The Cubic Formula: A Tale of Skulduggery and Intrigue

All good secondary students would remember solving quadratic equations at some point in their education. Some may even recognise the general formula for solving the quadratic equation in the form:

$$ax^2 + bx + c = 0$$

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

However, very few would have come across the corresponding formula to solve a general cubic equation:

$$ax^3 + bx^2 + cx + d = 0$$

$$x = \sqrt[3]{-\frac{b^3}{27a^3} + \frac{bc}{6a^2} - \frac{d}{2a} + \sqrt{\left(-\frac{b^3}{27a^3} + \frac{bc}{6a^2} - \frac{d}{2a}\right)^2 + \left(\frac{c}{3a} - \frac{b^2}{9a^2}\right)^3}} + \sqrt[3]{-\frac{b^3}{27a^3} + \frac{bc}{6a^2} - \frac{d}{2a} - \sqrt{\left(-\frac{b^3}{27a^3} + \frac{bc}{6a^2} - \frac{d}{2a}\right)^2 + \left(\frac{c}{3a} - \frac{b^2}{9a^2}\right)^3}} - \frac{b}{3a}$$

It is mouthful so it is easy to see why!

In this article, I will be looking at how these two formulae are derived based on the symmetries of these two curves, and then show how the quirks of the cubic formula led to important discoveries in the development of Mathematics.

1. The Quadratic Formula: An Alternative Derivation

Before we look at the derivation of the cubic formula, we should look at how the famous quadratic formula is derived and see if we can use the method to help us find a formula for cubic equations. Usually, the derivation of the quadratic formula involves completing the square; a method which involves writing the quadratic as a square of x plus a constant term that equals another constant term. However, this method does not help us much when applying it to a cubic. Instead, here is another approach to derive the quadratic formula that also helps us when applied to cubic equations.

First, consider the general quadratic equation $ax^2 + bx + c = 0$, where a, b and c are all constants. Dividing through by a gives the following equation:

$$x^2 + \frac{b}{a}x + \frac{c}{a} = 0$$

When we plot the equation $y = x^2 + \frac{b}{a}x + \frac{c}{a}$, we get a graph that looks like the red curve below:

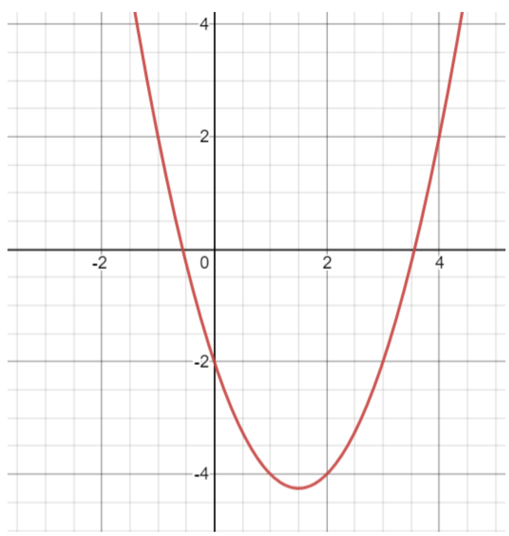

Figure 1. A quadratic equation

As we can see, the graph has a vertical line of reflective symmetry about the minimum point (the point at which the y coordinate of the graph is the lowest), with the roots lying an equal distance from this point. Therefore, if we "move" the curve so that the minimum point is on the y-axis, we will have an equation of the form $x^2 - k = 0$, where k is a constant, which will be much easier to solve. This is because the new curve will have roots that are an equal distance away from the origin and on opposite sides. The new curve looks like the blue curve in Figure 2 below:

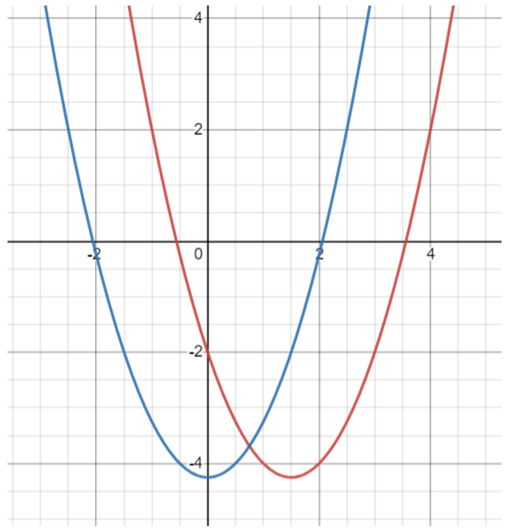

Figure 2. "Moving" the quadratic

To find this point, we differentiate the equation $y = x^2 + \frac{b}{a}x + \frac{c}{a}$ and set it equal to zero (as the gradient of the curve is zero at the minimum point), which results in the following:

