On the Algebraic Solution of the Cubic, version 2

Table of Contents

Copyright Notice

Preamble

Since I have the impression that my paper "On the Algebraic Solution of Polynomial Equations" wasn't particularly well-received (due to the terse and highly mathematical nature of the paper), I decided to rewrite the algebraic solution of the cubic in a manner which I hope is a bit more intuitively accessible.

Introduction: The Solution of the Quadratic

You should be familiar with the general quadratic equation

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

and the quadratic formula, which solves these equations:

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

Have you ever wondered how the quadratic formula was derived? Well, as it turns out, it's not too difficult, and in fact will prove to be very useful when studying how to solve the cubic. So let's have a look at how the quadratic formula was derived.

To begin with, suppose we try to solve

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

directly. After fiddling around with it for a while, eventually, you should realize that we have a big problem: there are two x's in the equation, making it impossible to solve for x, since there will always be another x on the other side after we're done.

How do we get around this? Well, ideally, we'd want some way of modifying the equation such that only a single x remains. I mean, if we think about it, what we really want is an equation that looks like

$$ax^2+k=0$$

which is the general quadratic with no linear (x) term and really easy to solve:

$$x=\pm\sqrt{-\frac{k}{a}}$$

The obvious problem is that only a small fraction of quadratic equations look like this – we need something that works for *all* quadratic equations.

Now, here's the key insight, the magic trick, if you will: **if we can't get an equation that easy, maybe we can get the next best thing** – something that looks like this:

$$a(x+h)^2+k=0$$

This equation looks a little more complicated than our last one, but it's still easily solvable, isn't it? Using basic algebra (I'll skip the steps), the answer is

$$x=-h\pm\sqrt{-\frac{k}{a}}$$

Great, you might say, but how does this help us? Well, if we expand out our "still easy to solve" equation, we get

$$ax^2+2ahx+ah^2+k=0$$

which, I admit, looks pretty ugly. But let's look at it anyway...doesn't it vaguely remind you of the general quadratic equation? It should. And, in fact, if we make the definitions

$$b=2ah$$
$$c=ah^2+k$$

then we get

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

which is exactly what we want to solve!

Now we can finish this off and derive the quadratic formula in all its glory. To begin with, we already know that

$$x=-h\pm\sqrt{-\frac{k}{a}}$$

which means that all we need is h and k in terms of b and c. From our set of definitions, we know that

$$b=2ah$$

which means

$$h=\frac{b}{2a}$$

and since

$$c=ah^2+k$$

we can substitute in our value for *h* and then solve for *k*:

$$\begin{aligned}
c&=a\left(\frac{b}{2a}\right)^2+k\\
c&=\frac{ab^2}{4a^2}+k\\
c&=\frac{b^2}{4a}+k\\
k&=c-\frac{b^2}{4a}
\end{aligned}$$

Now we substitute these to our formula for *x*:

$$\begin{aligned}
x&=-\frac{b}{2a}\pm\sqrt{\frac{-\left(c-\frac{b^2}{4a}\right)}{a}}\\
x&=-\frac{b}{2a}\pm\sqrt{\frac{\frac{-4ac}{4a}+\frac{b^2}{4a}}{a}}\\
x&=-\frac{b}{2a}\pm\sqrt{\frac{b^2-4ac}{4a^2}}\\
x&=-\frac{b}{2a}\pm\frac{\sqrt{b^2-4ac}}{2a}\\
x&=\frac{-b\pm\sqrt{b^2-4ac}}{2a}
\end{aligned}$$

and that's all there is to it! We have the quadratic formula!

The Solution of the Cubic, Part 1

Now that we know how to solve the general quadratic equation, let's turn our eyes to the general cubic equation:

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

As with the quadratic, what we really want is some equation that looks like

$$ax^3+k=0$$

for which we know the solution to:

$$x=\sqrt[3]{-\frac{k}{a}}$$

Obviously, only a handful of cubics really look like this. But, just as with the quadratic, we'll happily settle for the next best thing:

$$a(x+h)^3+k=0$$

The solution is still easy to get:

$$x=-h+\sqrt[3]{-\frac{k}{a}}$$

And now all we need to do is expand this out and create a set of simple equations to find *h* and *k* in terms of *a, b, c*, and *d*. So let's do so:

$$ax^3+3ahx^2+3ah^2x+ah^3+k=0$$

Matching up coefficients gives us

$$\begin{aligned} b&=3ah \\ c&=3ah^2 \\ d&=ah^3+k \end{aligned}$$

The first equation gives us

$$h=\frac{b}{3a}$$

and the second equation

$$h=\pm\sqrt{\frac{c}{3a}}$$

and now we...wait a minute. Don't we have two values of *h*? That could be a problem...unless the two values are equal, in which case life is good. But for them to be equal, we need the following to be true:

$$\frac{b}{3a}=\pm\sqrt{\frac{c}{3a}}$$

Simplifying this a bit:

$$\frac{b}{3a} = \pm\sqrt{\frac{c}{3a}}$$
$$\frac{b^2}{9a^2} = \frac{c}{3a}$$
$$\frac{b^2}{3a} = c$$
$$b^2 = 3ac$$

So the ultimate condition required for our trick to work is

$$b^2 = 3ac$$

which is a bit of a problem. And by a bit of a problem, I mean a really big problem – because the coefficients of the general cubic equation should all be *independent*, meaning that none of them should depend on the values of any others.

So what does all this mean? Well, in a nutshell, it means that our neat little trick, which we borrowed from the quadratic, doesn't work here. So, we'd better look for some other way to solve the cubic.

The Solution of the Cubic, Part 2

At this point, you might be looking at me and wondering why I intentionally tried a solution technique I knew wouldn't work in advance. Well, besides the fact that it seems natural to try it (a generalization of the solution to the quadratic), we still get some useful insight from it – in particular, we know that reduction to the form

$$a(x+h)^3 + q = 0$$

is impossible. But remember the fundamental idea that we got out of solving the quadratic:

If we can't get our ideal, we'll settle for the next best thing.

In our case, if we can't get rid of both the quadratic (x^2) and linear (x) terms (which would be our ideal), let's settle for just getting rid of the quadratic term, such that we have something that looks like

$$ax^3 + px + q = 0$$

Again, as with the quadratic, not every cubic looks like this. But also as with the quadratic, we can live with the next best thing:

$$a(x+h)^3 + p(x+h) + q = 0$$

Let's see if this proves to be of any use. Expanding this out, we get

$$ax^3 + 3ahx^2 + (3ah^2 + p)x + ph + q = 0$$

Now matching coefficients to the general cubic:

$$b = 3ah$$
$$c = 3ah^2 + p$$
$$d = ph + q$$

We now get

$$h = \frac{b}{3a}$$

as the only equation defining *h*, which is good – it suggests we're on the right track! Substituting this value of *h* into the second equation, we get

$$c = 3a\left(\frac{b}{3a}\right)^2 + p$$
$$c = \frac{3ab^3}{9a^2} + p$$
$$c = \frac{b^3}{3a} + p$$
$$p = c - \frac{b^3}{3a}$$

Finally, substituting these values of *h* and *p* into the final equation, we will get (I omit the algebra since it's a bit tedious):

$$q = \frac{2b^3}{27a^2} - \frac{bc}{3a} + d$$

So with the definition

$$y = x + h = x + \frac{b}{3a}$$

we can simplify the general cubic to

$$ay^3 + py + q = 0$$

or, after dividing all terms by *a*,

$$y^3 + my + n = 0$$

where

$$m = \frac{p}{a}, n = \frac{q}{a}$$

and the quadratic term is zero! This is called a *depressed cubic equation*, and while we don't have any ideas as to how to solve it yet, you have to admit that it sure *looks* easier than the general cubic (and it is, as we will soon see).

The Solution of the Cubic, Part 3

We're still a long ways off from solving the general cubic equation, but at least now we know that we can assume, without any loss of generality, that our cubic equation looks something like

$$y^3+my+n=0$$

Ideally, we'd want to get rid of the linear term. But, as we saw, the "tricks" (if you're curious, they're formally known as (linear) *Tschirnhaus transformations*) like the one we just tried only work on one term at a time. If we wish to stay rid of the quadratic term, we have to accept that the linear term is here to stay. So we'd better hunt for some way to attack this thing directly.

