Solving the General Cubic Equation

I'm not sure where to find a book that explains all this. I did read this stuff in a book I found in the school library about 55 years ago, but now I've forgotten what the title was. Maybe you learned some of this in algebra class. Probably not, though. Most modern algebra classes stop treating the solution of polynomial equations after working out the quadratic formula.

First, we should think about the historical background. I looked part of this up in the Encyclopedia Brittanica, under the heading *Mathematics, History of.* When I first read about the general cubic equation, the solution was credited to a guy named Cardano, who lived in Italy in the sixteenth century. His book about mathematics was published in 1545, and like all the scientific books of his day, it was written in Latin. According to the encyclopedia, historians now believe that this work was not original with Cardano, but that he effectively stole the idea from another guy named Tartaglia, who got it from Scippione del Ferro about 30 years before Cardano published his book.

Now sixteenth century Italy was the heart of the Renaissance, the period of time in which European science and culture came alive again after a thousand year period of stagnation known as the Dark Ages. Gutenberg invented the printing press along about 1450, so books were becoming more and more widely available. Martin Luther formed his breakaway church in Germany in 1517. Columbus made his first voyage to the new world in 1492. New ideas were the order of the day. So it's hardly surprising that Cardano used a new kind of number, what we now call an imaginary number, to solve the general cubic equation.

These imaginary numbers (which, when added to the more familiar real numbers form a hybrid we call complex numbers) are both the strength and the weakness of Cardano's solution. They're a strength because they make the solution possible, and because Cardano's effort spurred further innovations by later mathematicians in which complex numbers played a central role. Indeed, most of modern mathematics would not even be possible without complex numbers and the associated notion of the complex plane. Today even physics and chemistry and biology use complex numbers to describe physical reality. But in Cardano's day, an imaginary number, or a complex number, was literally unthinkable, because nobody else had thought of them before.

But complex numbers are also the weakness of Cardano's solution, because they're not particularly easy to work with, and also because they arise in almost every case to which Cardano's method can be applied. You'll see what I mean once we've worked our way through the solution. Since I'm not sure how familiar you might be with the notion of a complex number, I have divided this exposition into nine parts:

(1) A review of the solution to the general quadratic equation.
(2) A geometric interpretation of the discriminant, $b^2 - 4ac$.
(3) Some examples of a quadratic equation with complex roots.
(4) Some observations on polynomial equations in general.
(5) Some further observations on mathematical notation.
(6) A geometric interpretation of the general cubic equation.
(7) The derivation of Cardano's solution to the general cubic equation.
(8) Some examples of cubic equations with real roots.
(9) A note on the cubic discriminant function, $e^2 + \frac{4d^3}{27}$.

(1) The Solution to the General Quadratic Equation

The general quadratic equation, or polynomial equation of the second degree, is traditionally expressed in the form

$$ax^2 + bx + c = 0 \tag{1}$$

where x is unknown, and a, b, and c are called coefficients.

Presumably you are already familiar with the solution to this equation, which is written in the following form.

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

For simplicity we will assume that the coefficients a, b, and c are real numbers. At a later stage of our investigation we will show that the case in which any of a, b, or c is imaginary or complex can be reduced to the simpler case we are considering now.

Now we can easily see how to calculate the solution expressed by formula (2) so long as the quantity $b^2 - 4ac$ appearing under the square root sign is a positive number. All we have to do is add, subtract, multiply, and divide .... and also calculate that pesky square root. But for many purposes (as in answers to math problems in textbooks) simply specifying the square root as part of the solution will work perfectly well, since we can manipulate the square root algebraically without really worrying what the value of it is until we get to the end of our entire chain of reasoning. And there are many tricks for reducing the number of square roots appearing in an algebraic expression ... I'll show you some of those later.

No, our real problem arises when $b^2 - 4ac$ is a negative number. How can we calculate the square root then? After all, the square of every real number is a positive real number, since $(-1) * (-1)$ is $+1$. That's even why we write the $\pm$ sign in formula (2) to remind ourselves that every number has two square roots, one of them being the negative of the other.

Clearly the square root of a negative real number cannot be a real number. So mathematicians have invented a new kind of number, called *imaginary*, to serve in this circumstance. Formally, we write

$$\sqrt{-1} = i \tag{3}$$

where i represents the imaginary unit. (This notation using i to represent the square root of minus one was popularized by Karl Friedrich Gauss, possibly the greatest mathematician who ever lived. In Cardano's day people used other terms to refer to it. And many engineering textbooks today use j instead of i. No matter what we call it, though, it remains the same number.)

Why do we only need to assign a special symbol for the square root of minus one? What about other numbers, like the square root of minus three? Well, we just make some arbitrary assumptions about this new number i, namely that it behaves just like a familiar real number, except for the peculiar property that when it is squared, the result is -1. And then we work with it as if it were a real number ... we can add it to a real number to form a complex number, or we can multiply it or divide it by any real number, or even by a complex number. Here are a few examples.

$$i * i = i^2 = -1 \tag{4}$$
$$i^3 = i^2 * i = -1 * i = -i \tag{5}$$
$$i^4 = i^2 * i^2 = (-1) * (-1) = 1 \tag{6}$$
$$(2 + 3i) * (2 - 3i) = 2^2 - 9i^2 = 4 + 9 = 13 \tag{7}$$
$$(3 + i) * (3 - i) = 3 * 3 + 3 * i - 3 * i - i * i = 3 * 3 - i^2 = 9 + 1 = 10 \tag{8}$$

These examples illustrate the fact that every complex number has a *complement*, or mirror-image. In general, if $z = a + bi$ (where a and b are real numbers), the complement of z is $z^* = a - bi$. And, as the examples indicate, $zz^* = a^2 + b^2$.

