Mathematics 307—October 2, 1995

A formula for solving cubic equations

Suppose we are given a cubic equation

$$P(x) = x^3 + ax^2 + bx + c = 0$$

and want to find its roots. When x is large the term x^3 will be larger than the others, far larger if x is very large, and therefore $P(x)$ will be positive if x is large and positive, negative if x is large and negative. This guarantees that somewhere in between $P(x)$ will be 0. In other words, there is always at least one root which is a real number (as opposed to a complex number). There may be just one and no more, for example if $P(x) = x^3 + x$ the only root is $x = 0$, since

$$P(x) = x(x^2 + 1)$$

and the second factor has only imaginary roots.

If there is just one real root, then there will be two more roots, but they will be conjugate complex numbers, as in the example above.

The case we shall most often see is the other case, when there are three real roots. This is true when $P(x)$ is the characteristic polynomail of a symmetric matrix, for example, and this is in fact the kind of matrix for which you should most often expect to find eigenvalues.

When $P(x)$ has three real roots, there is a reasonable way to find its roots explicitly. I will just present the method without justification. It has some features in common with the formula for finding the roots of quadratic equatons, but this sort of thing does not work for polynomials of degree greater than three, with an exception of sorts for degree four. The process comes in a few steps.

Step 1. The first step is very similar to what happens with a quadratic equation. If we want to solve

$$x^2 + ax + b = 0$$

then we *complete the square* to write this as

$$\begin{aligned} x^2 + ax + a^2/4 - a^2/4 + b &= 0 \\ (x + a/2)^2 &= a^2/4 - b \\ x + a/2 &= \pm\sqrt{a^2/4 - b} \\ x &= -a/2 \pm \sqrt{a^2/4 - b} \end{aligned}$$

we can do something similar in a first step towards simplifying a cubic equation. We have

$$(x + y)^3 = x^3 + 3x^2y + 3xy^2 + y^3$$

so we can complete the cube by seeing the terms

$$x^3 + ax^2$$

as the first few terms of

$$(x + a/3)^3 .$$

This is what we are going to do, but the algebra will be simpler if we think of it in a different way. Completeing the square amounts to substituting $x + a/2$ for x, or in other words changing the polynomial in x to one in a new variable $y = x + a/2$, substuting $y - a/2$ for x. For the cubic, we set

$$x = y - a/3 .$$

We then replace $P(x)$ by

$$\begin{aligned}(y-a/3)^3 + a(y-a/3)^2 + b(y-a/3) + c &= [y^3 - 3(a/3)y^2 + 3(a/3)^2y - (a/3)^3] \\ &\quad + a[y - 2(a/3)y + (a/3)^2] + b[y - ab/3] + c \\ &= y^3 + [b - a^2/3]y + [c - ab/3 + 2a^3/27] \\ &= y^3 + b_\bullet y + c_\bullet \\ b_\bullet &= b - a^2/3 \\ c_\bullet &= c - ab/3 + 2a^3/27 .\end{aligned}$$

So now we want to solve a cubic equation

$$y + b_\bullet y + c_\bullet = 0$$

with no term of degree two in the unknown. For the quadratic equation, this single step turned out to be all that was necessary, but for a cubic equation there is usually more work to do.

Step 2. In this step we are also going to substitute to get a new variable z, but now the substitution is

$$y = \alpha z$$

where α is a suitable constant. The point is that there is a very special kind of cubic equation which can be solved in a very satisfactory way, and we want to reduce the one we have to one of these special ones.

The idea comes from trigonometry. The cosine sum formula tells us

$$\cos(p+q) = \cos p \cos q - \sin p \sin q$$

so that if θ is any angle then

$$\begin{aligned}\cos 2\theta &= \cos^2\theta - \sin^2\theta \\ &= \cos^2\theta - (1 - \cos^2\theta) \\ &= 2\cos^2\theta - 1 \\ \sin 2\theta &= 2\sin\theta\cos\theta \\ \cos 3\theta &= \cos 2\theta\cos\theta - \sin 2\theta\sin\theta \\ &= [2\cos^2\theta - 1]\cos\theta - [2\cos\theta\sin\theta]\sin\theta \\ &= 2\cos^3\theta - \cos\theta - 2\cos\theta\sin^2\theta \\ &= 2\cos^3\theta - \cos\theta - 2\cos\theta(1-\cos^2\theta) \\ &= 4\cos^3\theta - 3\cos\theta .\end{aligned}$$

