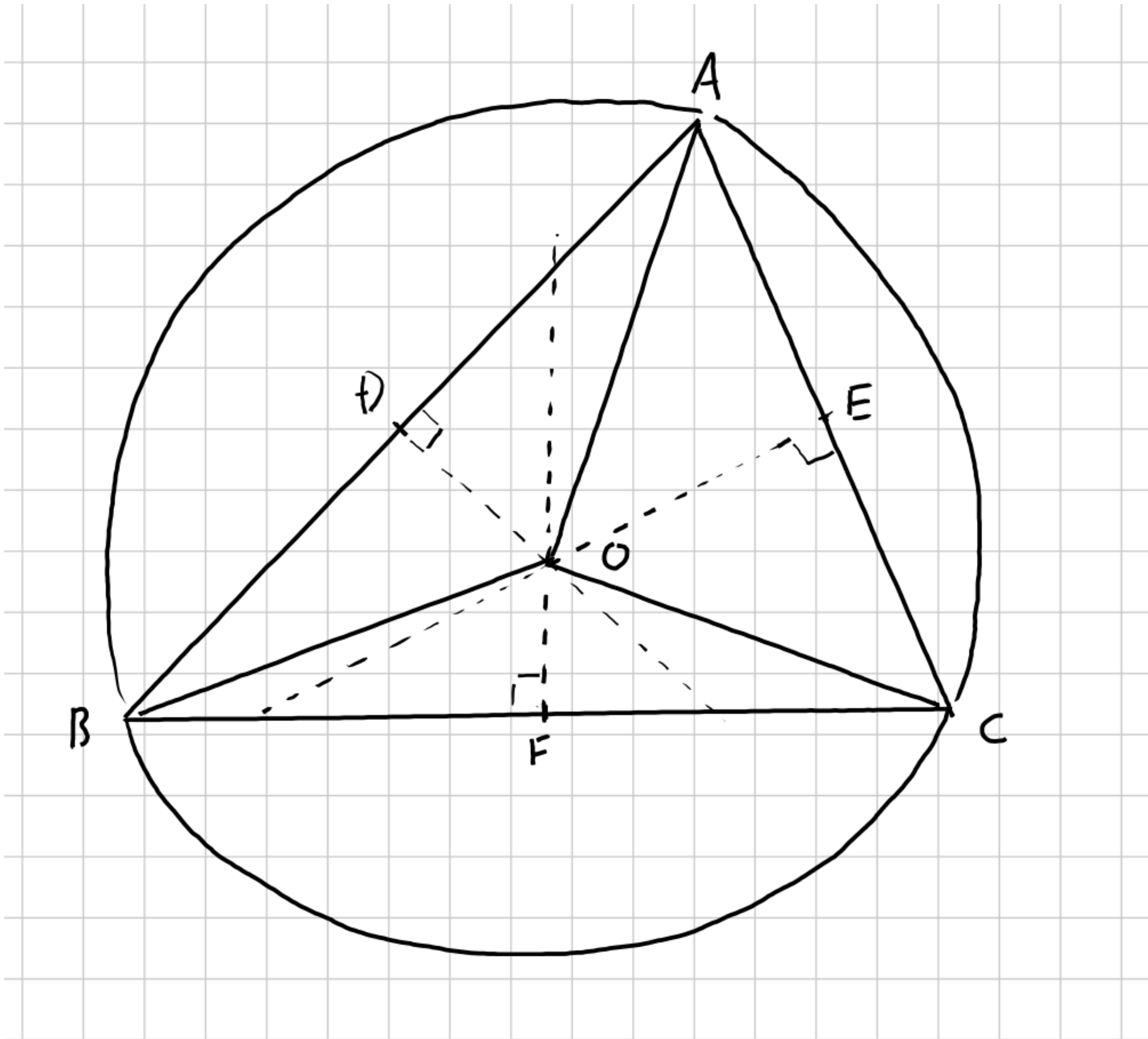
Connecting the Dots

Being able to draw a mathematically defined curve through a set of points is an ancient problem that has been pondered countless times. From Euclid to the recent breakthroughs by mathematicians Eric Larsen and Isabel Vogt on finding a curve through an arbitrary number of points in a set dimension, interpolation has a rich and unique place in mathematics. There are, of course, applications beyond pure mathematical enjoyment. Many subjects, such as engineering, physics and economics, use curve fitting to predict accurate trajectories for a system. Even computer graphics can require this branch of math to give birth to realistic images.

Euclid’s elements

The simplest form of interpolating data points is when there are only two points. The earliest recording of an interpolation between two points is given in the first proposition of the first book of Euclid’s Elements.

Logically, the next step would be to increase the number of dots. There is another solution from Euclid’s Elements concerning any three non-collinear points on a plane. Euclid suggests in Proposition 5 of Book 4 that for any triangle, a circle can be drawn around it, touching all three vertices.



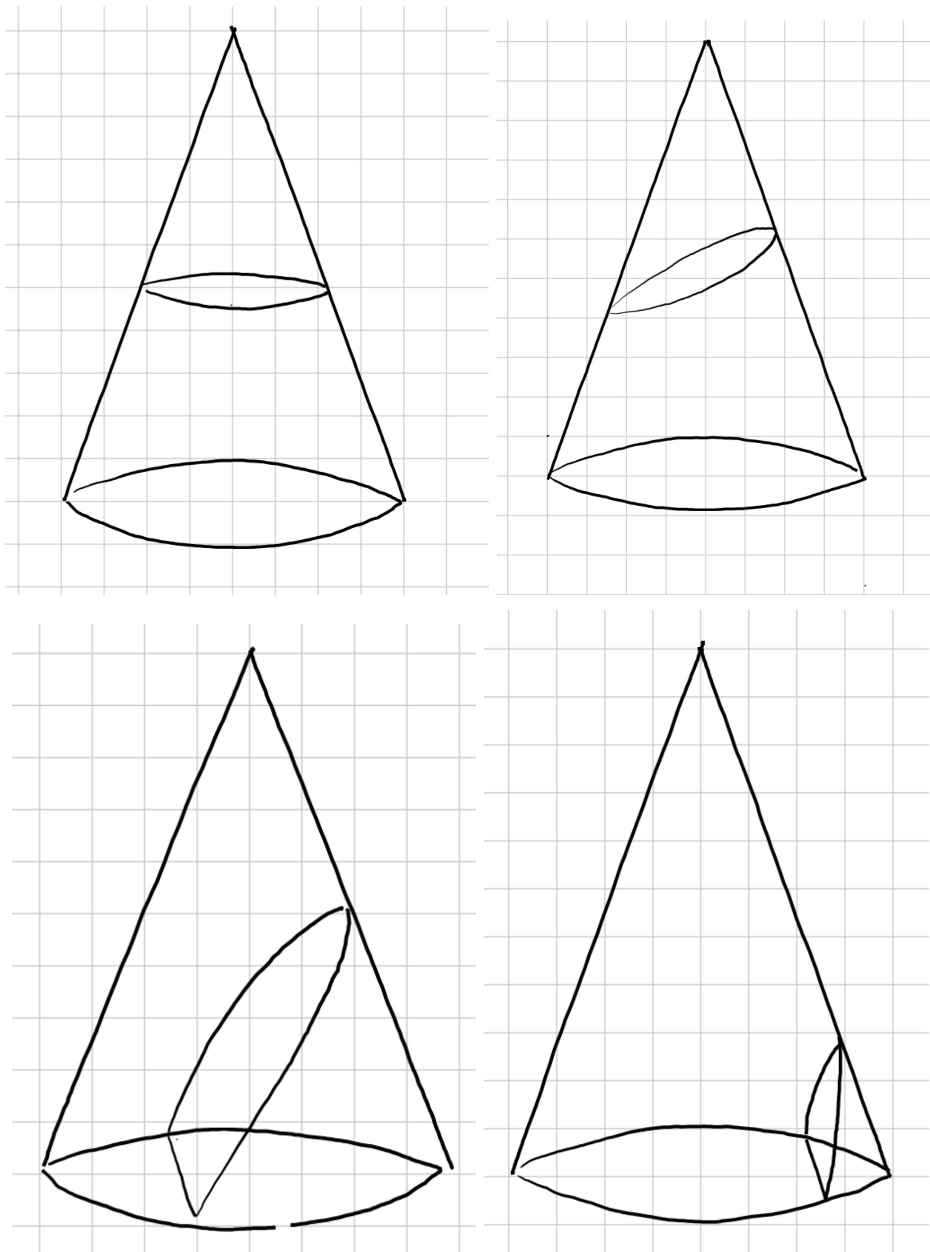
This is proven by the fact that the centre point of the triangle, also known as the circumcentre, is constructed by drawing perpendicular bisectors on each side. This means that the point in the middle is at an equal distance from every corner. Therefore, a circle of constant radius can be drawn around the triangle.