So what if we were asked to symbolize the square root of -3, or of -25? We would just resort to a simple algebraic property of square roots:

$$\sqrt{ab} \equiv \sqrt{a} * \sqrt{b} \tag{9}$$

so that

$$\sqrt{-3} = \sqrt{(-1) * (3)} = \sqrt{-1} * \sqrt{3} = \pm i\sqrt{3} \quad \text{and} \tag{10}$$
$$\sqrt{-25} = \sqrt{(-1) * 25} = \sqrt{-1} * \sqrt{25} = \pm 5i \tag{11}$$

OK, one more trick with complex numbers, and then we've got to move on to the next topic. What is the square root of i itself? Believe it or not, the answer is sort of easy to work out. First, let's look at a very simple complex number, $1 + i$:

$$(1 + i)^2 = (1 + i)(1 + i) = 1 + 2i + i^2 = 1 + 2i - 1 = 2i \quad \text{so clearly} \tag{12}$$
$$\sqrt{2i} = \pm(1 + i) \tag{13}$$

but then we can use the identity (9) above to assert that

$$\sqrt{2i} = \sqrt{2}\sqrt{i} \tag{14}$$
$$\sqrt{i} = \frac{\sqrt{2i}}{\sqrt{2}} = \frac{(1 + i)}{\sqrt{2}} = \pm\frac{(1 + i)\sqrt{2}}{2} \tag{15}$$

or, in decimal fractions, the square root of i is $\pm(.707 + .707i)$, very nearly.

(2) Geometric Interpretation of the Discriminant

Now the expression $b^2 - 4ac$ appearing in formula (2) above is commonly called the discriminant of the general quadratic equation (1). If the discriminant is positive, equation (1) has two solutions, or roots, both of which are real numbers. If the discriminant is negative, (1) still has two roots, both of which are complex numbers (or possibly pure imaginary numbers, if $b = 0$).

We can give this result a geometric interpretation by resorting to the simple expedient of considering the left hand side of equation (1) to be a function of the variable quantity x. In symbols we write

$$y = f(x) = ax^2 + bx + c \tag{16}$$

where we read "a function of x" for the symbols $f(x)$. Once we do that, the problem of solving equation (1) is the same as asking ourselves for what values of x is $f(x) = 0$? We can proceed to compute $f(x)$ for a few values of x, and then draw a graph of the resulting shape ... this will at least help us to guess what the solutions are, for we will be able to see the points where the curve representing $f(x)$ crosses the x-axis, and these are clearly the desired solutions.

Working our way through an example will make this process somewhat clearer. Let's choose particular coefficients for our quadratic function, then compute the values of $f(x)$ for a few points x, then draw the resulting graph. Let's keep it simple – I choose $f(x) = x^2 + 4x + 2$. Some values are tabulated below.

x	x^2	$4x$	$f(x)$
-5	25	-20	7
-4	16	-16	2
-3	9	-12	-1
-2	4	-8	-2
-1	1	-4	-1
0	0	0	2
1	1	4	7

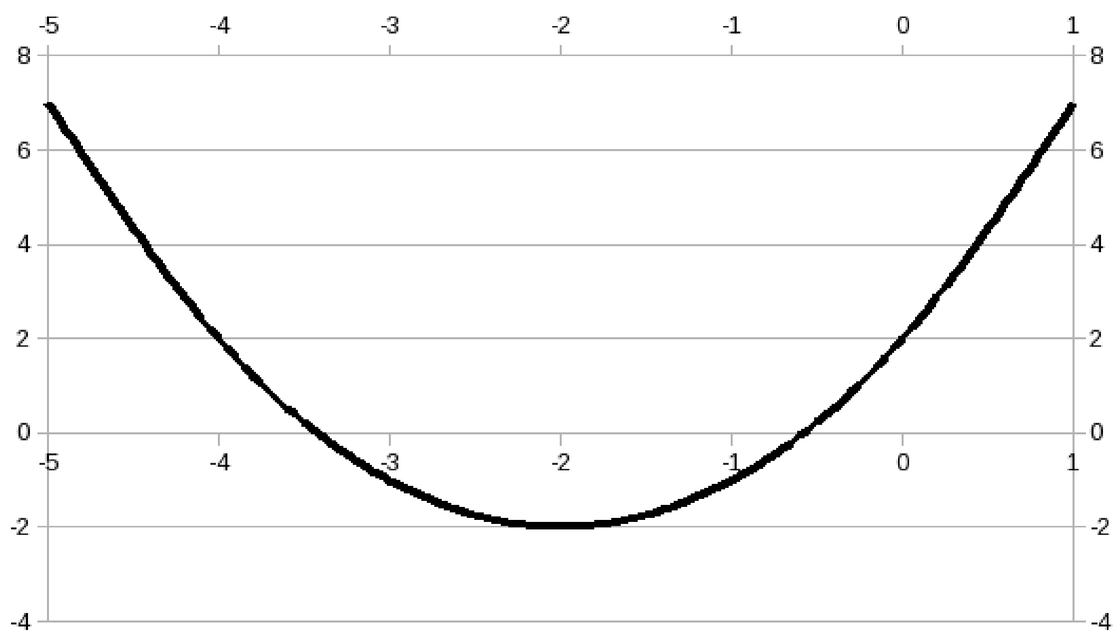

Figure 1: The function $f(x) = x^2 + 4x + 2$.

In Figure 1 I've drawn the graph of this function. From the picture we can see that the roots of the equation $x^2 + 4x + 2 = 0$ must lie somewhere near $x = -3.5$ and $x = -0.5$. And in fact if we use formula (2) we find that the roots of our example are given by

$$\frac{1}{2} * (-4 \pm \sqrt{4^2 - 8}) = -2 \pm \sqrt{2} \approx -3.41 \text{ or } -0.59 \tag{17}$$

Now as it turns out, every quadratic curve has the same general shape as the one sketched in figure 1. The only things that change from one quadratic to the next are its placement relative to the origin, and the direction it "points" (*i.e.*, if the coefficient of x^2 is a negative number, the parabola will have its nose up, not down as shown.)