The key insight, the magic trick, required here is to try a similar, but not quite identical, "trick" (substitution):

$$y=u+v$$

Using this, we have

$$(u+v)^3+m(u+v)+n=0$$

which eventually becomes

$$u^3+3u^2v+3uv^2+v^3+mu+mv+n=0$$

Now, under what conditions does this equation hold? (Remember that we don't know anything about what *u* and *v* should be.) Well, we can try using the coefficients of our depressed cubic (*m* and *n*) to determine their values – in particular, suppose we make the definition (again, a bit more magic here)

$$n=-(u^3+v^3)$$

In this case, the cubed terms and *n* disappear, as they are now negatives of each other:

$$3u^2v+3uv^2+mu+mv=0$$

Now what must *m* be to make everything else disappear? Well, trying to solve for *m*, we get

$$\begin{aligned}3u^2v+3uv^2+m(u+v)&=0\\3uv(u+v)+m(u+v)&=0\\(3uv+m)(u+v)&=0\\m&=-3uv\end{aligned}$$

(we know that because we cannot assume *u+v* is zero, we must make *3uv+m* equal to zero).

So we have a pair of simultaneous equations

$$\begin{aligned}u^3+v^3&=-n\\uv&=-\frac{m}{3}\end{aligned}$$

Let's try to solve this system. Firstly, as we have cubed terms, let's cube the second equation to make the two match up better:

$$u^3+v^3=-n$$
$$u^3 v^3=-\frac{m^3}{27}$$

Now let's solve the first equation for, say, v^3, and then substitute that into the second equation:

$$v^3=-n-u^3$$
$$u^3(-n-u^3)=-\frac{m^3}{27}$$
$$-nu^3-u^6=-\frac{m^3}{27}$$
$$u^6+nu^3-\frac{m^3}{27}=0$$

The sixth power looks pretty scary, but there's really nothing to be scared of; as you may have already noticed, all the powers of *u* are multiples of 3. So if we substitute

$$z=u^3$$

we then get

$$z^2+nz-\frac{m^3}{27}=0$$

which, amazingly, is a simple quadratic equation! And since we already know how to solve the quadratic, all we need to do is to solve this equation for *z*, and then make all the necessary backward substitutions!

The Solution of the Cubic, Part 4

After doing all the monkey work of tedious, backwards substitutions, the final result, the *cubic formula*, looks something like

$$x=-\frac{b}{3a}$$
$$+\sqrt[3]{\frac{-2b^3+9abc-27a^2d+\sqrt{(2b^3-9abc+27a^2d)^2-4(b^2-3ac)^3}}{54a^3}}$$
$$+\sqrt[3]{\frac{-2b^3+9abc-27a^2d-\sqrt{(2b^3-9abc+27a^2d)^2-4(b^2-3ac)^3}}{54a^3}}$$

which probably looks pretty scary to you (it doesn't to me, but that's because you haven't seen what “scary” really means, heh). But, more than its scariness, there are some fundamental problems with this formula, which explain why the formula is never taught to students:

1. A cubic equation has three solutions (not necessarily all real and/or distinct, though) – this formula only gives one solution. There do exist very similar formulas for the other two solutions, but they require the use of complex numbers, which most students are unfamiliar, or at least uncomfortable, with.

2. The expression underneath the square roots in the cubic formula is not guaranteed to be positive, meaning that we might not have real numbers in the formula. The really tricky and irritating thing about it, though, is that the expression underneath the square root *is only negative when the original cubic equation has three real solutions* (if the original cubic has only one real solution, the expression is positive, meaning nothing else needs to be explained). This phenomenon is known as "casus irreducibilis" (the irreducible case), and essentially means that we have a real solution (or rather, three of them) which can only be expressed algebraically by using imaginary numbers.
    - The expression underneath the square roots determines the nature of the roots? Does this sound familiar? It should. Like quadratics, cubics (and indeed all polynomials) maintain the concept of a discriminant – in fact, the expression underneath the square roots is equal to $-27a^2\Delta$, where Δ is the discriminant.
    - If the discriminant is positive, the cubic has three real solutions (and the *casus irreducibilis* issue). If the discriminant is zero, the cubic either has a double root and another distinct root, or a triple root (in either case, the roots are all real). If the discriminant is negative, the cubic has one real solution.