Apollonius and Conic Sections

What inevitably became crucial to the work produced today for curve fitting was the study of different curves represented by cross sections of a cone. By admiring the beauty of curves produced from a cone, the ancient Greeks laid the groundwork for future mathematicians to discover many methods to describe curves between points on a plane.

The circle is the first and most basic smooth shape that can be made from a cone cross-section.

This occurs when you cut in parallel to the base of the cone. The second is the ellipse, which happens when you cut the cone diagonally but not through the base. The third is a parabola, which occurs when you cut through the cone diagonally and through the base. The fourth is the hyperbola, created by cutting the cone perpendicular to the base.



All four of these shapes fit a general formula for the cross-sectional shape of a cone. To derive this form, we first have to look at the distance formula, which is essentially a form of Pythagoras's theorem on right-angled triangles:

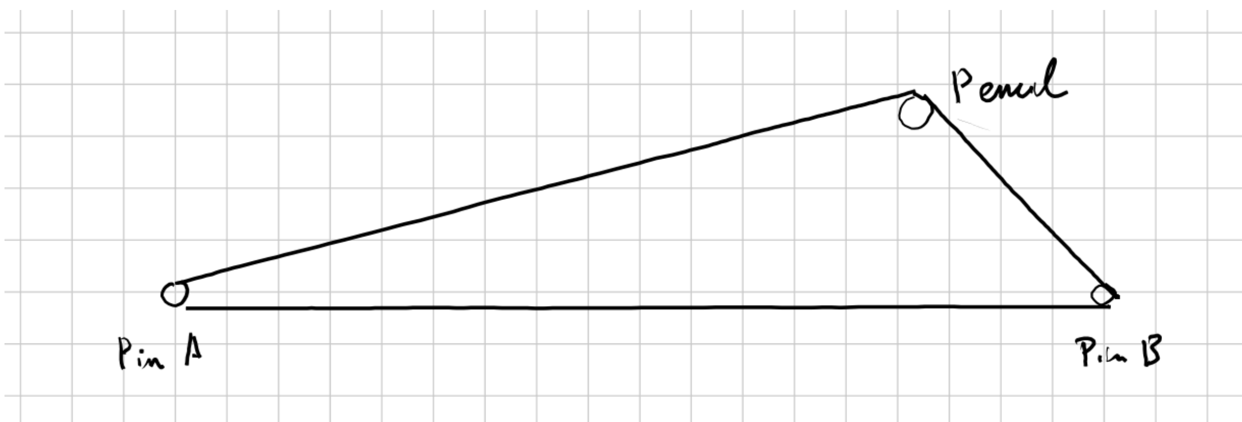
$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

The derivation for the general form of a circle is pretty simple. Since the radius is constant, the distance from the circle's centre to the function value must also be constant. Therefore, the value of d will always equal the radius. Replacing the x_1 and y_1 with the respective coordinates for the centre point and then squaring both sides, the general equation of the circle is:

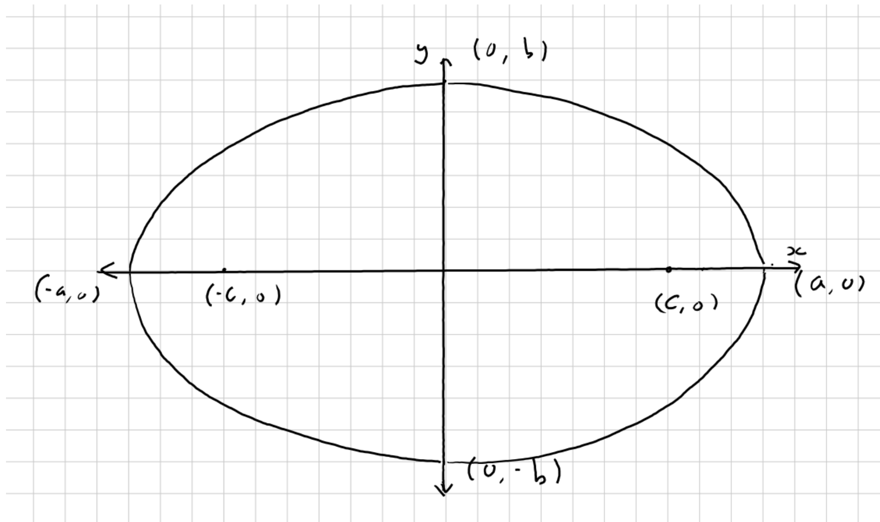
$$r^2 = (x - h)^2 + (y - k)^2$$

However, the distance for an ellipse is not constant. The ellipse, however, has some interesting properties that allow us to derive its general form. The one that is most useful for deriving an algebraic form to describe an ellipse is that the sum of the distances of the function from the two foci should always be constant.

This can be imagined by taking a piece of string and tying it around two pins that are placed firmly onto a board. An ellipse is formed when you use a pencil to draw the maximum area that the string restricts.