So far we've been working with a quadratic function for which the discriminant is a positive number. What happens when the discriminant is negative? Well, we can see that the square root in formula (2) will be the square root of a negative number. But what does this look like geometrically? Let's make up another quadratic function that has a negative discriminant – I choose $h(x) = x^2 + 4x + 5$.

x	x^2	$4x$	$h(x)$
-5	25	-20	10
-4	16	-16	5
-3	9	-12	2
-2	4	-8	1
-1	1	-4	2
0	0	0	5
1	1	4	10

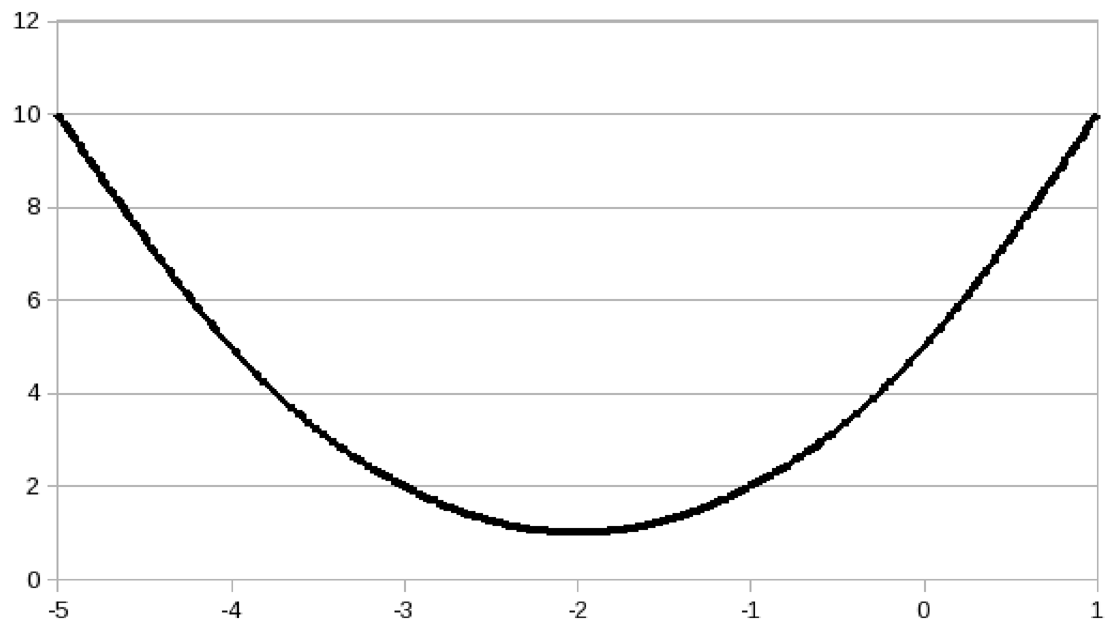

Figure 2: The function $h(x) = x^2 + 4x + 5$.

Figure 2 shows why there are no real roots of my chosen function $h(x)$ – the parabola is simply positioned too high up to cross the x-axis. What is harder to see, and something I won't even attempt to illustrate, is why this function has two zeroes when x is allowed to assume complex values. The points at which this occurs are $(-2+i)$ and $(-2-i)$. These are easily worked out using formula (2), and we can verify that they are indeed roots, or zeroes, of $h(x)$, as follows:

$$h(-2+i) = (-2+i)^2 + 4(-2+i) + 5 \tag{18}$$
$$= (4-4i-1) + (-8+4i) + 5 \tag{19}$$
$$= 3-8+5 = 0 \tag{20}$$
$$h(-2-i) = (-2-i)^2 + 4(-2-i) + 5 \tag{21}$$
$$= (4+4i-1) + (-8-4i) + 5 \tag{22}$$
$$= 3-8+5 = 0 \tag{23}$$

(3) Some Examples of Quadratic Equations with Complex Roots

So now we've worked our way through one example of a quadratic equation whose two roots are complex numbers. Here are a few equations for you to work out – every one can be solved by using formula (2). When you have time, try working some of these out, and be sure to express your answers in the form $(a \pm bi)$.

$$x^2 + 6x + 10 = 0 \tag{24}$$
$$3x^2 + 6x + 15 = 0 \tag{25}$$
$$7x^2 + 9x + 3 = 0 \tag{26}$$
$$x^2 + 2x + 5 = 0 \tag{27}$$

Can you explain why equations (25) and (27) have the same solutions?

(4) Some Observations on the General Polynomial Equation

By now you have a good idea of how to deal with the second degree polynomial equation $ax^2 + bx + c = 0$. And we're still working toward an understanding of Cardano's solution of the general cubic equation $x^3 + ax^2 + bx + c = 0$. To make the rest of our work a bit easier, I want to make a few assertions about polynomial equations of any degree. I'm going to offer these without proof. Please trust me on some of this stuff. Many of the proofs are pretty simple, and you can probably work them out for yourself. I just don't have time to put all the proofs in this document – it's getting too unwieldy as it is.

(1) We can reduce any polynomial equation to canonical form, in which the coefficient of the term of highest degree is unity, without affecting the roots of the polynomial equation $P(x) = 0$. In symbols, this is the same thing as saying that the two equations $ax^n + bx^{n-1} + ... + yx + z = 0$ and $x^n + \frac{b}{a}x^{n-1} + ... + \frac{y}{a}x + \frac{z}{a} = 0$ have the same solutions. Equations (25) and (27) provide an example.

(2) Every polynomial equation $P(x) = 0$ has at least one solution in the field of complex numbers if the coefficients of $P(x)$ are real numbers. This important result was first proved by Karl Friedrich Gauss about 200 years ago – it is known as the *Fundamental Theorem of Algebra*.