$$\frac{dy}{dx} = 2x + \frac{b}{a} = 0$$

Solving for x gets us:

$$x = -\frac{b}{2a}$$

Therefore, when we reflect the curve in the line $x = -\frac{b}{2a}$, it remains the same.

This means that, in order to "move" the curve such that the minimum point is on the y-axis, we need to make the substitution:

$$x = z - \frac{b}{2a}$$

This changes the equation to:

$$\left(z - \frac{b}{2a}\right)^2 + \frac{b}{a}\left(z - \frac{b}{2a}\right) + \frac{c}{a} = 0$$

Expanding the brackets gives the following equation:

$$\left(z^2 - \frac{bz}{a} + \frac{b^2}{4a^2}\right) + \left(\frac{bz}{a} - \frac{b^2}{2a^2}\right) + \frac{c}{a} = 0$$

Collecting like terms simplifies the equation to:

$$z^2 - \frac{b^2}{4a^2} + \frac{c}{a} = 0$$

Bringing the constant terms all in one fraction results in the equation:

$$z^2 - \frac{b^2 - 4ac}{4a^2} = 0$$

Now this is starting to look like the quadratic formula that we are familiar with. Solving for z gives the following equation:

$$z = \pm\frac{\sqrt{b^2 - 4ac}}{2a}$$

We are almost there! Now, to find the roots of the original equation we must “move” the solutions back to their original position. This means undoing the “movement” we performed on the curve originally (we need to “move” the roots from their positions on the blue curve back onto the red curve). To do this, we substitute x in for z using our previous definition of z, resulting in the equation:

$$x + \frac{b}{2a} = \pm\frac{\sqrt{b^2 - 4ac}}{2a}$$

Finally, solving for x by taking away $\frac{b}{2a}$ from both sides gives us the quadratic formula:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

This alternative derivation may seem a little more complicated than the usual method to prove the quadratic formula, but using a substitution like this in the cubic formula helps simplify the general cubic into something we can solve for more easily.

2. The Cubic Formula: History and Derivation

Although the quadratic formula has been known for millennia by several ancient civilisations such as the Ancient Greeks and Egyptians, the discovery of the cubic formula (also known as Cardano's formula) is a relatively recent one. It was originally discovered independently by Italian mathematicians Scipione del Ferro, and later Niccolo Tartaglia, in the 16th Century. Both kept the formula a secret so that they could win mathematical duels against other mathematicians; as only they knew how to solve cubic equations, they could gain an advantage over their opponents. Later, del Ferro passed away in 1526, without publishing his formula.

However, another mathematician Gerolamo Cardano convinced Tartaglia to reveal the cubic formula to him in 1539, under the condition that Cardano would not tell anyone else about it and not publish it. When Cardano found the cubic formula in the notes of the deceased del Ferro, he decided to publish the cubic formula in his book *Ars Magna* in 1545, which infuriated Tartaglia. He justified this by stating that he was publishing del Ferro's work and not Tartaglia's. This led to a long feud between the two. Eventually, Tartaglia challenged Cardano's student Lodovico Ferrari, who had derived the quartic formula based on the work of Tartaglia. Tartaglia lost the duel, costing him his reputation.

The derivation of the cubic formula is as follows, using the method we saw to derive the quadratic formula:

First of all, consider the general cubic equation $ax^3 + bx^2 + cx + d = 0$. Dividing the whole equation through by a gives the equation:

$$x^3 + \frac{b}{a}x^2 + \frac{c}{a}x + \frac{d}{a} = 0$$

Cubic equations, like quadratics, have a symmetry as well. Instead of reflective symmetry, they have rotational symmetry of order 2 about their inflection point. This is the point where the curve goes from being convex (so looking like a negative quadratic) to being concave (so looking like a positive quadratic) or vice versa. The general cubic will look like the graph below:

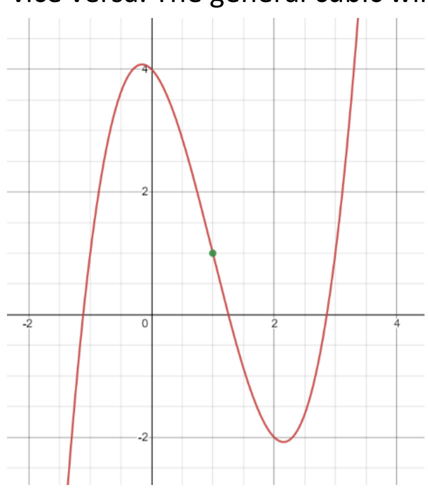

Figure 3. A cubic equation

Here, the green dot represents the point of infection. By "moving" the entire curve so that this point is on the y-axis, we can simplify the problem, as the second derivative of the new graph will be zero at the origin, so the constant term in the second derivative is zero, so there is no x^2 term in the equation of the new curve. To find the point, we need to set the second derivative of the curve equal to zero. This gives the following equation:

$$\frac{d^2y}{dx^2} = 6x + \frac{2b}{a} = 0$$

Solving for x gives that:

$$x = -\frac{b}{3a}$$

This looks quite similar to the x-coordinate of the minimum point of the quadratic equation. So, if we rotate the entire curve 180° about the point on it where $x = -\frac{b}{3a}$, the whole curve will be unchanged. Once again, we make the substitution $x = y - \frac{b}{3a}$ to "move" the point of inflection to the y-axis, resulting in the following equation that can be visualised by the blue curve below:

$$\left(y - \frac{b}{3a}\right)^3 + \frac{b}{a}\left(y - \frac{b}{3a}\right)^2 + \frac{c}{a}\left(y - \frac{b}{3a}\right) + \frac{d}{a} = 0$$

Figure 4. "Moving" the cubic

Expanding the equation out using the binomial theorem gives the following equation:

$$\left(y^3 - \frac{b}{a}y^2 + \frac{b^2}{3a^2}y - \frac{b^3}{27a^3}\right) + \left(\frac{b}{a}y^2 - \frac{2b^2}{3a^2}y + \frac{b^3}{9a^3}\right) + \left(\frac{c}{a}y - \frac{bc}{3a^2}\right) + \frac{d}{a} = 0$$

Collecting like terms and factorising gives the following equation:

$$y^3 + \left(\frac{c}{a} - \frac{b^2}{3a^2}\right)y + \left(\frac{2b^3}{27a^3} - \frac{bc}{3a^2} + \frac{d}{a}\right) = 0$$

To simplify things, we can let the coefficient of y in the equation be p and the constant be q, simplifying the cubic to the equation $y^3 + py + q = 0$. This is a depressed cubic, which means that it does not have a term in y^2. To transform this cubic into a form which we can solve, we make another substitution $y = z - \frac{p}{3z}$. Expanding the brackets gives:

$$\left(z^3 - pz + \frac{p^2}{3z} - \frac{p^3}{27z^3}\right) + \left(pz - \frac{p^2}{3z}\right) + q = 0$$

Collecting like terms gives us the equation:

$$z^3 + q - \frac{p^3}{27z^3} = 0$$

Multiplying through by z^3 gives the disguised quadratic:

$$z^6 + qz^3 - \frac{p^3}{27} = 0$$

In this equation, the value of z^3 can be found using the quadratic formula to be:

$$z^3 = \frac{-q \pm \sqrt{q^2 + \frac{4p^3}{27}}}{2}$$

By multiplying both the top and the bottom of the fraction by a half and taking the cube root of the whole thing we get:

$$z = \sqrt[3]{-\frac{q}{2} \pm \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}$$

Substituting this back into the equation for y gives the following equation:

$$y = \sqrt[3]{-\frac{q}{2} \pm \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \frac{p}{3\sqrt[3]{-\frac{q}{2} \pm \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}}$$

By multiplying both the top and the bottom of the fraction by $\sqrt[3]{-\frac{q}{2} \mp \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}$ and simplifying the expression we get to the following equation:

$$y = \sqrt[3]{-\frac{q}{2} \pm \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} \mp \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}$$

When we take half of the expression as $\sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}$ and the other as $\sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}$, we arrive at Cardano's formula for the roots of a depressed cubic:

$$y = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}$$

If we wanted to, we could substitute the values of y, p and q back into this equation, but the equation above is usually how the cubic formula is expressed.