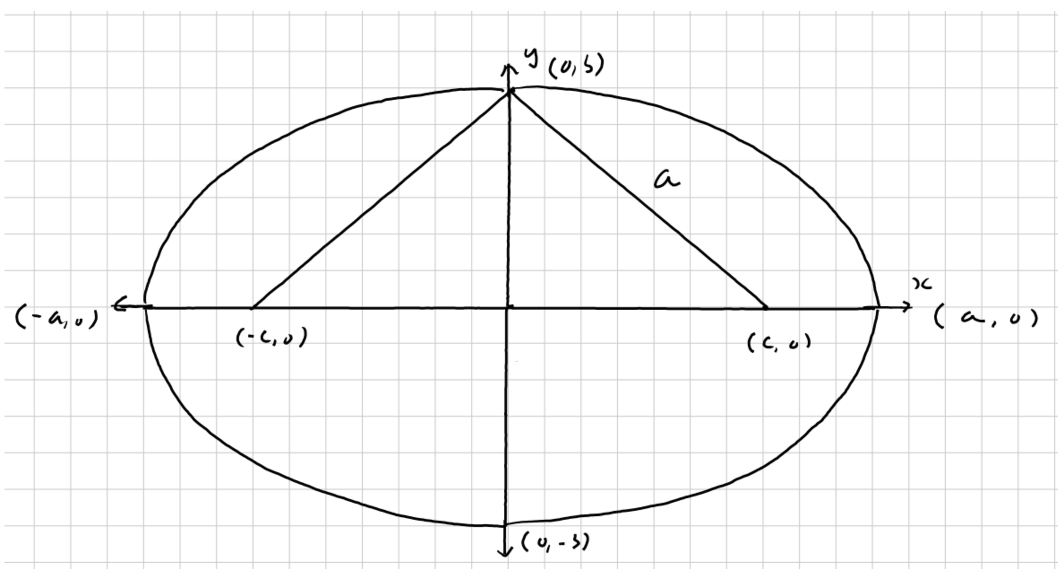
Using this fact, we can construct an equation for an ellipse. Let's say that the foci of the ellipse, i.e. the pins, are defined by (-c,0) and (c,0). The function's maximum and minimum y values are b and -b, and the maximum and minimum values for x are -a and a.



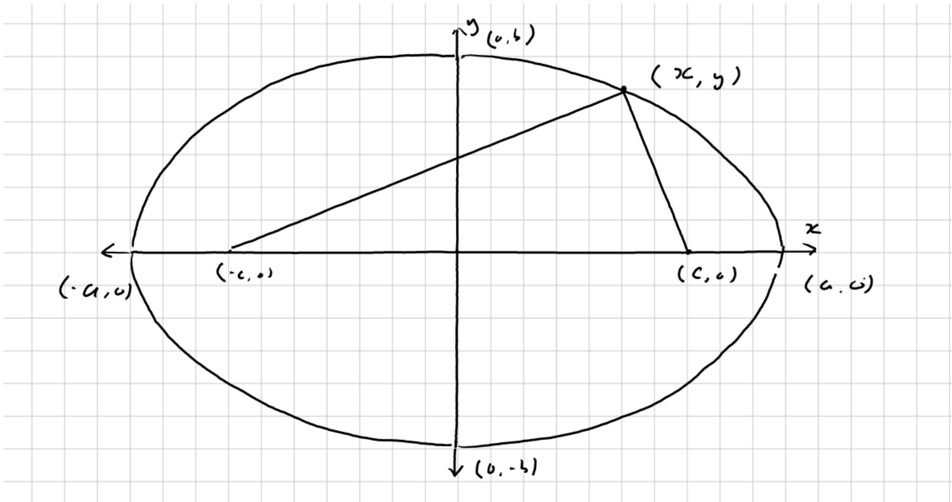
This diagram can be used to deduce two things. First, the length of the string must be 2a. Second, the values of a,b, and c are a Pythagorean triple.

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

This can be observed when considering the maximum y value of the function at the point (0,b). If we draw a triangle joining (0,b) and (c,0), a right-angle triangle is formed whose side length is a since the distance from (0,b) to either focus is the same. Moreover, as a right-angled triangle, we end up with the Pythagorean triple relating the foci, the maximum and minimum y values and the maximum and minimum x values.



Now, we can form an equation for an ellipse using this relation and taking an arbitrary point. First, we can relate the distances from the foci to the arbitrary point and the constant length of the string.



After a series of calculations, we produce the desired equation describing an ellipse.

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

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

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

$$0 = 4a^2 - 4a\sqrt{(x-c)^2+y^2} + (x+c)^2 - (x-c)^2$$

$$0 = 4a^2 - 4a\sqrt{(x-c)^2+y^2} + 4xc$$

$$0 = a^2 - a\sqrt{(x-c)^2+y^2} + xc$$

$$a\sqrt{(x-c)^2+y^2} = a^2 + xc$$

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

$$a^2x^2 + 2a^2xc + a^2c^2 + a^2y^2 = a^4 + 2a^2xc + x^2c^2$$

$$a^2x^2 + a^2c^2 + a^2y^2 = a^4 + x^2c^2$$

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

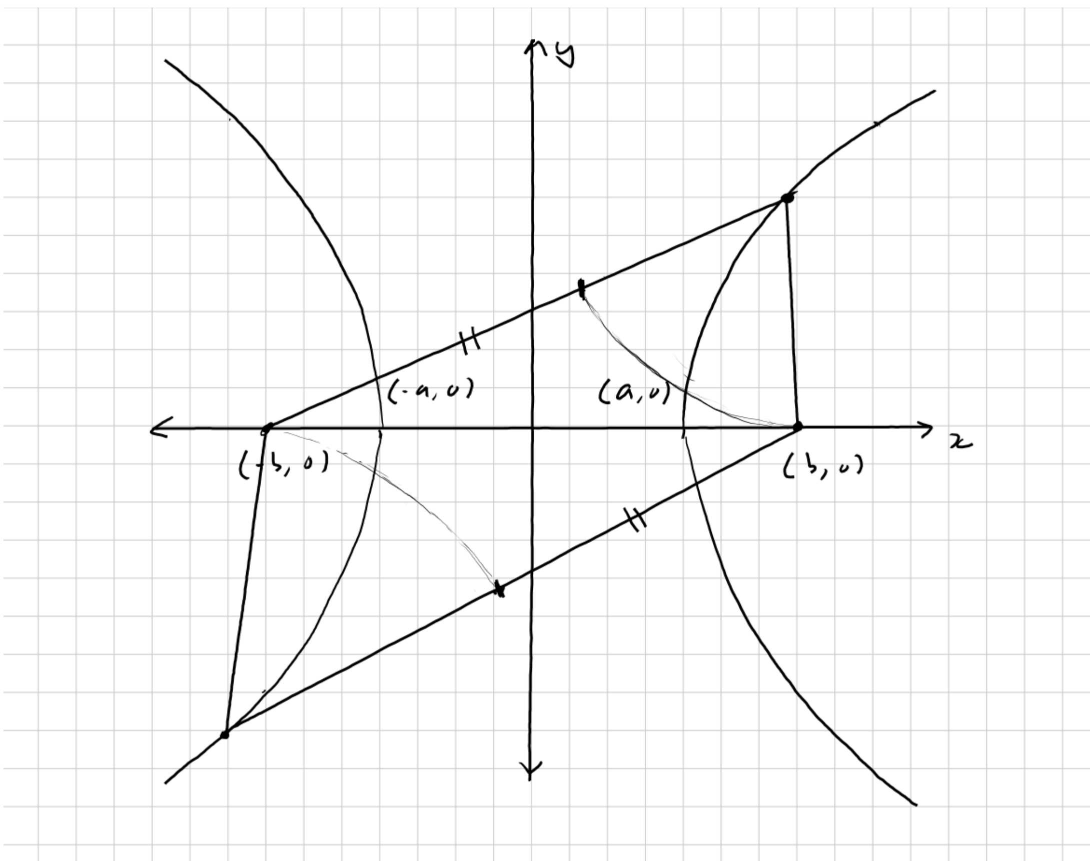
$$\frac{x^2}{a^2} + \frac{y^2}{a^2-c^2} = 1$$