(3) There is a process known as synthetic division by means of which we may divide one polynomial expression by another to obtain a quotient $Q(x)$ and a remainder $R(x)$. This procedure is in many ways directly analagous to the procedure of long division which is in common use.

(4) If we divide a polynomial $P(x)$ by the simple linear polynomial $(x - t)$, then the remainder polynomial $R(x)$ will take on the same value at the point t as $P(x)$ does. This is known as the *Remainder Theorem.* In symbols, if $P(x) = (x - t)Q(x) + R(x)$, then $[P(t)] = [R(t)]$ where the square brackets [ ] indicate that we are to calculate the value of the polynomial expressions P and R at the point $x = t$.

(5) We call the highest power of x appearing in the polynomial $P(x)$ the *degree* of P. If $P(x)$ has real coefficients and is of odd degree, it must have at least one real root, but if it is of even degree, it may or may not have any real roots. We will actually approach the proof of this assertion when we look at the general cubic polynomial in more detail a bit later.

(5) Some Further Observations on Mathematical Notation

No doubt you're already familiar with the concept of exponentiation, and with the idea of extracting a square root. Extending the idea of a square root to roots of greater degree (as a cube root, a fourth (or *biquadratic*) root, a fifth root, etc.) should be fairly easy. For example, we write $\sqrt[3]{3}$ to denote the cube root of 3, which is any number z such that $z^3 = 3$. Similarly, we write $\sqrt[5]{5}$ to denote the fifth root of 5, which is any number z such that $z^5 = 5$.

All this seems obvious enough. But have you ever stopped to consider the idea that the two processes of exponentiation (raising a number to a power) and of extracting roots are inverse processes? Just like addition and subtraction, or multiplication and division, these two processes go together. An example might help.

Clearly, $a^2 = \sqrt[3]{a^6}$, since $(a^2)^3 = a^2 * a^2 * a^2 = a^6$. (For now, let's not think too hard about the fact that a^6 may have other cube roots besides a^2 ... I'm trying to make a point about notation here.)

In fact, mathematicians have seized on two facts (that are easy to prove when a and b are positive integers) to generalize the laws of exponents. If x is any non-zero number and a and b are any real or complex numbers, then

$$(x^a) * (x^b) \equiv x^{a+b} \tag{28}$$

$$(x^a)^b \equiv x^{ab} \tag{29}$$

$$x^0 \equiv 1 \tag{30}$$

Equation (30) is easy to understand if we stop to think that $x^{-a} = \frac{1}{x^a}$. Do you see why?

The point of this whole digression into the laws of exponents is simple – once we understand that we can write a root of a number as a fractional power of that same number, a lot of equations become easier to write down. For instance, we don't need to talk about square roots any longer – we can talk about the $\frac{1}{2}$ power instead. And the cube root can be expressed as the $\frac{1}{3}$ power. Using our new notational convention, formula (2) can be rewritten as

$$x = \frac{-b \pm (b^2 - 4ac)^{\frac{1}{2}}}{2a} \tag{31}$$

Please forgive me for being so lazy. Writing all those formulas with the $\sqrt{}$ sign is really a lot of work. I hope the notation using fractional exponents seems clear to you – it sure is a lot easier for me to put together on my PC! One other thing. Once you get the idea of fractional exponents clearly in mind, it will be easy to understand logarithms, which are closely related to exponents. Maybe I'll write about those some day, when I've got a little spare time.

(6) A Geometric Interpretation of the General Cubic Equation

With all those preliminaries out of the way, we're finally ready to tackle Cardano's solution to the general cubic equation. Let's start with a simpler equation, one that is very easy to solve:

$$x^3 + a = 0 \tag{32}$$

Apparently, one solution of (32) is given by the cube root of $-a$. Using our new notation, we can also write $x = -a^{\frac{1}{3}}$ and so long as a is a real number, we can be sure that at least one value of $\sqrt[3]{-a}$ is also a real number, since the cube of -1 is -1. Does that make sense?

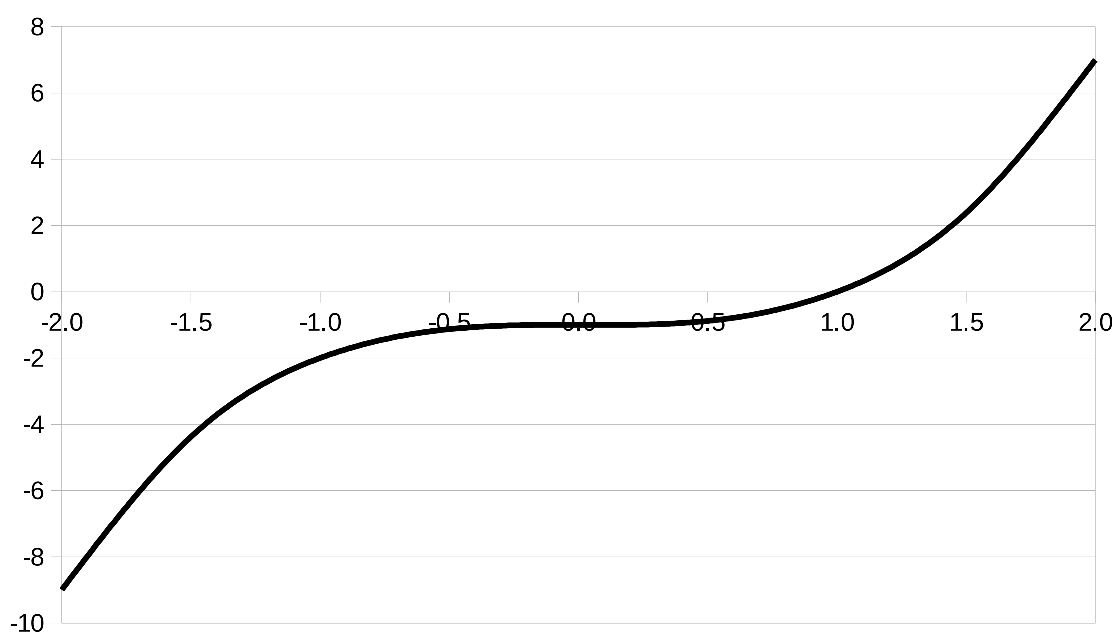

Figure 3: The function $g(x) = x^3 - 1$.

