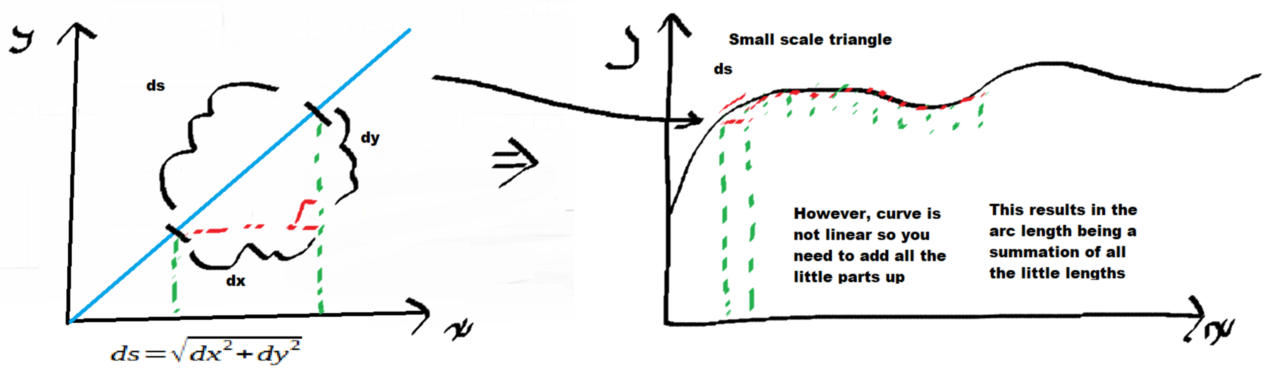
Insight into Elliptic Curves and use in Cryptography:

1. Origin

The origin of elliptic curves stems back to the 18th century. In this day, elliptic arc lengths were intriguing to many and required calculation, perhaps due to the arrival of machinery and advances in sciences. The arc length of an ellipse can be found using a combination of pythagoras and integration.


y
ds
dy
dx
x
$ds=\sqrt{dx^2+dy^2}$
⇒
y
Small scale triangle
ds
However, curve is not linear so you need to add all the little parts up
This results in the arc length being a summation of all the little lengths
x

$$\Delta s=\sqrt{\Delta x^2+\Delta y^2}$$

or:

$$ds=\sqrt{dx^2+dy^2}$$

This can be simplified as $dx \sim \frac{dx}{dx}$:

$$ds^2=dx^2+dy^2, \rightarrow \frac{ds^2}{dx^2}=1+dy^2 \rightarrow \frac{ds}{dx}=\sqrt{1+dy^2} \rightarrow ds=\sqrt{1+dy^2}\,dx$$

The arc length across a graph area can be represented as the summation from *r=1 to n=∞* as $\frac{dy}{dx}$ represents a change in y with respect to x*:*

$$\lim_{n\rightarrow\infty}\sum_{r=1}^{n}\sqrt{1+\frac{dy^2}{dx^2}}\,dx\,, arc\,length=\int_a^b\sqrt{1+\frac{dy^2}{dx^2}}\,dx\,where\,f'(x)=\frac{dy}{dx}$$

Since '*x*' and '*y*' are 'constant' variables, we can use another variable '*t*' which when used in the functions *x(t) , y(t)* giving the 'x' and 'y' coordinate when 't' is used i.e.

$$y = f(t)\,,\,x = g(t)$$

To find change in 'x' , 'y' as 't' changes, we get:

$$\Delta y = f'(t)\,,\,\Delta x = g'(t)$$

If the curve follows a simple ellipse,

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

then:

$$y=f(x)=\pm\sqrt{b^2(1-\frac{x^2}{a^2})},x=g(x)=x$$

Now if we let *y = f(t) , x = g(t)* then these values can be substitued into the Arc length equation:

$$\text{Arc length} \quad \int\sqrt{(f'(t))^2+(g'(t))^2}\,dt$$

Using chain rule to differentiate f(x):

$$y=\pm\sqrt{b^2(1-\frac{x^2}{a^2})}$$

$$h(x)=\sqrt{x},j(x)=b^2-b^2\frac{x^2}{a^2}$$

$$h'(x)=\frac{1}{2\sqrt{x}},j'(x)=\frac{-2b^2x}{a^2}$$

$$y'=h'(j(x))j'(x),y'=\frac{\frac{\frac{-2b^2x}{a^2}}{2\sqrt{-2b^2x}}}{a^2}\rightarrow-\frac{bx\sqrt{2x}}{2aix}$$