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

The next conic section is the parabola. The general form of a parabola that intersects the origin and has a focus point at (0,p) is:

$$y = \frac{1}{4p}x^2$$

The last of the four conic sections is the hyperbola. What defines this shape? If we go back to the image of the conic section, it is easy to see that a hyperbola is defined by a set of points where the difference in distance to the foci is constant.



Using this property, an equation can be formed. Interestingly, the equation is of the same form as the ellipse.

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$

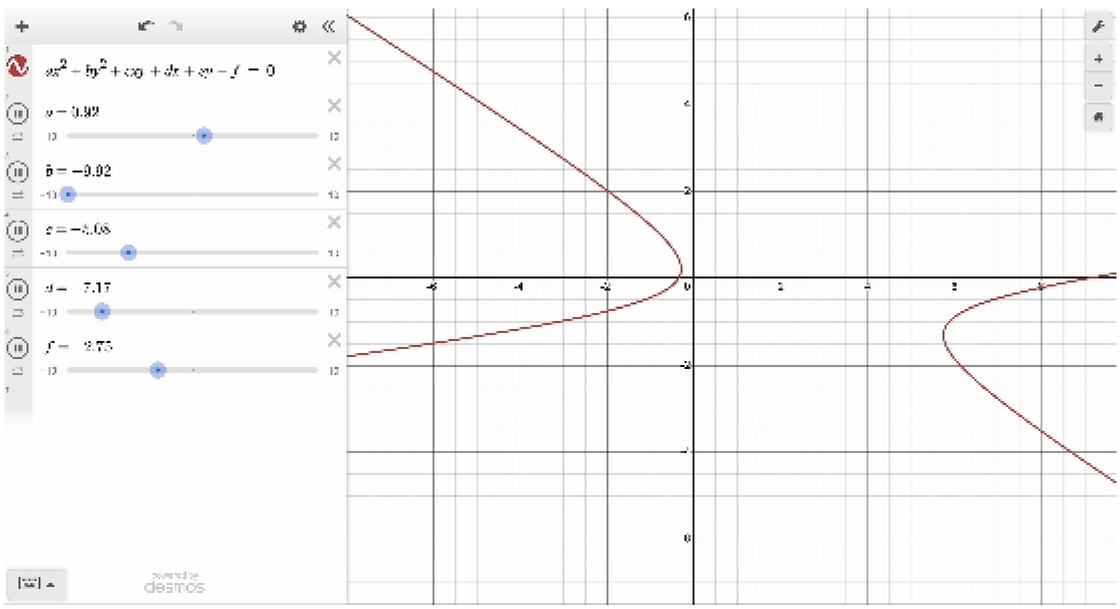
The asymptotes that these hyperbolic functions follow are given by:

$$y = \frac{b}{a}x \quad \text{and} \quad y = \frac{-b}{a}x$$

It is incredible and beautiful how a series of different shapes could be created from a simple cone; interestingly, they all have a similar form. The general form for a conic section is:

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$$

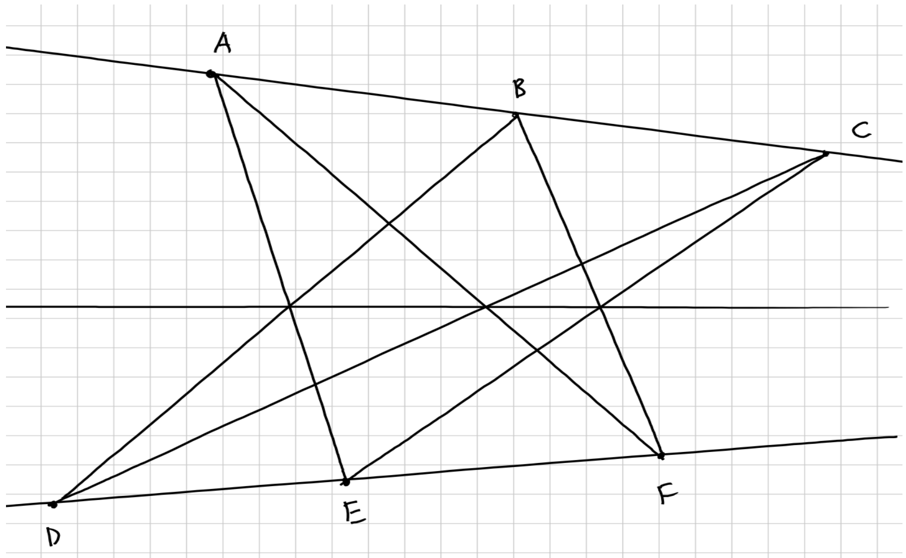
This can be simulated on Desmos, as shown below. By varying the values of A~E, the formed shape changes back and forth!



Describing a set of points with conic sections

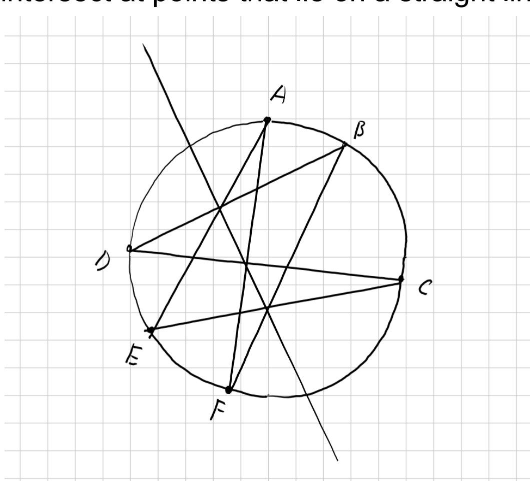
The general form of a conic section can be used to derive polynomials that describe the relation between multiple data points. However, since there is not enough information, a way of relating the points together other than through conic sections must exist. This is where the collinearity of intersections can help.