Figure 3 illustrates the shape of the polynomial function $x^3 - 1$. As expected, this curve crosses the x-axis at the point $x = 1$. More importantly, this curve runs off toward plus and minus infinity as x assumes arbitrarily large values. Intuitively, this is why equation (32) must have a real root so long as a is a real number. Since the quantity x^3 assumes a positive value when x is positive and a negative value when x is negative, and since x^3 increases (decreases) without limit as x varies, we can be sure that no matter how great a number a may be, we can find a value of x such that x, and therefore x^3, exceeds it.

Notice the region $-1 < x < 1$ in Figure 3. In this interval the function $x^3 - 1$ first curves to the right, then straightens out and curves to the left before heading north at a great rate. Why does this happen? (You can gain insight by thinking about the behavior of the absolute value of x^3, or $|x^3|$, when $|x| < 1$ and then contrasting that with $|x^3|$ when $|x| > 1$. What if you visualize x^3 as a mapping which carries each point $x \to x^3$? When x is zero (or one, or minus one), the mapping carries x into itself. When x is near zero (or 1, or -1), the mapping doesn't move x very far. But when x is a fairly large number, say 10, the mapping moves x a great distance.)

What happens when a cubic polynomial $P(x)$ includes terms in x and in x^2? Well, the flattish region in Figure 3 near $x = 0$ gets trickier. In general, it will get a hump in it, wiggling up and down one time before making the hard left turn and heading almost straight north. Depending on its coefficients, $P(x)$ may have three real roots. Or it may only have one. You can visualize this by thinking about the bumpy region in the curve. All the bumps may lie below the x-axis. Or one may fall above and the other below it. Or both bumps may be above the x-axis.

Figure 4 below is included to make these ideas clearer. I have charted three different polynomials in Figure 4 – once we have developed Cardano's method, we will use these three polynomials as actual examples.

$$P_1(x) = x^3 - 2x^2 - 5x + 6 \tag{33}$$

$$P_2(x) = x^3 - 2x^2 - 5x - 5 \tag{34}$$

$$P_3(x) = x^3 - 2x^2 - 5x + 11 \tag{35}$$

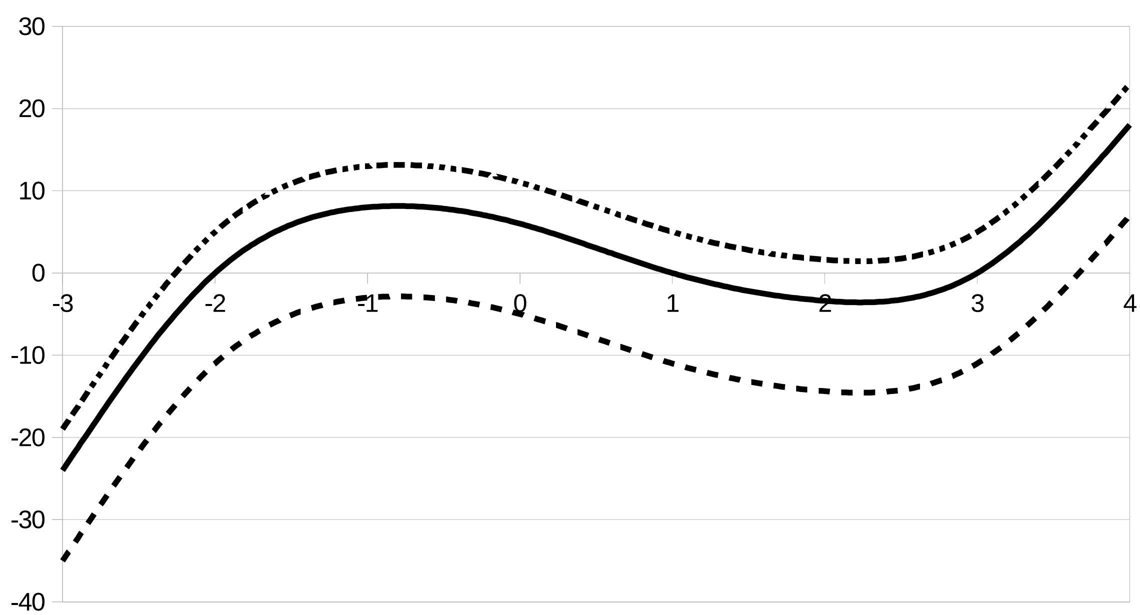

Figure 4: Three cubic polynomial functions

In Figure 4 above, the solid black curve is $P_1(x)$, the dashed curve is $P_2(x)$, and the dashed / dotted curve is $P_3(x)$. Notice how all three curves have the same shape. This should be obvious – the only difference in the three polynomials is in the constant term.

Now let's look more closely to pick out the values of x for which each of these polynomials takes the value zero. P_1 clearly has three zeroes at or near -2, +1, and +3. Each of P_3 and P_2 has a single zero, near -2.5 and +3.5, respectively.

In general, these three curves show how every real-valued cubic polynomial function behaves. The bumps may get bigger or smaller, but there are always two of them. And for values of x which are sufficiently positive or negative, the polynomial curve looks just like the tail on the simple cubic illustrated in figure 3. When the bumps are in the right position, the cubic polynomial has three real roots. Shift the curve up or down a bit, and two of the roots disappear. But every cubic polynomial with real coefficients has at least one real root, for the graph of such a polynomial must cross the x-axis an odd number of times.