Substitute into Arc length:

$$arc\ length=\int_a^b\sqrt{t'+y'}=\int_a^b\sqrt{1-bx\frac{\sqrt{2x}}{2aix}}$$

During this era, the arc lengths would have possibly been used for many applications in science such as in Planetery Motion, predicting the movements of the planets and plotting their distances relative to the Earth.

They also would have applications in marine engineering at the time in the form of boat/ship keels and hull shaping. Although these physical applications do not need rigorous mathematics and may be solved much more easily using geometrical approximation in a design plans.

Actually calculating the above expression is beyond the scope of this investigation, but its difficulty led to the naming of similar functions to these to be called elliptical integrals.

Elliptical integrals may take many forms but the modern definition is:

$$f(x)=\int_{c}^{x} R(t,\sqrt{P(t)})$$

Here R() is a rational function and in context with the above equations for arc length, *t* is *g(x)* and P(t) is f(x) but the highest polynomial orders now have to be larger than 3.

These elliptical integrals in turn led to elliptical curves in that they are essentially the same form of expression where there is no integration taking place such as:

$$y^2=ax^3+bx^2+cx+c$$

This is equivalent to what may be an elliptic integral, in addition it is also very similar to elliptic curves in that when plotted produces this example:

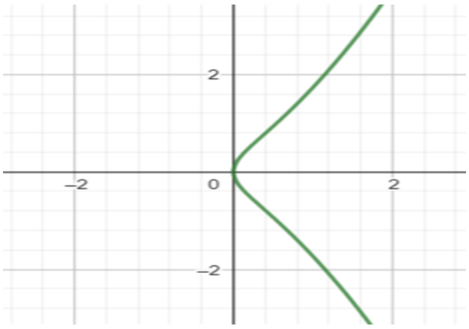


Figure 1: $y^2 = ax^3+bx^2+cx+d$

This is similar to a type of elliptic function of particular interest.
We simply change the values of *a,b,c,d* to produce different curves, such that the curve is non – singular; it has no intersections, cusps (where the curve 'violently' changes direction) or has any isolated points.

This means that the discriminant of the curves satisfies:

$$4c^3+27d^2\neq 0$$

This discriminant is derived from the discriminant of a cubic:

$$\Delta=18abcd-4b^3d+b^2c^2-4ac^3-27a^2d^2.$$

Where *b = 0, a = 1*
This resultant curve where the discriminant satisfies the above criteria follows the equation

$$y^2=x^3+cx+d$$

This can be rewritten as

$$y^2=x^3+ax+b$$

This function has its own *a* and *b* as it is a special non-singular function. This is known as an Elliptic Curve which has some special properties and looks like:

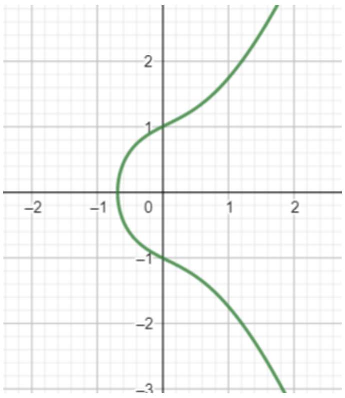


Figure 2: $y^2 = x^2 + x + 1$

or

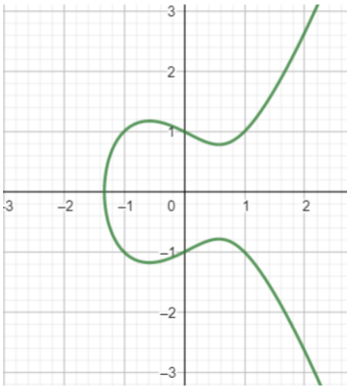
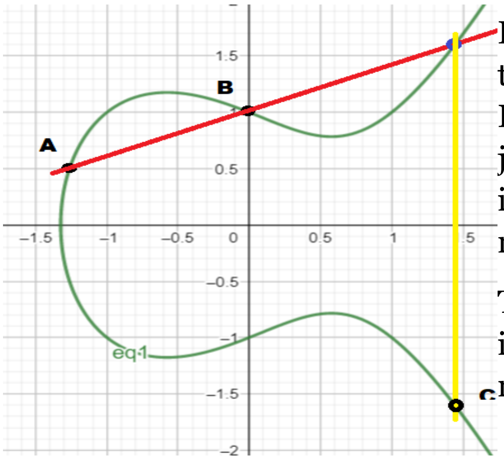


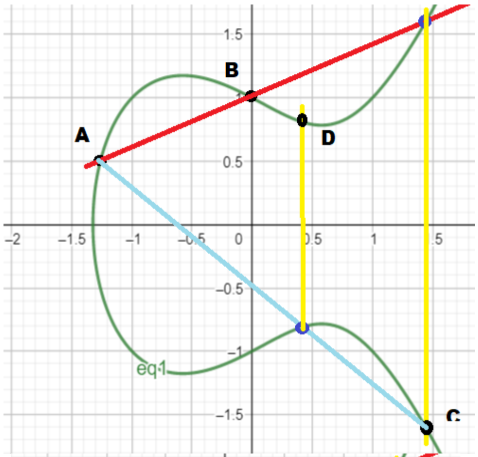
Figure 3: $y^2 = x^3 - x + 1$

2. Properties of Elliptic Curves



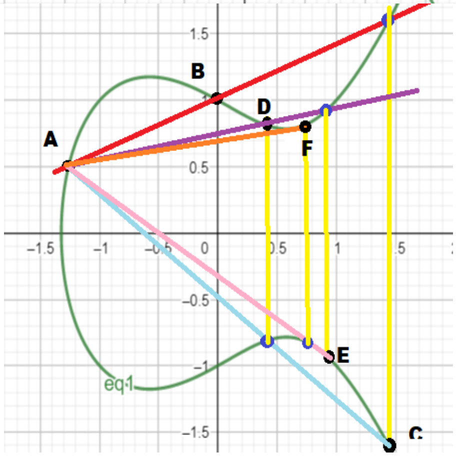
Ellliptical Curves are symmetrical around the *y* axis due to the y^2 having two roots, positive and negative. Furthermore, if you were to put a point on the curve an join it with another, this will create a line which may intersect with the curve. This resultant point can then be reflected in the negative axis.

This point is represented as the logic A + B = C, where A is the first point, B is the second point and C is the reflected resultant point of intersection.



If we continue connecting this point C back to A, the plotted line will again intersect with the elliptic curve. This will give another resultant point, again this is reflected in the x- axis and its reflection will give a new point, we may call it D.

The black points are intersections, the blue points are resultant ‘to be reflected’ points.



As you can see, the line from C to A has intersected the curve and the resultant reflection has been marked D.

This can be iterated many times, moving the final point along the curve i.e. from A to F.

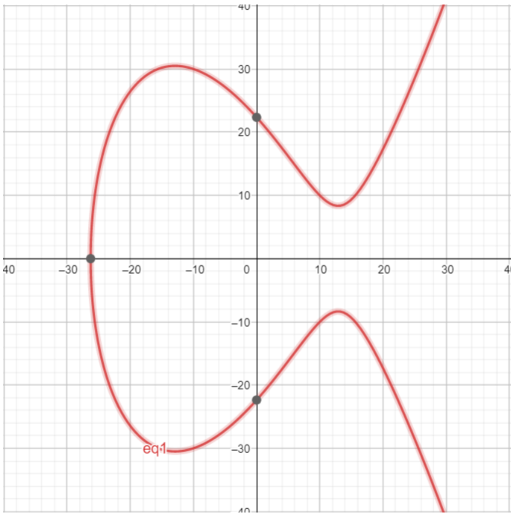
This means that from these two input points, and ‘running’ this function a multiple of times you can return a ‘random’ number somewhere on the curve. Although the ‘severity’ of the displacement of the final point does depend on the equation you begin with. If the equation is rather simple like the one used above, the points are rather ‘close’ together. A different equation such as $y^2=x^3-5x+5$ will return a different curve. Some even return a ‘circle’ and a curve:

Since the elliptic curve forms over such a small area, we can enlarge it.

