_{n+1}\left(1 + \frac{A_r^{(1)}}{n+2} + \frac{A_r^{(0)}}{(n+2)(n+1)}\right) \le S_{n+1}\left(1 + \frac{A_r^{(1)} + A_r^{(0)}}{n+2}\right).$

Call $A_r^{(1)} + A_r^{(0)}$ a, then $S_{n+2} \le S_1 \prod_{k=0}^{n}\left(1 + \frac{a}{k+2}\right)$, where this symbol is a product like how sigma is a sum. We can turn this product into a sum by taking logs on both sides to obtain

$$\ln(S_{n+2}) \le \ln(S_1) + \sum_{k=0}^{n}\ln\left(1 + \frac{a}{k+2}\right)$$

Now $\ln(1+x) \le x$ for all x>-1 because $\ln(1+x) - x$ has a derivative of $\frac{1}{1+x} - 1$ which is positive when x is between -1 and 0 and negative when x is >0. $\ln(1+x) - x$ has a stationary point at (0, 0), and because it is increasing before that and decreasing after that, it is never positive. Therefore,

$$\ln(S_{n+2}) \le \ln(S_1) + \sum_{k=0}^{n}\frac{a}{k+2} < \ln(S_1) + \sum_{k=0}^{n}\frac{a}{k+1} = \ln(S_1) + \sum_{k=1}^{n+1}\frac{a}{k}$$

H_n is commonly used as shorthand for the quantity $\sum_{r=1}^{n}\frac{1}{r}$. We have that

$$\ln(S_{n+2}) \leq \ln(S_1) + a(H_{n+1}) \leq \ln(S_1) + a(H_n) + a$$

$H_n \leq 1 + \ln(n)$. Why? Suppose, for example, n=5. Then $1 + \ln(n)$ is given by this blue area with the graph being y=1/x (easily shown by integration)

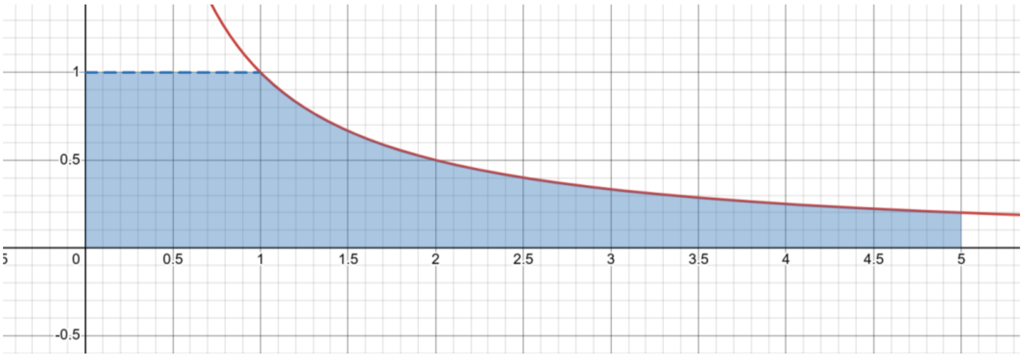


Image: Shows an area equal to 1+ln(n) for n=5 as an example using the area bounded by y=0, y=1, y=1/x, x=0 and x=5.

And H_n is this area which is clearly smaller:

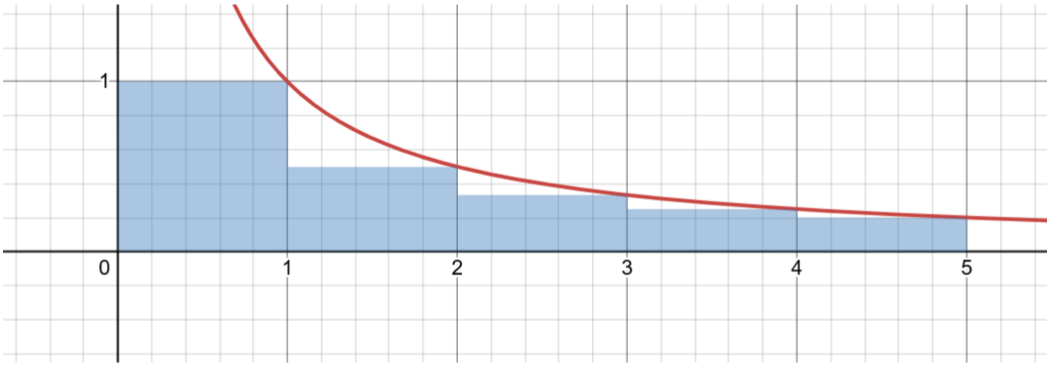


Image: Shows an area equal to H(n) for n=5 as an example by using the area bounded by x=0, x=5, y=0, y=1/ceiling(x), visually demonstrating that H(n)<1+ln(n)