(7) Deriving Cardano's Solution to the General Cubic Equation

So now we're finally ready to derive Cardano's solution to the general cubic equation for ourselves. We begin by stating the problem – find the roots of a third degree polynomial, which we will write in canonical form for convenience:

$$x^3 + ax^2 + bx + c = 0 \tag{36}$$

The first step in our procedure is to eliminate the term ax^2 from equation (36). We may accomplish this by a substitution of variables. Let us write $x = y + k$, where k is a quantity we don't yet know, but wish to determine. Substituting this expression into (36) we obtain

$$(y+k)^3 + a(y+k)^2 + b(y+k) + c = 0 \quad \Longrightarrow \tag{37}$$

$$(y^3 + 3ky^2 + 3k^2y + k^3) + a(y^2 + 2ky + k^2) + by + bk + c = 0 \tag{38}$$

Gathering terms in like powers of y together we see that

$$y^3 + (3k + a)y^2 + (3k^2 + 2ak + b)y + (k^3 + ak^2 + bk + c) = 0 \tag{39}$$

Now we equate the coefficient of y^2 in equation (39) to zero and solve for k:

$$(3k + a) = 0 \quad \Rightarrow \quad k = -a/3; \text{ or } x = y - a/3 \tag{40}$$

and substituting this value of k back into equation (36) we obtain Cardano's first reduced equation:

$$y^3 + (-a + a)y^2 + \left(3\left(\frac{a}{3}\right)^2 - 2\frac{a^2}{3} + b\right)y + \left(-\frac{a^3}{27} + \frac{a^3}{9} - \frac{ab}{3} + c\right) = 0 \quad \Longrightarrow \tag{41}$$

$$y^3 + \left(b - \frac{1}{3}a^2\right)y + \left(c - \frac{1}{3}ab + \frac{2}{27}a^3\right) = 0 \tag{42}$$

Before we proceed I'd like to make a few observations. First, we employed the binomial theorem in (37) and (38) above to expand the term $(x+k)^3$ – in case you're not familiar with that, we may simply observe that $(m+n)^3 = m^3 + 3m^2n + 3n^2m + n^3$. Second, it might be wise at this point to make the substitution $x = y - (1/3)a$ in equation (36) on your own to be sure you can obtain the result in (42) without using the intermediate variable k.

Finally, we will rewrite equation (42) to make it a bit simpler before proceeding with the tricky part:

$$y^3 + dy + e = 0 \qquad \text{where} \qquad d = b - \frac{1}{3}a^2 \qquad \text{and} \qquad e = c - \frac{1}{3}ab + \frac{2}{27}a^3 \tag{43}$$

So now we've eliminated the term of second degree in our original equation. If we could only eliminate the term in the first degree as well (dy in (43)) we'd be home free, since the equation $y^3 = k$ is easy to solve. So let's make another substitution of variables and see where that leads us.

$$w + k = y \quad \Longrightarrow \tag{44}$$

$$(w+k)^3 + d(w+k) + e = 0 \quad \Longrightarrow \tag{45}$$

$$w^3 + 3kw^2 + 3k^2w + k^3 + dw + dk + e = 0 \quad \Longrightarrow \tag{46}$$

$$w^3 + \left(3kw^2 + dw\right) + \left(3k^2w + dk\right) + k^3 + e = 0 \tag{47}$$

Now if we could just get the two expressions enclosed in parentheses to go away, we would have a much simpler looking equation. But if we assume that neither w nor k is equal to zero in equation (47) we can write

$$3kw^2 + dw = (3kw + d)w \quad = 0 \quad \text{if and only if} \quad 3kw + d = 0 \tag{48}$$

$$3k^2w + dk = (3kw + d)k \quad = 0 \quad \text{if and only if} \quad 3kw + d = 0 \tag{49}$$

and so we are motivated to write $k = -\frac{1}{3}\frac{d}{w}$ and make the substitution in (43) all over again.

$$w - \frac{d}{3w} = y \quad \Longrightarrow \tag{50}$$

$$\left(w - \frac{d}{3w}\right)^3 + d\left(w - \frac{d}{3w}\right) + e = 0 \quad \Longrightarrow \tag{51}$$

$$w^3 - 3w^2\frac{d}{3w} + 3w\left(\frac{d}{3w}\right)^2 - \left(\frac{d}{3w}\right)^3 + dw - \frac{d^2}{3w} + e = 0 \quad \Longrightarrow \tag{52}$$

$$w^3 - wd + \frac{d^2}{3w} + dw - \frac{d^2}{3w} - \left(\frac{d}{3w}\right)^3 + e = 0 \quad \Longrightarrow \tag{53}$$

$$w^3 - \left(\frac{d}{3w}\right)^3 + e = 0 \tag{54}$$

Multiplying this last equation by $w^3 = w^3$ we obtain

$$w^6 + ew^3 - \left(\frac{d}{3}\right)^3 = 0 \tag{55}$$

But equation (55) is just a quadratic equation in the variable w^3! Since e and d are already known (see formula 43), we may rewrite (55) and solve it using the familiar quadratic formula (2):

$$z = w^3 \quad \Longrightarrow \quad z^2 + ez - (1/27)d^3 = 0; \quad \therefore \quad w^3 = \frac{1}{2}\left(-e \pm \sqrt{e^2 + \frac{4}{27}d^3}\right) \tag{56}$$

And this formula (56) is Cardano's famous solution. All we need to do now is to compute the value of w^3, from that extract the cube root to find w, then substitute w back through equation (50) to find y, and then run y back through formula (40) to find x, a root of our original equation (36). Whew!