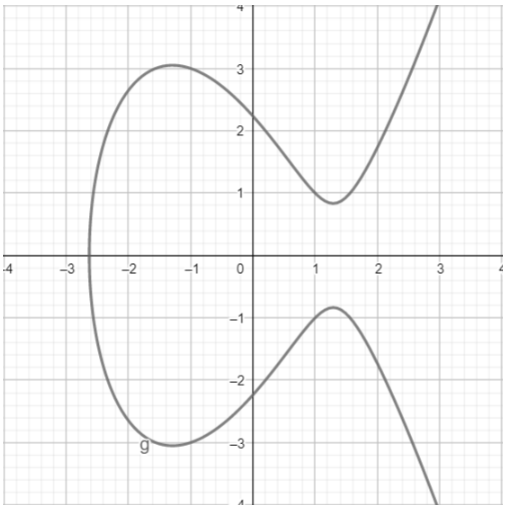
If we want a 'larger' *x* axis, we can simply use $f(\frac{x}{10})$, this makes the effective x axis ten times larger. To keep to the aspect ratio we can enlarge the *y* axis by ten, however since *y* is squared, we must enlarge it by 10^2 = 100. Therefore enlargement by a factor of ten is represented by:

$$y^2 = 10^2 f(\frac{x}{10})$$

This results in the elliptic curve above being shown as:



Rather than:



This leads onto some uses of elliptic curves as this enlargement is helpful.

3. Uses of elliptic curves in cryptography:

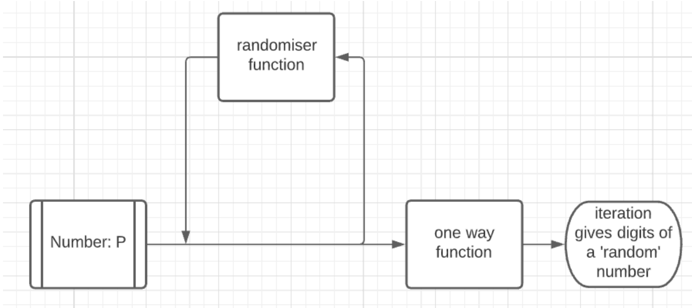
One use of Elliptic curves that I find very impressive is in cryptography. Usually very large products of prime numbers are used to encrypt data. This works as very large prime numbers are calculated and known. With two prime numbers (with 100s to 1000s of digits) they can be multiplied together giving a very large number. For simplicity's sake I will say the each key is a prime number and only two people know these numbers, the people that want to access the information.

The information will be encrypted using the massive prime number, since either of the people have the key, they can decrypt the information. However for intruders trying to decrypt the data, they find it is encrypted using a very large number. They must factorise the number to gain access. Since they do not know the primes to begin with, it may take a very long time to manually brute force the decyption by trying hundreds and thousands of prime numbers.

Although you can technically get around this the computing time will be so large that it would take many years to crack the prime number effectively making this a 'one way' function.

You can easily make the large encryption number, but it is very difficult to undo if you do not know the starting numbers.

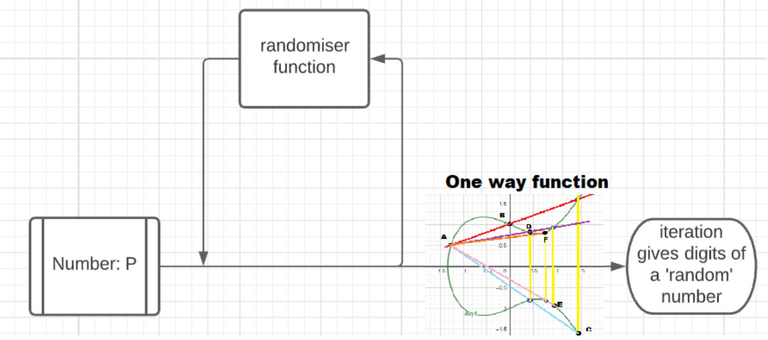
Other forms of one way functions are random number generator which repeatedly iterate a function producing ‘random’ numbers. This is iterated multiple times through various funtions to give a random number:



A number P is entered, it becomes encrypted forming part of a key, then P is passed through the randomiser multiple times each time producing a new digit to form the encryption key, this may be a video of something random such as a dice or lava lamps. Eventually as this has cycled enough times as necessary ie. a 200bit key may cycle 200 times (although binary only has 2 states so this is useless in this sense), a full key have been produced.

Now imagine this same case but instead of using a randomiser function, we use elliptic curves.

The logic, A + B = C is iterated multiple times, possible even multiple times for one iteration of the encryption process. The seeminly random movement of the points on the curve now have a use: They are to create a seemingly random string of numbers to be used in encryption. The key is that after you have iterated the elliptic function multiple times, the point may have moved anywhere on the curve even onto a previous point. How is a hacker meant to undo an elliptic function if it is so ‘random’ in nature. In addition to this, a use as a one-way function is highly effective, how are you meant to undo an elliptic function?

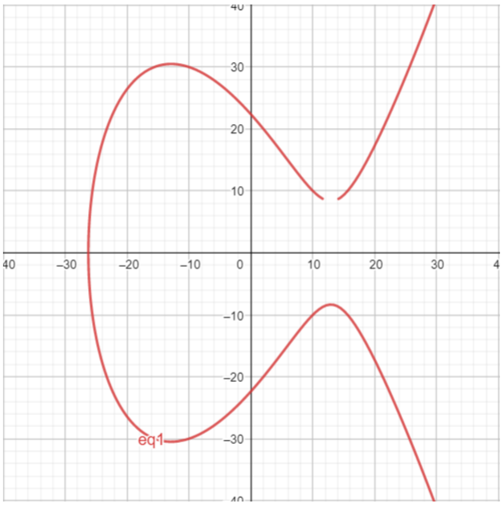


A very common analogy with most concepts are billiard balls.

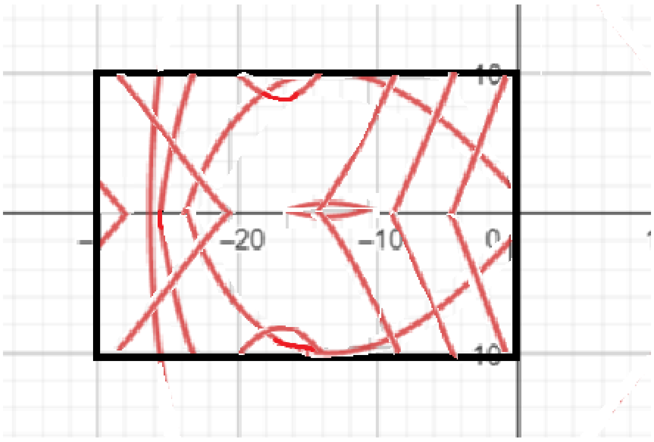
Think of the elliptic curve as a corner of the table, if you hit the ball onto the side it may bounce multiple times and end up in a ‘random’ place on the table. If you did not know what movements took place and saw the ball, how are you meant to figure out where its original position was?

This is the same with an elliptic function when used for cryptography. During actual use, the enlargement from earlier is important. Computers cannot store infinitely large numbers, they have limits, therefore the curve can only go so high, and points can only be determined to so many decimal places.

Therefore the curve must be within its limits and much like the classic game asteroids simply goes to the opposite end and starts again.

