CUBICS

DAVID SMYTH

1. Solving Polynomial Equations: High School Approach

Most of modern algebra was constructed in order to come to grips with the following problem: Given a polynomial

$$f(x) = a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n,$$

how can we write down a number α such that $f(\alpha) = 0$. For concreteness, let's think of $a_0, \ldots, a_n$ as rational numbers.

Linear Equations ($n = 1$). I know that you know how to solve a linear equation, but humor me. Consider the equation:

$$f(x) = a_0x + a_1 = 0.$$

We can divide through by a_0, then subtract a_1/a_0 from both sides.

$$x + \frac{a_1}{a_0} = 0$$
$$x = -\frac{a_1}{a_0}$$

Even though this is completely trivial, I would like to make two observations.

- If a_0 and a_1 are rational numbers, then our solution $-a_1/a_0$ is still a rational number. Similarly, if a_0 and a_1 are real numbers, then our solution is still a real number. You probably remember from high school that simple quadratic polynomials sometimes have complex solutions, but nothing like that happens here. In modern terminology, we can say that *no field extension is necessary* in order to find solutions of linear equations.

- Our first bit of algebra - dividing through by the leading coefficient a_0 - is actually a completely general recipe for reducing an arbitrary polynomial to a *monic* polynomial, i.e. a polynomial with $a_0 = 1$. If we can solve monic polynomials, we can solve all polynomials. Thus, from now on, I'll simply assume our polynomial is monic to begin with.

Quadratic Equations ($n = 2$).

$$f(x) = x^2 + a_1x + a_2 = 0$$

The basic trick here, probably known to most of you, is to make a substitution which cancels the a_1-term:

$$x = y - \frac{a_1}{2}$$

If we simply plug this in, we get

$$\begin{aligned}\left(y - \frac{a_1}{2}\right)^2 + a_1(y - \frac{a_1}{2}) + a_2 &= 0\\ \left(y^2 - a_1 y + \frac{a_1^2}{4}\right) + \left(a_1 y - \frac{a_1^2}{2}\right) + a_2 &= 0\\ y^2 + \left(a_2 - \frac{a_1^2}{4}\right) &= 0\\ y &= \pm\sqrt{\left(\frac{a_1^2}{4} - a_2\right)}\end{aligned}$$

Now that we have found y, we can go back and find x. We get

$$x = -\frac{a_1}{2} \pm \sqrt{\left(\frac{a_1^2}{4} - a_2\right)}$$

Once again, I would like to make two observations.

- It is no longer true that if our coefficients are rational, then we will necessarily have a rational solution. On the other hand, we can see precisely what we need to "add" in order to get solutions, namely $\sqrt{\left(\frac{a_1^2}{4} - a_2\right)}$. This is the only piece of the solution that might not be defined over the original field of definition. In modern terminology, we'll say that you always have a solution after passing to a degree two extension.
- Once again, the key trick here generalizes in an obvious way. Using a linear substitution of the form $x = y - a_1/n$, we can reduce an arbitrary degree-n equation to an equation satisfying $a_1 = 0$. We are making progress! Perhaps with enough tricks, we can solve any polynomial.

Cubic Equations ($n = 3$). Applying our previous two tricks, we may assume that we have an equation of the form:

$$f(x) = x^3 + a_2 x + a_3 = 0$$

Any ideas? As with the quadratic equations, it's easy to see that if we could cancel the a_2-term we'd be in business - we could simply take cube roots. The new trick is to recognize the possibility of a non-linear substitution:

$$x = y - \frac{a_2}{3y}$$

Let's see what happens when we plug this in:

$$\begin{aligned}\left(y - \frac{a_2}{3y}\right)^3 + a_2\left(y - \frac{a_2}{3y}\right) + a_3 &= 0\\ \left(y^3 - a_2 y + \frac{a_2^2}{3y} - \frac{a_2^3}{27y^3}\right) + \left(a_2 y - \frac{a_2^2}{3y}\right) + a_3 &= 0.\\ y^3 + a_3 - \frac{a_2^3}{27y^3} &= 0\end{aligned}$$