To illustrate how the cubic formula is used, take the example $x^3 + 9x - 26 = 0$. Here, $p = 9$ and $q = -26$. Substituting these values into the formula gives:

$$x = \sqrt[3]{-\frac{(-26)}{2} + \sqrt{\frac{(-26)^2}{4} + \frac{9^3}{27}}} + \sqrt[3]{-\frac{(-26)}{2} - \sqrt{\frac{(-26)^2}{4} + \frac{9^3}{27}}}$$

This simplifies to:

$$x = \sqrt[3]{13 + \sqrt{169 + 27}} + \sqrt[3]{13 - \sqrt{169 + 27}} = \sqrt[3]{13 + 14} + \sqrt[3]{13 - 14} = 3 - 1 = 2$$

As $x = 2$ does satisfy the cubic equation, we can see that the formula works.

It was a great achievement for Tartaglia and del Ferro to have derived this formula in the 16th Century, as it allowed cubic equations that did not have rational roots to be solved for the first time. However, the cubic formula starts to pose some problems when it is applied in some situations.

3. Using Cardano's Formula and the *Casus Irreducibilis*

Although the cubic formula does give us the roots of cubic equations, we run into problems when we try to use it for certain cubic equations. This is because, when $\frac{q^2}{4}+\frac{p^3}{27}$ is less than zero, the formula involves the square root of negative numbers. Under the assumption that the square root of a negative number cannot exist, the formula would not give a valid result.

All cubic equations have at least one real root. This is because, when y is very large and negative, the cubic $f(x)=x^3+\frac{b}{a}x^2+\frac{c}{a}x+\frac{d}{a}$ is negative as the x^3 term "dominates" the other terms in the expression. Similarly, when x is very large and positive, $f(x)$ is positive. As the cubic has a value for all real numbers, it is continuous as well. Therefore, the change in sign of $f(x)$ from negative to positive implies that the cubic equation must have at least one real root, as it must cross the line $y=0$. This fact means that Cardano's formula should give a valid root for every cubic equation, which contradicts what we observe when we use the formula.

For example, take the equation:

$$x^3+3x^2-3x-1=0$$

It has a root at $x=1$ (as $1^3+3(1)^2-3(1)-1=0$), so we can factorise the equation to:

$$(x-1)(x^2+4x+1)=0$$

By using the quadratic formula to solve the quadratic equation $x^2+4x+1=0$, we can see that the three roots of the cubic are $x=4$ or $x=-2+\sqrt{3}$ or $x=-2-\sqrt{3}$, which are all real numbers.

To test this cubic in Cardano's formula, we need to modify the cubic into a depressed cubic, so we make the substitution $z=x-1$ (as $-\frac{b}{3a}=-1$). Making the substitution and expanding the brackets gives:

$$(z^3-3z^2+3z-1)+3(z^2-2z+1)-3(z-1)-1=0$$

Collecting like terms results in the depressed cubic $z^3-6z+4=0$. When we substitute in $p=-6$ and $q=4$ into Cardano's formula we get:

$$z=\sqrt[3]{-2+\sqrt{-4}}+\sqrt[3]{-2-\sqrt{-4}}$$

This would indicate that the equation had no real roots as the formula contains the square root of -4 (under the assumption that the square roots of negative numbers cannot exist), much like if the discriminant of a quadratic is negative. Despite this, we have already found that the cubic has 3 real roots ($x=1, x=-2+\sqrt{3}$ and $x=-2-\sqrt{3}$).

So how do we make sense of this result from Cardano's formula?

This problem of having square roots of negative numbers in the results of Cardano's formula was known as the *casus irreducibilis*, which remained unsolved until another Italian mathematician, Raphael Bombelli, solved the issue. He did this by allowing the square root of minus one be its own number, which became known as the imaginary number (usually denoted by i). Numbers involving i became known as complex numbers and they were represented on an Argand diagram: a graph which has the horizontal axis containing all the real numbers; and a vertical axis, which contains multiples of i. On Figure 5 below, the two points $-2+2i$ and $-2-2i$ are represented by the black dots A and A' respectively.