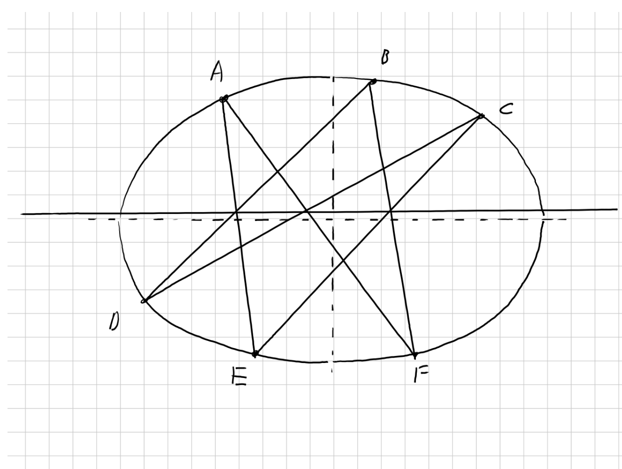
Pappus's Theorem states that given two sets of 3 colinear points, e.g. {A, B, C} and {D, E, F}, the intersections of AE and DB, AF and DC, BF and CE are collinear.

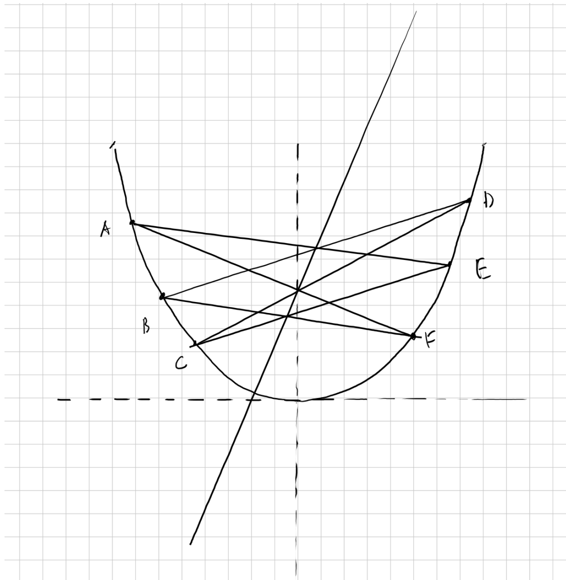


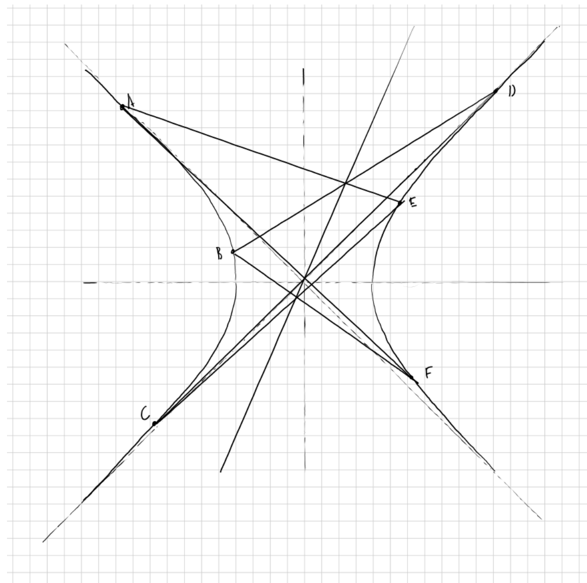
The fact that the three points are collinear will become very important when defining a conic section that goes through a set of 6 data points. The case of having two straight is a degenerate case since there are two sets of collinear points, making some of the points unnecessary in portraying the function.

Pascal generalised this rule to sets of 6 points on conic sections. Known in Latin as the hexagrammum mysticum theorem, Pascal's theorem states that if 6 points lie on a conic and are connected by line segments that form a hexagon, then the three pairs of opposite sides intersect at points that lie on a straight line called the Pascal line.









This can be proved using the Cayley Bacharach theorem. The theorem asks how many points are necessary to determine a polynomial. This is the reverse of the original question of finding an elegant polynomial that relates a set of points and may give further insight into how a set of points may be interpolated. The Cayley Bachrach theorem states that given a set of 9 points and two cubics that go through all nine, if a third curve goes through 8 of those points, it must also go through the 9th point.

The converse of this theorem is known as the Braikenridge-Maclaurin theorem and states that if the three points are collinear, the shape on which the points lie is a conic section. This is important to note because this implies that if a set of points is colinear, then you can interpolate through the points using a conic section.

Given a set of points, how would one develop a conic that goes through all points? The commonly used method is to put the data points into a system of equations of the general equation for a conic section. A matrix form is perfect for this purpose.

$$
\begin{matrix}
\mathrm{A}x_1^2 & \mathrm{B}x_1y_1 & \mathrm{C}y_1^2 & \mathrm{D}x_1 & \mathrm{E}y_1 & 1 \\
\mathrm{A}x_1^2 & \mathrm{B}x_1y_1 & \mathrm{C}y_1^2 & \mathrm{D}x_1 & \mathrm{E}y_1 & 1 \\
\mathrm{A}x_2^2 & \mathrm{B}x_2y_2 & \mathrm{C}y_2^2 & \mathrm{D}x_2 & \mathrm{E}y_2 & 1 \\
\mathrm{A}x_3^2 & \mathrm{B}x_3y_3 & \mathrm{C}y_3^2 & \mathrm{D}x_3 & \mathrm{E}y_3 & 1 \\
\mathrm{A}x_4^2 & \mathrm{B}x_4y_4 & \mathrm{C}y_4^2 & \mathrm{D}x_4 & \mathrm{E}y_4 & 1 \\
\mathrm{A}x_5^2 & \mathrm{B}x_5y_5 & \mathrm{C}y_5^2 & \mathrm{D}x_5 & \mathrm{E}y_5 & 1 \\
\mathrm{A}x_6^2 & \mathrm{B}x_6y_6 & \mathrm{C}y_6^2 & \mathrm{D}x_6 & \mathrm{E}y_6 & 1
\end{matrix}
$$