So $\ln(S_{n+2}) \leq c + a(H_n) \leq c + a(1 + \ln(n)) \leq c + a(\ln(n))$

Where in the last step I have renamed c: What matters is it's still a constant. Un-logging both sides gives

$$S_{n+2} \leq cn^a$$

Where again, c is still some constant.

Why do all this? Because the radius of convergence of our power series is given by

$\frac{1}{\limsup_{n\to\infty} |c_n|^{1/n}}$.

Now using the definitions, this result, and the fact that $S_n \geq M_n \geq 0$, we have that

$$|c_n|^{1/n} \leq \left(\frac{M_n}{|x|^n}\right)^{\frac{1}{n}} \leq \left(\frac{S_n}{|x|^n}\right)^{\frac{1}{n}} \leq \left(\frac{cn^a}{|x|^n}\right)^{\frac{1}{n}} = \frac{c^{1/n}n^{a/n}}{|x|}$$

So, since $c^{1/n}n^{a/n} \to 1$ (You can take logs and use L'hopital, proved in the video, to check this), the lim sup of this is no larger than $\frac{1}{|x|}$. Since x could be anything less than R (where R is the common radius of convergence that we've assumed |x| is less than the entire time), the lim sup is no larger than $\frac{1}{R}$, so the reciprocal is no smaller than R, so done (at last). The series solution does in fact converge within the common radius of convergence of the coefficients of a 2nd order linear differential equation.

Distance between point and plane formula

This is because if a plane is defined by ax+by+cz+d=0 then we know that any point's distance from the plane is given by the length of the line segment normal to the plane, from the plane to the point. We can also rewrite the equation of the plane as (a,b,c).(x,y,z)=-d. We know that the normal has direction (a, b, c) since for any points P, Q on the plane, (P-Q).(a,b,c)=-d-(-d)=0 and also it is visually obvious that P-x is perpendicular to the normal of the plane. So the distance from any point R to the plane is the length of the component R-P in the direction of n, where n is normal to the plane and P is a point on the plane. This means it is given by $\frac{|n.(R-P)|}{|n|}$ but n.P is -d so it is given by $\frac{|n.R-d|}{|n|}$, and the result follows.

Matrix transpose relation

We want to show that $(AB)^T = B^T A^T$ for matrices A and B. This is true because the row i, column j element of AB is given by doing (row i of A).(column j of B) and so the row j, column i element of $(AB)^T$ is given by that same formula. The row j, column i element of the right hand side of the equation above is equal to (row j of B^T).(column i of A^T) which by the definition of the transpose is indeed equal to (row i of A).(column j of B). The matrices are equal element-wise so they are equal.

Cross product properties

We want to show that:

- AxB is perpendicular to A and B
- AxB=-(BxA) (anticommutitivity)
- Ax(B+C)=AxB+AxC (distributivity)
- |AxB|=area of parallelogram spanned by A and B

It will be helpful to think of the cross product as the determinant of the following matrix, since if you actually compute this determinant you can see that it is equal to the formula for the cross product that we assume is true and want to prove these properties from. (Note: i, j, k are the standard unit vectors, and the subscript denotes the coordinate)

$$AxB = det\begin{bmatrix} i & A_x & B_x \\ j & A_y & B_y \\ k & A_z & B_z \end{bmatrix}$$

This proves AxB=-(BxA) as swapping two columns changes the sign of the determinant, and that Ax(B+C)=AxB+AxC since the determinants are linear in columns as shown earlier. Now we can consider the determinant of the following matrix:

$$\begin{bmatrix} A_x & A_x & B_x \\ A_y & A_y & B_y \\ A_z & A_z & B_z \end{bmatrix}$$

Clearly it is zero because it has two columns the same, but another way to see this is to consider expanding it out by the first column, ie the determinant is

$$A_x(A_yB_z - B_yA_z) + A_y(B_xA_z - A_xB_z) + A_z(A_xB_y - B_xA_y)$$

But notice, this is actually equal to A.(AxB), and so we have that A.(AxB)=0. Applying the same argument to B gives that AxB is perpendicular to both A and B.