(8) Some Examples of Cubic Equations with Real Roots

So now let's apply Cardano's formula to one of the three polynomials we met a while ago, namely $P_1(x)$:

$$x^3 - 2x^2 - 5x + 6 = 0 \tag{57}$$

First we must transform this equation to eliminate the term in x^2.

$$x = y - \frac{-2}{3} = y + \frac{2}{3} \quad \Longrightarrow \tag{58}$$

$$\left(y + \frac{2}{3}\right)^3 - 2\left(y + \frac{2}{3}\right)^2 - 5\left(y + \frac{2}{3}\right) + 6 = 0 \quad \Longrightarrow \tag{59}$$

$$y^3 - \frac{19}{3}y + \frac{56}{27} = 0 \tag{60}$$

(See equations (40) through (42) for a reminder of how we got to this point.)

Now we plug the coefficients from the auxiliary equation (60) back into formula (56):

$$w^3 = \frac{1}{2}\left[-\frac{56}{27} \pm \sqrt{\left(\frac{56}{27}\right)^2 + \frac{4}{27}\left(\frac{-19}{3}\right)^3}\right] \tag{61}$$

$$= \frac{1}{27}\left[-28 \pm \sqrt{28^2 - 19^3}\right] \tag{62}$$

$$= \frac{1}{27}\left[-28 \pm \sqrt{784 - 6859}\right] \tag{63}$$

$$= \frac{1}{27}\left[-28 \pm i\sqrt{6075}\right] \tag{64}$$

$$= \frac{1}{27}\left[-28 \pm 45i\sqrt{3}\right] \tag{65}$$

Now this looks a bit messy. But let's not give up yet – the factor $\frac{1}{27}$ is easily disposed of, since $3^3 = 27$, and we are left to ponder a problem: how to extract a cube root from $\left[-28 \pm 45i\sqrt{3}\right]$

We must find a number which, when cubed, yields the value shown. Clearly the number we seek must be a complex number, since every real number yields another real number when it is cubed. So let's stop for a minute and think about the properties of complex numbers. Let's apply the binomial theorem to expand $(a+bi)^3$ and see if that gives us any ideas.

$$(a+bi)^3 = a^3 + 3a^2bi + 3a(bi)^2 + (bi)^3 \quad (66)$$
$$= (a^3 - 3ab^2) + (3a^2b - b^3)i \quad (67)$$
$$= a(a^2 - 3b^2) + b(3a^2 - b^2)i \quad (68)$$

This at least gives us an idea of how to proceed. Let's find a number that divides the real part of expression (65), then assume it's equal to a in (68). We can then factor the real part in (65), set $a^2 - 3b^2$ equal to the quotient, solve for b^2, and then substitute our hypothetical values for a and b into the imaginary part of expression (68) to be sure that it's going to work. Let's try it.

For our first try, let's choose $a = -4$, since (-4) * 7 = -28.

$$((-4)^2 - 3b^2) = 7 \quad (69)$$
$$3b^2 = 16 - 7 = 9 \quad (70)$$
$$b^2 = 3 \quad \text{or} \quad b = \sqrt{3} \quad (71)$$

Now we need to substitute this value for b back into the formula for the imaginary part of $(a+bi)^3$:

$$b(3a^2 - b^2) = \sqrt{3}(3*(-4)^2 - 3) = \sqrt{3}(3*16 - 3) = 45\sqrt{3} \quad (72)$$

Hoorah! This lucky guess works out just right. One of the desired solutions to equation (65) must be given by

$$w = \frac{-4 \pm i\sqrt{3}}{3} \quad (73)$$

Now one of the peculiar features of complex numbers is that every one of them, except zero, has exactly three distinct cube roots. So there are in fact two more solutions for equation (65). Rather than spend time on them here, I'll leave it to you to find them – but I'll give you a hint. You need to consider these two relationships to get there:

$$(7/2)*(-8) = -28 \qquad (1/2)*(-56) = -28 \quad (74)$$

Let's proceed to find the roots of our original cubic equation (57) by substituting the value of w we've already located back into equation (50): $y = w - \frac{d}{3w}$; w is given by (73), and $d = -\frac{19}{3}$ (see equation (60)).

Let's take the plus sign in (73) and observe that $3w = -4 + i\sqrt{3}$. Then

$$y = \frac{-4 + i\sqrt{3}}{3} - \frac{-19}{3} * \left[-4 + i\sqrt{3}\right]^{-1} \quad (75)$$

$$\text{But} \quad \frac{1}{-4 + i\sqrt{3}} = \frac{1}{-4 + i\sqrt{3}} * \frac{-4 - i\sqrt{3}}{-4 - i\sqrt{3}} = \frac{-4 - i\sqrt{3}}{19} \quad (76)$$

$$\text{So} \quad y = \frac{-4 + i\sqrt{3}}{3} + \frac{19}{3} * \frac{-4 - i\sqrt{3}}{19} = -\frac{8}{3} \quad (77)$$

$$\text{and then} \quad x = y + \frac{2}{3} \implies x = \frac{2-8}{3} = -2 \tag{78}$$

So at long last we have extracted a root from equation (57), namely, $x = -2$. Do you see why I said at the beginning that Cardano's formula is interesting, but not very useful? That was a lot of work!