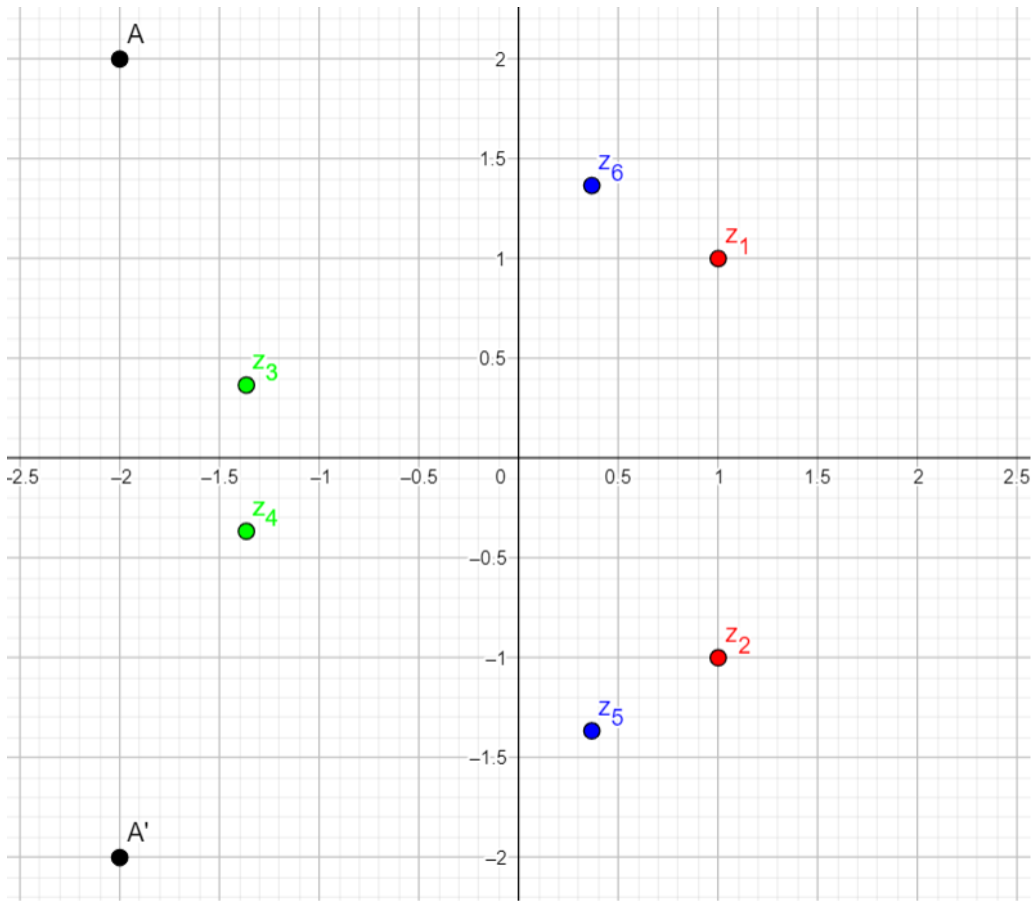

Figure 5. The Argand diagram with the cube roots of $-2+2\mathrm{i}$ and $-2-2\mathrm{i}$

First, we can write $-2+2i$ and $-2-2i$ in modulus-argument form; a form which expresses the complex number as its distance from the origin (the modulus) and the angle it makes with the positive real axis (the argument) between π and $-\pi$ radians (where π radians equal 180°). The modulus of both these numbers is $\sqrt{(-2)^2+(\pm 2)^2}=\sqrt{8}$ using Pythagoras' Theorem. The angle (in radians) they make with the positive real axis are:

$$\pi - tan^{-1}\left(\frac{2}{2}\right)=\frac{3\pi}{4} \quad and \quad -\pi + tan^{-1}\left(\frac{2}{2}\right)=-\frac{3\pi}{4}$$

for $-2+2i$ and $-2-2i$ respectively (if the complex number is below the real axis then we write its angle as negative rather than as a multiple of π larger than 1). Therefore, we can express the two complex numbers as:

$$\sqrt{8}(\cos\frac{3\pi}{4}+i\sin\frac{3\pi}{4}) \quad and \quad \sqrt{8}(\cos(-\frac{3\pi}{4})+i\sin(-\frac{3\pi}{4}))$$

respectively, with the cosine representing the real coordinate of the complex number and the sine representing the imaginary coordinate of the number. To cube root complex numbers, we take the cube root of their modulus and divide their arguments by 3. This gives us:

$$\sqrt[3]{-2+2i}=\sqrt{8}(\cos\frac{\pi}{4}+i\sin\frac{\pi}{4}) \quad and \quad \sqrt[3]{-2-2i}=\sqrt{8}(\cos\left(-\frac{\pi}{4}\right)+i\sin(-\frac{\pi}{4}))$$