To prove the thing about area, we notice that the sum of the squares of the components of the cross product is the square of the cross product by pythagoras, and that value can also be shown (by a direct computation) to be given by this formula:

$$det\begin{bmatrix} A_yB_z - B_yA_z & A_x & B_x \\ B_xA_z - A_xB_z & A_y & B_y \\ B_xA_z - A_xB_z & A_z & B_z \end{bmatrix}$$

So now picture this geometrically: We have a vector AxB perpendicular to vectors A and B and we want to find the length of AxB. We know that the volume of the parallelopiped spanned by A, B and AxB (By the volume property of the determinant proven earlier and the determinant of the matrix above) equals the square of the magnitude of the cross product, but the volume of this parallelopiped also equals the magnitude of AxB times the area of the parallelogram spanned by A and B. Cancelling a factor of |AxB| on both sides gives the desired result.

Symmetric matrices stuff

How do we know that the normalized eigenvector matrix of a symmetric matrix has an inverse equal to its transpose? That's what we'll answer here. It is sufficient to show that the eigenvectors are perpendicular, as then we see that if the eigenvalue matrix is A then $A^TA = AA^T = I$ by a dot product argument, so the transpose of A is the inverse of A. Here we assume that the symmetric matrix has real values, otherwise the claim about the transpose being the inverse, the more general result that works for complex numbers is about Hermitian matrices, where if you take the complex conjugate of the transpose of A you get the inverse of A, and for real numbers taking the complex conjugate simply does nothing, and the proof for that result goes exactly like the proof shown below.

If S is our symmetric matrix and v is an eigenvector with eigenvalue λ then by the definition of eigenvectors we have that Sv= λv.

So, suppose we have eigenvalues λ and μ (which are not equal to eachother, this is another assumption used in the mathematical statement we are proving) with eigenvectors v and u respectively, then we have

$Sv = \lambda\mathrm{v}$ (1) and $Su = \mu\mathrm{u}$ (2). We can multiply both sides of (1) on the left by the row vector u^T and both sides of (2) on the left by the row vector v^T. This will give $u^TSv = \lambda u^T\mathrm{v}$ and $v^TSu = \mu v^T\mathrm{u}$, where the constants can be moved out to the front. Now we will take the transpose of both sides of this last equation so we get $(v^TSu)^T = \mu(v^Tu)^T$ and then use the relation about matrix transposes from earlier to simplify this to $u^TS^Tv = \mu u^T\mathrm{v}$. It is here that we use the assumption that S is symmetric, as by definition we have that $S = S^T$ so $u^TSv = \mu u^T\mathrm{v}$.

Since we have $u^TSv = \lambda u^T\mathrm{v}$ from earlier as well, it means $\mu u^T\mathrm{v} = \lambda u^T\mathrm{v}$ as if they are both equal to u^TSv they must be equal to eachother. Therefore, we have 2 options: Either $\lambda = \mu$ (so the eigenvalues would not be distinct as per our assumption), or $u^T\mathrm{v} = 0$. Because of how matrix multiplication works, finding $u^T\mathrm{v}$ is actually the same as taking the dot product u.v, so we are done and we have proven the perpendicularity as required.

Cayley hamilton theorem

We define Adj(M) as the transpose of the cofactor matrix. We have $\det(M)\, M^{-1} = Adj(M)$ if M is invertible, and we always have that $\det(M)\, I = MAdj(M)$.

We define B=Adj(tI-A) where A is an nxn matrix with a characteristic equation in t that is det(tI-A)=0. Since the elements of an adjugate matrix are built by taking the original matrix and deleting a row and column and finding the determinant (which in the case of tI-A will be polynomials in t with a power of t not exceeding n-1), we know that the elements of B must be polynomials in t of degree up to n-1, so we can write B as

$\sum_{k=0}^{n-1} t^k B_k$ for matrices B_k with coefficients not depending on t.