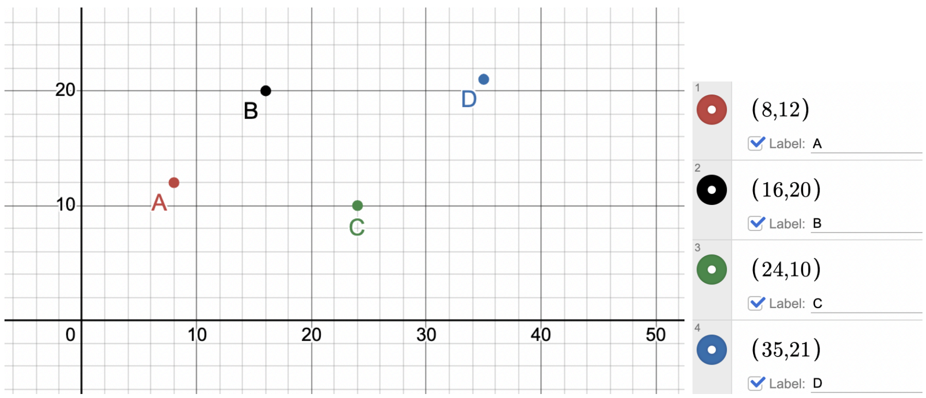
The Braikenridge-Maclaurin theorem shows that if the lines created by joining opposite sides create three collinear points, the points must lie on a conic section. Therefore, to find the conic section equation that unifies all points, we must first assume that the intersecting points are collinear. We can do this by thinking about the area created by the intersecting points. If they are collinear, the area must be 0! Thus, we can assume that the determinant of the matrix holding the equations of the conic section defined by each point will have a determinant of 0.
Since there are a maximum of 5 variables, i.e. A~E in a conic section formula, up to only 5 points are necessary. This means that anything above 5 is redundant data.

Beyond the Conic Section

So far, we have only examined methods of defining a curve that fits a set number of points. For example, a circle is a foolproof method of interpolating a set of three points. The Brackeridge-Maclaurin method is unsuitable for a set of data that is not representative of a conic section or has a number of elements of over five points.

Polynomial interpolation is the answer to this. The basic principle of any polynomial interpolation method is to find the curve that fits all set data points with the lowest possible degree. One thing to notice is that for any set of n+1 data points, a unique polynomial of degree n will pass through all these points. Take, for example, a set of 2 data points. The function of the lowest degree that interpolates the data points is a linear interpolation that produces a straight line of degree 1. For a set of 3 data points, a quadratic or a conic section can be drawn with a degree of 2. This is called the Riesz Thorin Interpolation Theorem, which is crucial for deriving some of the many interpolation methods. One such method is the Lagrange Interpolation Method.

Say you are given a set of nodes you want to interpolate into a single polynomial.



In Lagrange Interpolation, a set of polynomials is created for each node. Each polynomial is made so that the function value at the node's x value is 1 and 0 everywhere else. Constructing such a polynomial may be easy if you use the roots.

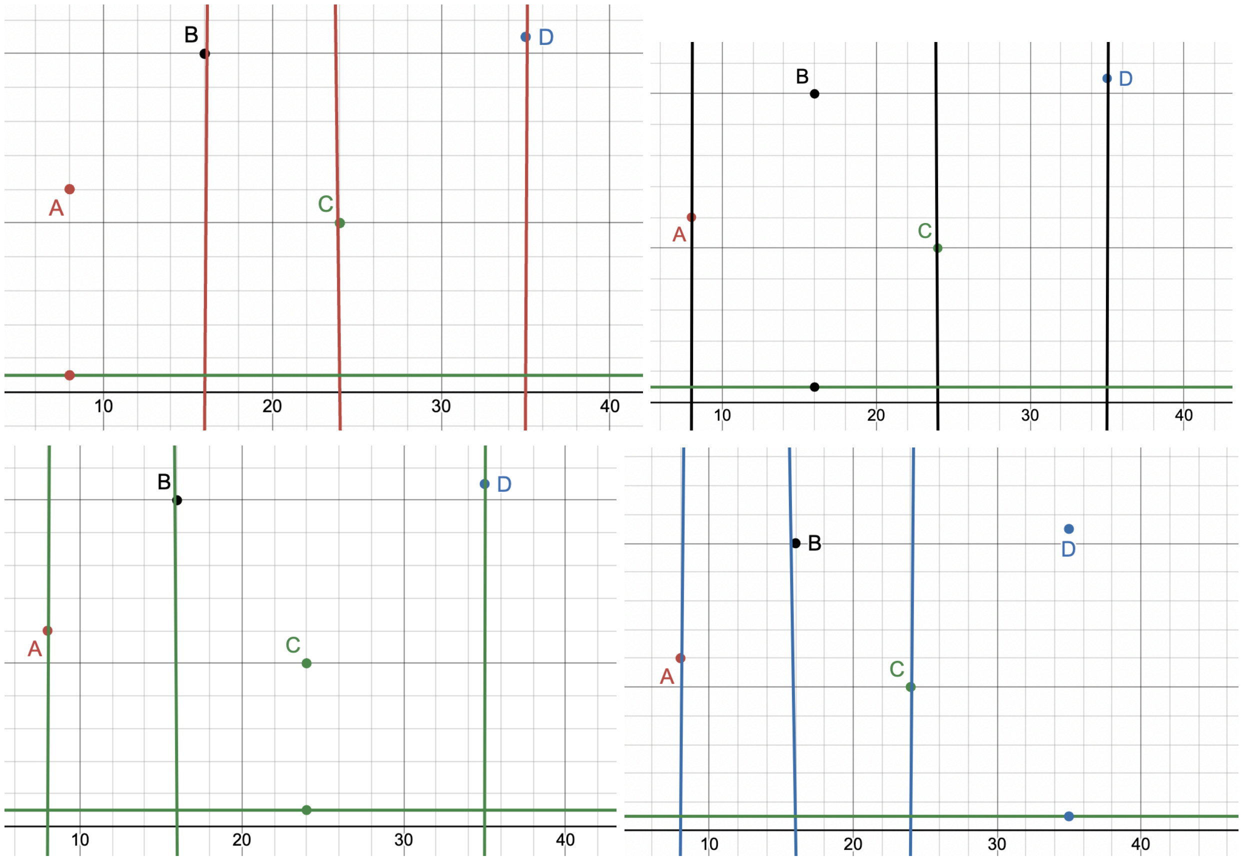
$$f(x) = (x - 16)(x - 24)(x - 35)$$

$$f(x) = (x - 8)(x - 24)(x - 35)$$

$$f(x) = (x - 8)(x - 16)(x - 35)$$

$$f(x) = (x - 8)(x - 16)(x - 24)$$

Since for each node, we want to construct a polynomial that intercepts the x-axis at the x value of every other node, the deconstruction above is logical. However, this produces a set of curves where the value of y at the given x value is not 1 for each node.