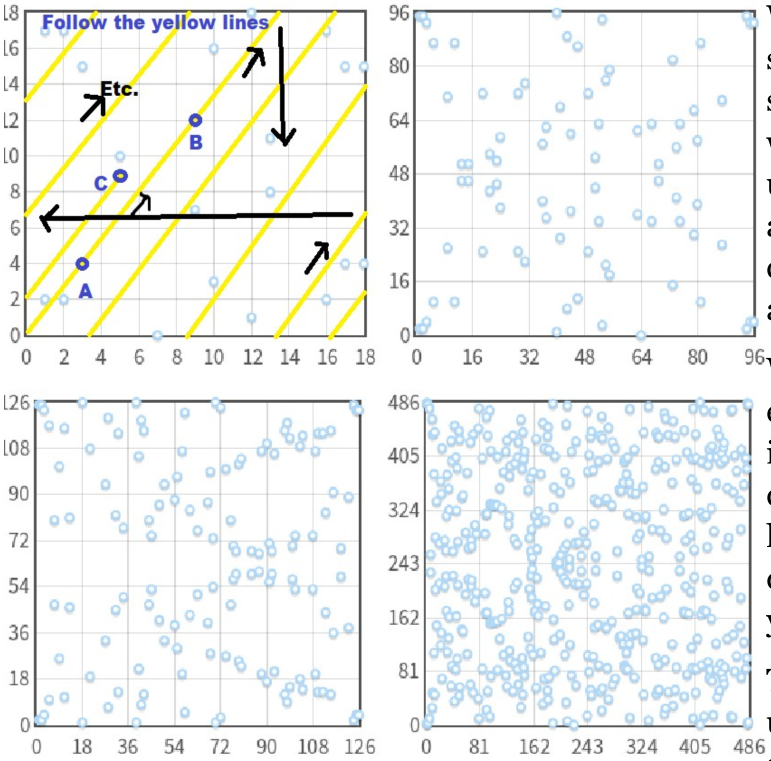


Becomes:



This is the curve on the left ‘folded’ into a small limiting box, if it goes too far to the right is starts again from the left, if it goes too high it starts again from the middle and vice versa. For a ‘whole’ elliptic curve, there would be many more lines as all the lines that entend beyond the view of the left image will still be packed into the right image.

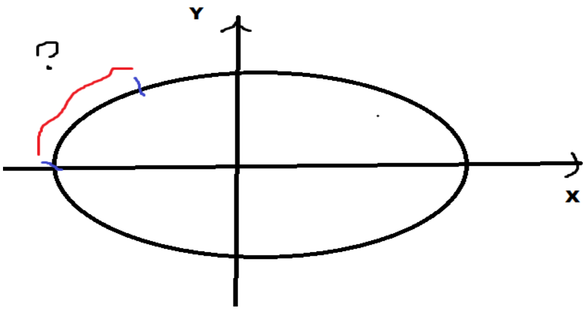
Since computers would rather use integer numbers, a factor of ten enlargement would mean that more of the curve’s points are on integers. If you were to plot these integer point you would get a similar curve to this one picture I found online, as you can see they get very complex.

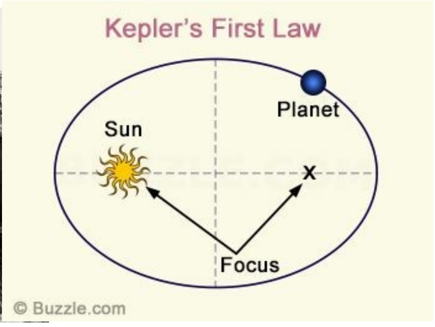


When joining points, the line drawn simply starts again at opposite the side that it went out of. The line would then continue being drawn until it intersects with another point, akin to drawing intesections on the curve. You can follow the yellow lines and black arrows from A,B to find C

With this visual representation, it is easy to understand how this is useful in cryptography. Pick a point, how can you find the original two points, how did you get there? It is extremely difficult, especially the more points you have, such as in the bottom right.

Therefore, elliptic curves may prove useful in cryptography as they make an excellent one-way function.





In conclusion, elliptic curves begin with much in relation to ellipses, they stem from calculus related to their arc lengths again sourced from uses in astronomy and engineering. This in turn led to developments in these integrals and subsequently to the elliptical curves as a function and group. Currently elliptical curves provide an niche but important part of the modern world and other amazing uses such as in the proof of Fermat's Last Theorem.

Sources/Materials:

Myself,

Wolfram Alpha,

Brown University Lecture Notes,

Computerphile,

Wikipedia.

Programs for graphs/images:

MS Paint,

LucidChart

Geogebra