We know that det(tI-A)I=(tI-A)Adj(tI-A) so $\det(tI - A)\, I = (tI - A) \sum_{k=0}^{n-1} t^k B_k$

On the left hand side, we have $t^n I + c_{n-1} t^{n-1} I + c_{n-2} t^{n-2} I + \cdots + c_1 tI + c_0 I$ where the c's are the coefficients of the characteristic polynomial of A and on the right hand side we have, by expanding,

$t^n B_{n-1} + t^{n-1}(B_{n-2} - AB_{n-1}) + t^{n-2}(B_{n-3} - AB_{n-2}) + \cdots + t(B_0 - AB_1) - AB_0.$

$$= t^n B_{n-1} + \sum_{k=1}^{n-1} (t^k (B_{k-1} - AB_k)) - AB_0$$

Now we equate the coefficients of powers of t from our two equations. We get that

$$I = B_{n-1}$$

And that

$$c_k I = B_{k-1} - AB_k$$

For k from 1 to n-1, and

$$c_0 I = -AB_0$$

Now by multiplying both sides of each equation by A^n, A^k and nothing respectively we get

$$A^n = A^n B_{n-1}$$

$$c_k A^k = A^k B_{k-1} - A^{k+1} B_k$$

$$c_0 I = -AB_0$$

So we have that

$$A^n + c_{n-1} A^{n-1} + \cdots + c_1 A^1 + c_0 I =$$

$$A^n B_{n-1} + A^{n-1} B_{n-2} - A^n B_{n-1} + A^{n-2} B_{n-3} - A^{n-1} B_{n-2} + \cdots + AB_0 - A^2 B_1 - AB_0$$

The latter expression is 0 because clearly every negative term cancels every positive term, and this is equal to the characteristic equation of A of A, where A is an arbitrary matrix. This completes the proof.

L'hopital's rule

I have a video on this which shows a visual argument for the 0/0 case and then at the end shows images of both the technical justification for that case and the proof of the other cases from that case.

A tricky integral problem

We know and can derive using partial fractions a formula that says that $\int \frac{1}{x^2-a^2}dx = \frac{1}{2a}\ln\left(\frac{x-a}{x+a}\right) + c.$ So, putting in a=i we know that $\int \frac{1}{x^2+1}dx = \frac{1}{2i}\ln\left(\frac{x-i}{x+i}\right) + c$. But this integral is also equal to arctan(x)+c. So we have two different answers. However, it turns out these are secretly equivalent. Another example of this is how the integral of 1/2x gives both ln(x)/2+c and ln(2x)/2+c. The resolution to this second example is that the two answers actually differ by ln(2)/2 by logarithm properties, which is a constant. We can do the same for the first example: Since $\left|\frac{x-i}{x+i}\right| = \frac{x^2+1}{x^2+1} = 1$ if x is real,

$$\ln\left(\frac{x-i}{x+i}\right) = \mathrm{i} * \arg\left(\frac{x-i}{x+i}\right) = \mathrm{i} * \arg\left(\frac{ix+1}{ix-1}\right) = \mathrm{i} * \arg(ix+1) - \mathrm{i} * \arg(ix-1)$$

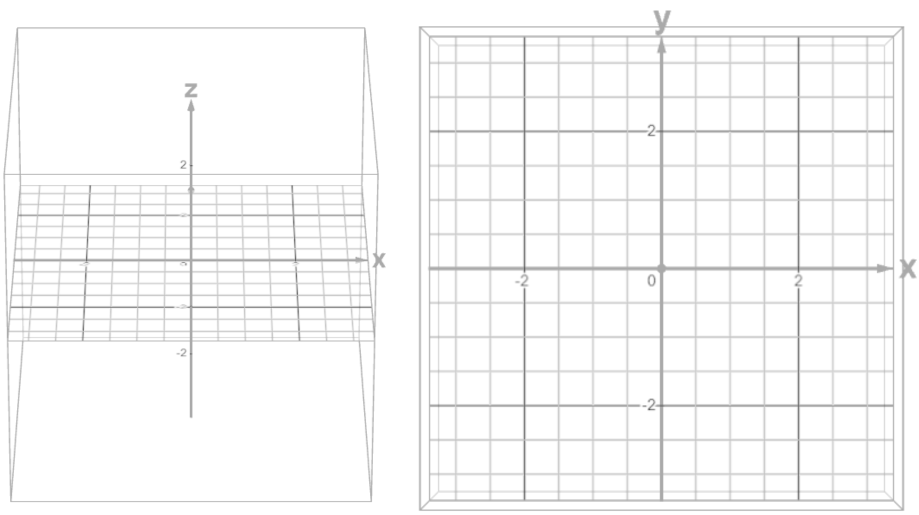
$$= \mathrm{i} * \arctan(x) - \mathrm{i} * \arctan(-\mathrm{x}) = 2\mathrm{i}\arctan(x)$$