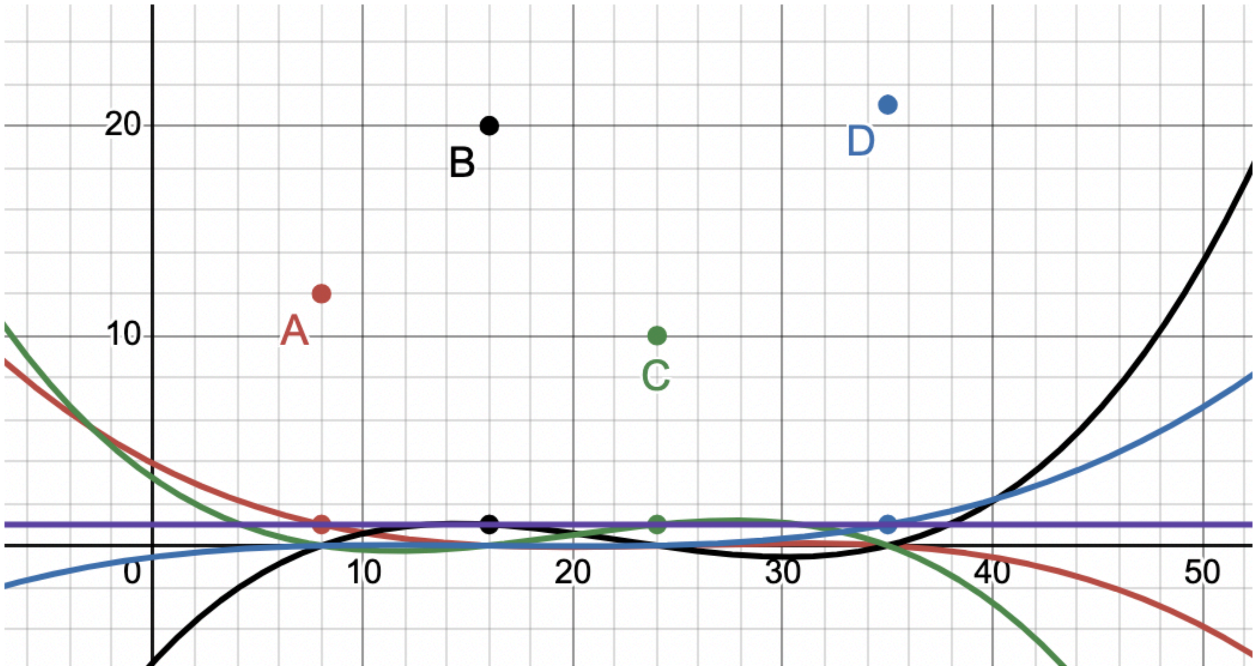
As visible in all four graphs, the functions do not pass through the desired y value when this approximate function is used. To overcome this, we need to change the function's magnitude when the value of x is the value of x at the node to 1. This can be done by dividing each function by the respective values of the function at their x values.

$$f(x) = \frac{(x-16)(x-24)(x-35)}{(8-16)(8-24)(8-35)}$$

$$f(x) = \frac{(x-8)(x-24)(x-35)}{(16-8)(16-24)(16-35)}$$

$$f(x) = \frac{(x-8)(x-16)(x-35)}{(24-8)(24-16)(24-35)}$$

$$f(x) = \frac{(x-8)(x-16)(x-24)}{(35-8)(35-16)(35-24)}$$



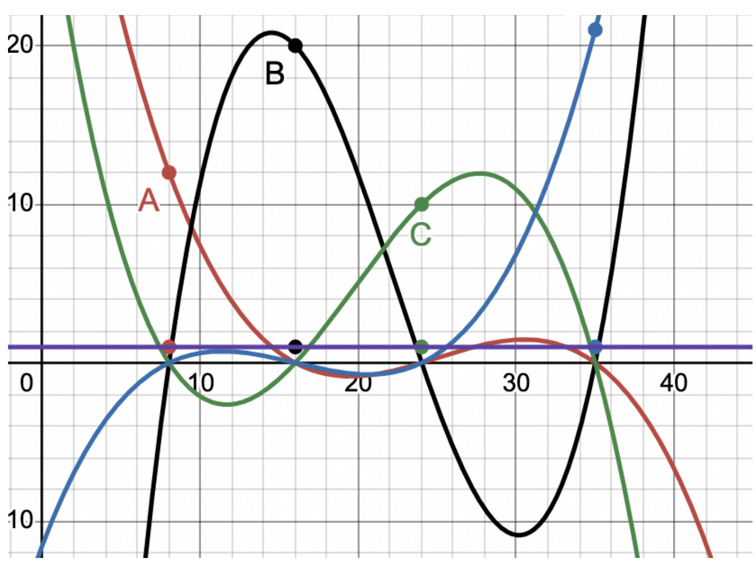
Now that we have made each function take the value of 1 at each node, the next step is to change the magnitude of the functions so that the y value matches the node's y value.

$$y=\frac{(x-16)(x-24)(x-35)}{(8-16)(8-24)(8-35)}\cdot 12$$

$$y=\frac{(x-8)(x-24)(x-35)}{(16-8)(16-24)(16-35)}\cdot 20$$

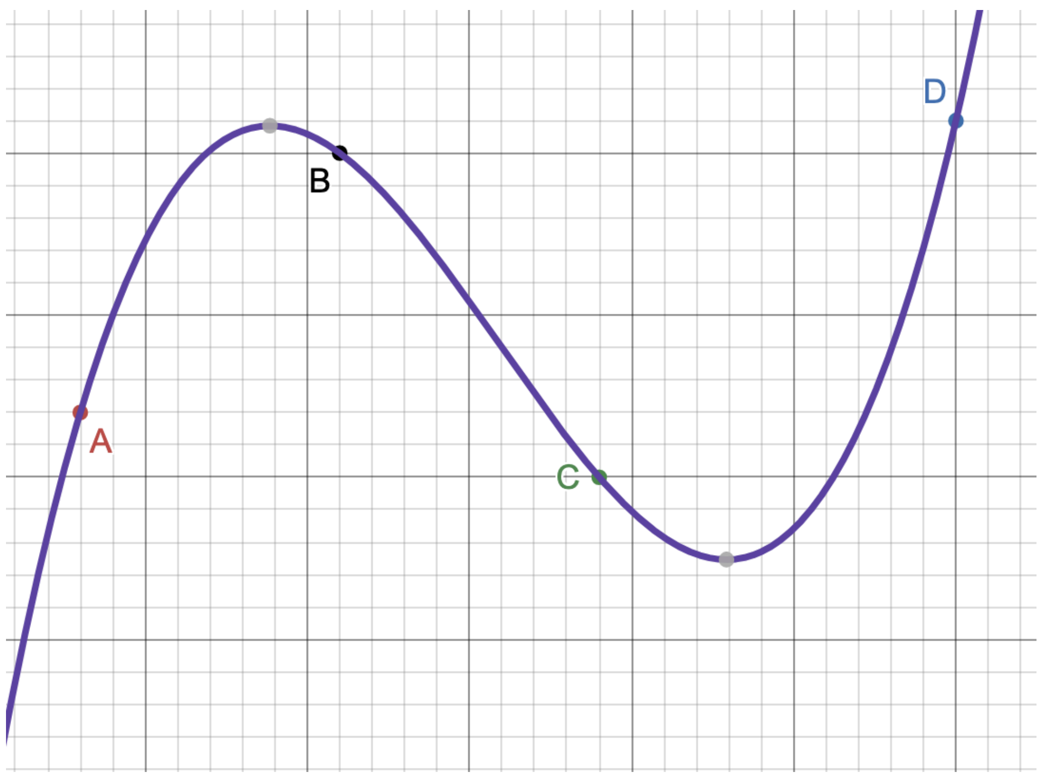
$$y=\frac{(x-8)(x-16)(x-35)}{(24-8)(24-16)(24-35)}\cdot 10$$

$$y=\frac{(x-8)(x-16)(x-24)}{(35-8)(35-16)(35-24)}\cdot 21$$



We can then superpose these waves together by taking a sum.