These are shown by the red dots z_1 and z_2 respectively on the Argand diagram above. To finally get a value for z, we add the two cube roots together, resulting in:

$$z=\sqrt{2}(\cos\left(-\frac{\pi}{4}\right)+\cos\left(\frac{\pi}{4}\right))=2$$

As these two complex numbers are conjugates (numbers which have the same modulus but the opposite sign of the argument), their imaginary parts are of equal magnitude but have the opposite sign, so adding them together removes the imaginary part of the resulting number, leaving us with a real answer. Substituting this back into our formula for x gives $x=2-1=1$, which was one of the roots of the original equation!

But what happened to the other two roots? Well, there are actually three cube roots for each complex number, as there is more than one argument that we can multiply by 3 to get $\frac{3\pi}{4}$. Take $\frac{11\pi}{12}$ for example; multiplying it by 3 gives $\frac{11\pi}{4}$, which is the same argument as $\frac{3\pi}{4}$, as going around the Argand diagram a full turn extra does not change the position of the resulting complex number. This gives us another cube root of the complex numbers:

$$\sqrt[3]{-2+2i}=\sqrt{2}(\cos(\frac{11\pi}{12})+i\sin(\frac{11\pi}{12}))$$
$$\sqrt[3]{-2-2i}=\sqrt{2}(\cos(-\frac{11\pi}{12})+i\sin(-\frac{11\pi}{12}))$$

These are represented by the green dots z_3 and z_4 respectively. Together with the other possible argument $\frac{19\pi}{12}$ (which is usually written as $-\frac{5\pi}{12}$), we get the third cube root of the complex numbers:

$$\sqrt[3]{-2+2i}=\sqrt{2}(\cos(-\frac{5\pi}{12})+i\sin(-\frac{5\pi}{12}))$$
$$\sqrt[3]{-2-2i}=\sqrt{2}(\cos(\frac{5\pi}{12})+i\sin(\frac{5\pi}{12}))$$

These are represented by the blue dots z_5 and z_6 respectively. These arguments are usually worked out by adding a multiple of $\frac{2\pi}{3}$ to the argument of one cube root, as multiplying by 3 converts this into a multiple of 2π, which can be taken away.

Summing these two other cube roots gives:

$$z = \sqrt{2}(\cos\left(-\frac{11\pi}{12}\right) + \cos\left(\frac{11\pi}{12}\right)) = -1 - \sqrt{3}$$
$$z = \sqrt{2}(\cos\left(-\frac{5\pi}{12}\right) + \cos\left(\frac{5\pi}{12}\right)) = -1 + \sqrt{3}$$

Substituting these values in for x results in the solutions $x = -2 + \sqrt{3}$ and $x = -2 - \sqrt{3}$, which are the other two solutions of the cubic that we found earlier. The use of complex numbers not only has given us a valid solution from Cardano's formula, but has given us all three real solutions of the cubic as well.

4. Conclusion

The use of the number $i = \sqrt{-1}$ in finding the real roots of cubic equations gave rise to the idea of complex numbers as well as the validity of their existence in Mathematics, thus introducing a powerful tool to modern mathematicians. This is why it is one of my favourite parts of Mathematics, as its connection with complex numbers was rather unexpected to me. Like the cubic formula, there exists a quartic formula as well, which is derived using a similar method as the derivation of the cubic. Instead of using the term $-\frac{b}{3a}$ to transform the polynomial, we use the term $-\frac{b}{4a}$ to convert a quartic to its depressed form, as we set the third derivative of the quartic equal to zero. What symmetry of the curve does this represent though? I will leave this as an exercise for the reader to figure out. Furthermore, the existence of a general formula for quintic equations and higher degree polynomials using radicals was disproven using something called Galois Group Theory which you can have a look at here.

References

- https://math.vanderbilt.edu/schectex/courses/cubic/
- https://www.youtube.com/watch?v=N-KXStupwsc
- https://mathworld.wolfram.com/CubicFormula.html
- https://www.storyofmathematics.com/16th_tartaglia.html
- https://mathworld.wolfram.com/QuinticEquation.html

Written By
Anirudh Khaitan