So $\frac{1}{2i}\ln\left(\frac{x-i}{x+i}\right) = \arctan(x).$

So, two different answers could be secretly equivalent: This often happens when trig identities involved. Note that if |x|=1 then x can be written as $e^{iarg(x)}$, and arg(ix+1)=arctan(x) because this can be shown geometrically.

Conic section properties

They are called conic sections because you get them from slicing a cone, but typically at A level this is not proven. If a cone has equation $c_1 z^2 = x^2 + y^2$ and a plane has equation $x + c_2 z = 1$ (where we have rotated and scaled the figure so that the plane intersection with z=0 closest to the origin has been standardized to the point (1,0) in the xy plane). Then we can make a substitution (specifically we rearrange the second equation for z and substitute that into the first equation): $A(1-x)^2 = x^2 + y^2$ where A is some constant. If a=1 this cancels to give $y^2 = 1 - 2x$ which is a parabola. This means a parabola is what you would see if you were to look at the plane from the z axis (ie, looking down from a point like (0, 0, 10000)). If you were to look at the plane from its perpendicular axis, it would stretch out, below is an illustration of what I mean, we see that if we ignore perspective issues and act like we are simply seeing a projection, the plane looks like a contracted version of itself when viewed from an axis not perpendicular to it.



Images: Shows a plane viewed from above and from the side, with its gridlines, so you can see my point.

Stretching a parabola still gives a parabola, if the x or y is scaled it's still a parabola.

Similarly, if A<1, we get an ellipse equation, if A>1 we get a hyperbola equation. In both cases the stretching argument still applies. Therefore we know that all three conic sections are gotten by slicing a cone.

Also, you can consider a parabola to be like a "limit" of an ellipse as it gets more and more stretched out, by slowly tilting the plane intersecting the cone so A is 0.99, 0.999, 0.9999, etc

We can also easily verify, using trig identities, that the parametric equations for an ellipse, parabola, or hyperbola satisfy the cartesian equations for them. The only thing we need to be careful of is that if we use (Acosh(t), Bsinh(t)) as the parametric equation for a hyperbola, then cosh can only take positive values if t is real, so we should either allow for complex t or use tan and sec.

Now we will prove that in an ellipse and hyperbola, the sum and differences of the distances of any point on the curve from the foci respectively is constant, this is the defining feature of the foci. For an ellipse we start by assuming the sum of the distances from the foci at (-c, 0) and (c, 0) is constant and equal to 2a with a>c then we prove that the equation we get is the equation for an ellipse.

$$\sqrt{(x+c)^2+y^2}+\sqrt{(x-c)^2+y^2}=2a$$
$$\sqrt{(x+c)^2+y^2}=2a-\sqrt{(x-c)^2+y^2}$$
$$(x+c)^2+y^2=\left(2a-\sqrt{(x-c)^2+y^2}\right)^2$$
$$x^2+2cx+c^2+y^2=4a^2-4a\sqrt{x^2-2cx+c^2+y^2}+x^2-2cx+c^2+y^2$$
$$4cx=4a^2-4a\sqrt{x^2-2cx+c^2+y^2}$$
$$a\sqrt{x^2-2cx+c^2+y^2}=a^2-cx$$

$$a^2((x-c)^2+y^2)=a^4-2a^2cx+c^2x^2$$
$$a^2(x^2-2cx+c^2+y^2)=a^4-2a^2cx+c^2x^2$$
$$a^2x^2+a^2c^2+a^2y^2=a^4+c^2x^2$$
$$x^2(a^2-c^2)+a^2y^2=a^2(a^2-c^2)$$
$$\frac{x^2}{a^2}+\frac{y^2}{a^2-c^2}=1$$

Note: the steps where we square both sides can be reversed since we always assume both sides to be positive. Therefore we have equivalence between the equation for the ellipse in "pythagorean" form and focus distance form.