$$f(x) = 12\frac{(x-8)(x-24)(x-35)}{(8-16)(8-24)(8-35)} + 20\frac{(x-8)(x-24)(x-35)}{(16-8)(16-24)(16-35)}$$
$$+ 10\frac{(x-8)(x-16)(x-35)}{(24-8)(24-16)(24-35)} + 10\frac{(x-8)(x-16)(x-24)}{(35-8)(35-16)(35-24)}$$



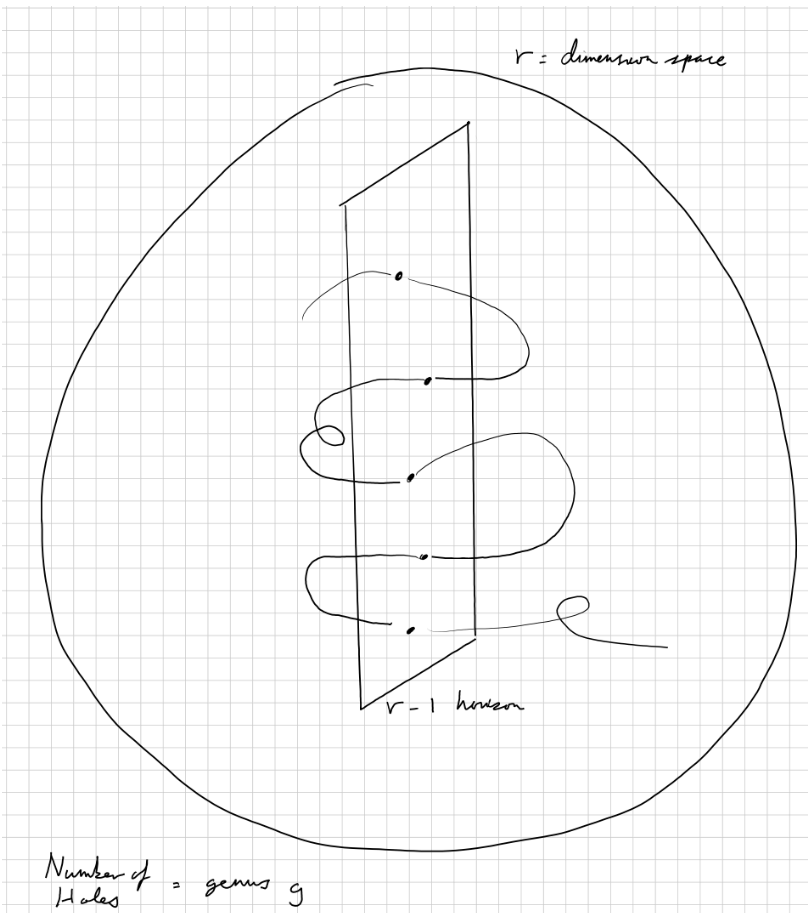
This creates a beautiful polynomial that goes through all the points! Formally, the method for producing such a polynomial given a set of data points can be provided by Sylvester's formula:

$$L(x) = \frac{(x-x_1)(x-x_2)\ldots(x-x_n)}{(x_0-x_1)(x_0-x_2)\ldots(x_0-x_n)}y_0$$
$$+ \frac{(x-x_0)(x-x_2)\ldots(x-x_n)}{(x_1-x_0)(x_1-x_2)\ldots(x_1-x_n)}y_1$$
$$+ \ldots$$
$$+ \frac{(x-x_0)(x-x_1)\ldots(x-x_{n-1})}{(x_n-x_0)(x_n-x_1)\ldots(x_n-x_{n-1})}y_n$$
$$= \sum_{i=0}^{n}\left(\prod_{0\le j\le n,\, j\ne i}^{\boxtimes} x - x_j x_i - x_j\right)y_i = \sum_{i=0}^{n}\frac{L(x)}{L'(x_i)(x-x_i)}y_i$$

There are other other methods for interpolation such as Gauss's method, the Stirling formula and the Bessel formula. They each have their own strengths and weaknesses but the Lagrange polynomial is the most efficient in terms of interpolating several data sets on the same data points.

Beyond 2D Interpolation

Up until now, I have only discussed methods to interpolate a set of points on a 2D plane. What if there was a way to connect dots in 3D, or even higher dimensions! When considering any curve, we need to first think of the dimensional space that it is in, the degree and the genus. The genus is a way of describing how many intersections the curve has with a plane of 1 dimension lower than what the curve is in.



The question that arises is, what is the maximum number of points that this curve can interpolate? The number and the method for such an interpolation was provided in 2023 by the two mathematicians Eric Larson and Isabel Vogt in their paper *Interpolation for Brill-Noether Curves*. Brill Noether curves are an interesting method of revealing the connections between algebraic surfaces and geometry through unexpected patterns. The theorem that arises when considering a Brill-Noether curve of degree d, genus g and dimension r, n general points can be interpolated in $\mathbb{P}^r$ (The Real Projective Space) if and only if

$$(r-1)n \leq (r+1)d - (r-3)(g-1)$$

except in the four following cases:

$$(d, g, r) \in \{(5, 2, 3), (6, 4, 3), (7, 2, 5), (10, 6, 5)\}$$

The reason why some of the curves don't hold for the equation is very complicated to explain and can be found in the research paper previously mentioned. A simple way to think of the reason why is by considering the surface that each of these curves live on. If the surface doesn't pass through the set interpolation points, then the curve itself won't either.

Conclusion

Although I still don't understand the proof fully for the breakthrough by Eric Larson and Isabel Vogt, I find it quite incredible how the history of interpolation has spread from Ancient Greece and will probably continue far into the future. Connecting the dots is a tricky puzzle that has continued and will continue to provoke curiosity from many mathematicians and I am glad that I came across this subject.

Not only does interpolation theory take a valuable space in the world of pure maths, but is greatly valued in the world of engineering and physics. From function approximation in computer graphics and digital cameras to computer aided design, interpolation theory is a fundamental pillar supporting much of the world around us today. Therefore, I encourage you reader to dive into some interpolation theory and appreciate the beauties and mystery that come with it!