In other words, if we are given an equation

$$4z^3 - 3z - C = 0$$

or

$$z^3 - (3/4)z - C/4 = 0$$

where $|C| \leq 1$ then we can solve it by setting $x = \cos\theta$ where

$$\cos 3\theta = C, \quad \theta = \frac{\cos^{-1} C}{3} .$$

So now we set

$$y = \alpha z$$

to get the equation

$$\alpha^3 z^3 + b_\bullet \alpha z + c_\bullet = 0$$
$$z^3 + \frac{b_\bullet}{\alpha^2} z + \frac{c_\bullet}{\alpha^3} = 0 .$$

To get it in the form we want we set

$$\frac{b_\bullet}{\alpha^2} = -\frac{3}{4}$$
$$\alpha = \sqrt{\frac{-4b_\bullet}{3}}$$

to get

$$z^3 - (3/4)z + \frac{c_\bullet}{\alpha^3} = 0$$

Set now

$$-\frac{\cos 3\theta}{4} = \frac{c_\bullet}{\alpha^3}$$
$$\theta = \frac{1}{3} \cos^{-1} \frac{-4c_\bullet}{\alpha^3}$$

to get a root

$$z_1 = \cos\theta .$$

How do we get the other roots? The angle 3θ is not really determined uniquely, since 3θ, $3\theta + 2\pi$, $3\theta + 4\pi$ are all essentially the same angle so we can also write down roots

$$z_2 = \cos(\theta + 2\pi/3)$$
$$z_3 = \cos(\theta + 4\pi/3)$$

and this gives all the roots.

There is something to be checked to make this business work. (1) We must have $b_\bullet \leq 0$ or we cannot take its square root; (2) we must also have $|4c_\bullet/\alpha^3| \leq 1$ in order to apply $\cos^{-1}$. It turns out that these conditions both hold precisely when we know that there are three real roots, but this is not quite a trivial fact.

I cannot resist making the historical remark that this method of solving cubic equations goes back to the Renaissance of the early 16th century, and that solving cubic equations was essentially the first new mathematical result contributed by Western civilization after the Greeks of the Alexandrian era more than one thousand years earlier.

Once we have the roots z_1 etc. we go backwards to find the roots x_1 etc.

Summary. We start with $x^3 + ax^2 + bx + c$, which we know to have all real roots. Then we set in succession

$$\begin{aligned}
b_\bullet &= b - a^2/3 \\
c_\bullet &= c - ab/3 + 2a^3/27 \\
\alpha &= \sqrt{-4b_\bullet/3} \\
C &= c_\bullet/\alpha^3 \\
\theta &= \frac{\cos^{-1}[-4C]}{3} \\
z_1 &= \cos(\theta) \\
z_2 &= \cos(\theta + 2\pi/3) \\
z_3 &= \cos(\theta + 4\pi/3)
\end{aligned}$$

to get finally the roots to the original equation:

$$\begin{aligned}
x_1 &= \alpha z_1 - a/3 \\
x_2 &= \alpha z_2 - a/3 \\
x_3 &= \alpha z_3 - a/3 \; .
\end{aligned}$$

Example.

Try

$$x^3 - 6x^2 + 11x - 6 = 0 \; .$$

In this case

$$\begin{aligned}
b_\bullet &= (11) - (-6)^2/3 \\
&= -1 \\
c_\bullet &= 0
\end{aligned}$$

so the new equation is

$$y^3 - y = 0$$

which factors as $y(y-1)(y+1) = 0$, so its roots are $y = 0, 1, -1$, and $x = 1, 2, 3$.

Example.

$$x^3 + 2x^2 - 5x - 6 = 0$$

Here

$$\begin{aligned}
b_\bullet &= -6.333333 \\
c_\bullet &= -2.074074 \\
\alpha &= 2.905933 \\
\cos 3\theta &= 0.338086 \\
\theta &= 1.225914/3 \\
z_1 &= 0.917663, \; etc. \\
x_1 &= 2 \\
x_2 &= -3 \\
x_3 &= -1
\end{aligned}$$

This is perfect for a programmable calculator, not quite so much fun to do by hand. The second example shows that even when the roots are integers, the intermediate calculations may involve numbers of a more complicated kind.