In a hyperbola we do nearly the identical work. You should be able to figure it out by following the same steps as in the ellipse proof, but here it is anyway.

$$\sqrt{(x+c)^2+y^2}-\sqrt{(x-c)^2+y^2}=2a$$
$$\sqrt{(x+c)^2+y^2}=2a+\sqrt{(x-c)^2+y^2}$$
$$(x+c)^2+y^2=\left(2a+\sqrt{(x-c)^2+y^2}\right)^2$$
$$x^2+2cx+c^2+y^2=4a^2+4a\sqrt{x^2-2cx+c^2+y^2}+x^2-2cx+c^2+y^2$$

$$4cx = 4a^2 + 4a\sqrt{x^2 - 2cx + c^2 + y^2}$$

$$a\sqrt{x^2 - 2cx + c^2 + y^2} = cx - a^2$$

$$a^2\left((x-c)^2 + y^2\right) = c^2x^2 - 2a^2cx + a^4$$

From here, the steps are the same, except now the equation we get at the end is a hyperbola equation because c>a.

In a parabola $y^2 = 4ax$ the distance fron the focus (a,0) to a point on the parabola is given by $\sqrt{(x-a)^2 + y^2}$, and $x = \frac{y^2}{4a}$ so we have $\sqrt{(\frac{y^2}{4a} - a)^2 + y^2}$. We can simplify this to $\sqrt{\frac{y^4}{16a^2} - \frac{y^2}{2} + a^2 + y^2}$ which is $\sqrt{\frac{y^4}{16a^2} + \frac{y^2}{2} + a^2}$ which is $a + \frac{y^2}{4a}$ which is a+x which is the distance from the relevant point on the parabola to the directrix.

Now, we prove properties of eccentricity. Suppose that the eccentricity of an ellipse or hyperbola is the value of e such that $x^2 + \frac{y^2}{1-e^2} = 1$, then we will prove that e is both the ratio between the distance between the foci and the length of the long diameter of the ellipse or the distance between the parts of the hyperbola, and the ratio of the distance of a focus and the corresponding directrix. For the first one, note that if we consider e to be $\frac{c}{a}$ in the proofs above the final equations can be simplified to be $x^2 + \frac{y^2}{1-e^2} = 1$ and also that by picking the point on the curve in question to be a point on the x axis it becomes clear that 2a, our constant distance is indeed twice the the length of the long diameter of the ellipse or the distance between the parts of the hyperbola.

For the other property, if we have a point on the curve, we let d_1 be the distance from that point to the focus and d_2 be the distance from that point to the corresponding directrix. Note also that even if the denominator of x is not 1 both properties hold by scaling. We have that $d_1 = \sqrt{(x-e)^2 + y^2}$ by pythagoras and that $d_2 = x - \frac{1}{e}$. Substituting in $y^2 = (1-x^2)(1-e^2)$ which comes from rearranging the equation for the curve we get that $d_1 = \sqrt{(x-e)^2 + (1-x^2)(1-e^2)} = \sqrt{x^2 - 2xe + e^2 + 1 - x^2 - e^2 + x^2e^2} = \sqrt{-2xe + 1 + x^2e^2} = xe - 1 = ed_2$ as required.

Recurrence relations

We will not really do any work here because the parts of this that are used as a black box in some specifications can be proven using the same ideas as the differential equations from earlier. Perhaps this is unsurprising since the methods are very similar for solving both types of problems. If you use the analagous substitution for the second order case, then you will get the right complementary function, and the same result that says the thing about the general solution being complementary function plus particular solution also applies and can be proven the same way. It remains to verify the results in the tables of particular solutions given in some textbooks. However, it is easy to verify that you will get a solvable equation to solve for the constants as a result of plugging them in, and anyway using educated guesses and substitutions is more proper than memorizing tables.