Clearing denominators, we get:

$$y^6 + a_3 y^3 - \frac{a_2^3}{27} = 0.$$

In most cases, going from a degree 3 equation to a degree 6 equation would not be considered progress. But here (rather miraculously), we can recognize this equation as a quadratic equation in y^3, i.e. we can rewrite it as:

$$(y^3)^2 + a_3(y^3) - \frac{a_2^3}{27} = 0.$$

Thus, we can solve for y^3 using the quadratic formula:

$$y^3 = -\frac{a_3}{2} \pm \sqrt{\frac{a_3^2}{4} + \frac{a_2^3}{27}}.$$

Taking cube roots, we get:

$$y = \sqrt[3]{-\frac{a_3}{2} \pm \sqrt{\left(\frac{a_3}{2}\right)^2 + \left(\frac{a_2}{3}\right)^3}}.$$

Now we should be a little careful here - just as one needs to account for positive and minus square roots, one needs to account for all possible cube roots. Thus, one really has the following solutions:

$$y = \omega^i \sqrt[3]{-\frac{a_3}{2} \pm \sqrt{\left(\frac{a_3}{2}\right)^2 + \left(\frac{a_2}{3}\right)^3}}, i = 0, 1, 2,$$

where $\omega = e^{2\pi i/3}$ is a third root of unity. (The fact that we are suddenly using complex numbers here is a little unsettling - eventually we will understand roots of unity in a purely algebraic way without any complex numbers entering the picture.) Plugging these solutions back into the formula $x = y - \frac{a_2}{3y}$, we obtain solutions of our original equation.

Optional Aside. You might notice something a little fishy here: We apparently have six possible solutions for y, each of which should give a solution for x. But this would give six possible solutions for our original cubic! In fact, there is no contradiction here. With some elementary but tedious algebra one can check that these six different value y, break into three pairs, with each pair giving the same value for x. The final solutions for x are:

$$\sqrt[3]{-\frac{a_3}{2} + \sqrt{\left(\frac{a_3}{2}\right)^2 + \left(\frac{a_2}{3}\right)^3}} - \sqrt[3]{-\frac{a_3}{2} - \sqrt{\left(\frac{a_3}{2}\right)^2 + \left(\frac{a_2}{3}\right)^3}}$$

$$\omega\sqrt[3]{-\frac{a_3}{2} + \sqrt{\left(\frac{a_3}{2}\right)^2 + \left(\frac{a_2}{3}\right)^3}} - \omega^2\sqrt[3]{-\frac{a_3}{2} - \sqrt{\left(\frac{a_3}{2}\right)^2 + \left(\frac{a_2}{3}\right)^3}}$$

$$\omega^2\sqrt[3]{-\frac{a_3}{2} + \sqrt{\left(\frac{a_3}{2}\right)^2 + \left(\frac{a_2}{3}\right)^3}} - \omega\sqrt[3]{-\frac{a_3}{2} - \sqrt{\left(\frac{a_3}{2}\right)^2 + \left(\frac{a_2}{3}\right)^3}}$$

What lessons can we draw from the cubic? The good news is that we are beginning to see what it means to "write down a number α such that $f(\alpha) = 0$." What we would like to do is find some formula for α in terms of the given coefficients $a_0, a_1, \ldots, a_n$. Based on the pattern we've seen so far, we would expect this formula to involve nothing more than the usual operations of addition/subtraction, multiplication/division, and also taking roots.

But there's also some bad news:

- The algebraic complexity of this problem is rapidly ballooning.
- The trick we used with the cubic does not seem to generalize in the way that our first two tricks did. The first two tricks serve to reduce us to the equation.
$$x^4 + a_2x^2 + a_3x + a_4$$
But it is not at all clear whether there is a substitution of the form $x = f(y, 1/y)$ which puts this in a simpler form.

Evidently, there are problems with the high school approach - these algebraic tricks feel unmotivated and to check that they work one has to do many unilluminating calculations. We need a different perspective on this problem. The basic shift in emphasis is to recognize the connection between finding roots and factoring polynomials.