Anyway, now that we've located one of the roots of equation (57), the rest of them will be pretty easy to find. A while ago I mentioned a process called synthetic division, and the Remainder Theorem. Since $P_1(-2) = 0$, we can be sure that $(x + 2)$ divides $P_1(x)$ evenly. I've laid the process out below kind of like a long division problem. I hope this makes sense.

$$\begin{array}{llrrrr}
 & & x^2 & -4x & +3 & \\
(x+2) & \text{into} & x^3 & -2x^2 & -5x & +6 \\
 & & x^3 & +2x^2 & & \\
 & & & -4x^2 & -5x & \\
 & & & -4x^2 & -8x & \\
 & & & & 3x & +6 \\
 & & & & 3x & +6 \\
 & & & & & 0
\end{array}$$

And we can indeed verify our division by observing that

$$(x+2)(x^2 - 4x + 3) = x^3 - 4x^2 + 3x + 2x^2 - 8x + 6 = x^3 - 2x^2 - 5x + 6 = P_1(x) \tag{79}$$

So now we can find the other two roots of equation (57) by solving the quadratic equation

$$x^2 - 4x + 3 = 0 \tag{80}$$

which is easily resolved into $(x - 1)(x - 3)$, implying that the remaining roots of equation (57) are 1 and 3.

I had intended to work through the solution of two more examples involving the polynomials $P_2(x)$ and $P_3(x)$ introduced above. But that's just too much work for me right now ... I'll write about those two another time. For now, let's take a couple of simpler equations to illustrate an important point about Cardano's formula.

$$x^3 + 3x^2 - 9x + 5 = 0 \tag{81}$$

$$x^3 - x^2 - 8x + 12 = 0 \tag{82}$$

Substituting $x = y - 1$ into equation (81) we obtain

$$y^3 - 12y + 16 = 0 \tag{83}$$

Substituting $y = w + \frac{12}{3w}$ into equation (83) and applying equation (56) we see that

$$w^3 = \frac{1}{2}\left[-16 \pm \sqrt{16^2 + \frac{4}{27}(-12)^3}\right] = \frac{1}{2}\left[-16 \pm \sqrt{256 - 4\frac{3^3}{27}(4^3)}\right] = -8 \tag{84}$$

$$\text{But} \quad w^3 = -8 \implies w = -2 \implies y = -2 + \frac{12}{-6} = -4 \tag{85}$$

And since $x = y - 1$, we have $x = -5$ as one of the roots of equation (81). But then we may divide the polynomial in (81) by $(x + 5)$ to obtain a reduced quadratic equation $x^2 - 2x + 1 = 0$, which evidently has two roots, both of which are equal to 1. And indeed, we can verify the three roots of equation (81) by noting that $(x-1)^2(x+5) = x^3+3x^2-9x+5$.

I'll bet you can work equation (82) ... it's just about as simple as (81). Why not give it a try?

(9) A Note on the Cubic Discriminant Function

The chief difficulty in Cardano's formula (56) involves the occurrence of complex numbers in the solution for w. Although it may seem a little odd, the fact is that formula (56) almost always generates complex values for w^3, even when the desired roots of the original cubic equation are real numbers. Why do you think that happens?

We can gain real insight by thinking about the fundamental theorem of algebra and the Remainder Theorem. Let's suppose that we're given a cubic equation with three distinct real roots, none of which is zero, and let's call those roots r, s, and t. For convenience let's further suppose that the equation has already been reduced so that the term in y^2 has been removed, as in formulas (41) and (42) above. Using the Remainder Theorem we can prove that

$$P(y) = (y - r)(y - s)(y - t) \tag{86}$$

so that the equation $P(y) = 0$ can also be written as

$$(y - r)(y - s)(y - t) = 0 \tag{87}$$

Expanding the expression on the left we obtain

$$y^3 - (r + s + t)y^2 + (rs + rt + st)y - rst = 0 \tag{88}$$

In other words, the coefficients of the polynomial $P(y)$ can be expressed in terms of the roots of the equation. But since we've assumed that the coefficient of y^2 in our given equation is zero, we must have

$$r + s + t = 0 \quad \Longrightarrow \quad r + s = -t \tag{89}$$

In other words, the condition that our equation has no term in y^2 implies that one of the roots must be negative, and the other two must be positive, or else one of the roots is positive, and the other two are negative. Let's assume that r and s are the positive roots – the case where there are two negative roots is entirely analogous.

So r and s are positive, and $t = -r-s$; let's assume that s is the greater of the two positive roots (remember, we assumed that all three roots are distinct), so that we can write

$$s = (1 + k)r \quad \text{where} \quad k > 0 \quad \text{and} \quad t = -r - s = -r - (1 + k)r = -(2 + k)r \tag{90}$$

Now let's express the coefficients in equation (88) above solely in terms of r and k, then relate them back to Cardano's solution in formula (56). Let's start with the coefficient of

y, also called d in formula (56):

$$d = (rs + rt + st) = r^2(1+k) - r^2(2+k) - r^2(1+k)(2+k)$$
$$= r^2[1 + k - 2 - k - 2 - 3k - k^2] = r^2[-3 - 3k - k^2] = -r^2[3 + 3k + k^2]$$

Also, $e = -rst = (-r) * (1+k)r * [-(2+k)]r = r^3[2 + 3k + k2]$, where e denotes the constant term in Cardano's solution. Now that solution contains the expression $\sqrt{e^2 + \frac{4}{27}d^3}$ in the formula for w^3, so if $e^2 + \frac{4}{27}d^3$ is a negative number, we will be stuck with computing the cube root of a complex number when we apply the formula. But notice what happens when we substitute these expressions for d and e into this discriminant function for the cubic equation.

$$e^2 + \frac{4}{27}d^3 = (r^3[2 + 3k + k^2])^2 + \frac{4}{27}(-r^2[3 + 3k + k^2])^3$$

$$= r^6[(4 + 6k + 2k^2 + 6k + 9k^2 + 3k^3 + 2k^2 + 3k^3 + k^4)$$
$$+ \frac{-4}{27}(9 + 9k + 3k^2 + 9k + 9k^2 + 3k^3 + 3k^2 + 3k^3 + k^4)(3 + 3k + k^2)]$$

$$= r^6[(4 + 12k + 13k^2 + 6k^3 + k^4) - \frac{4}{27}(3 + 3k + k^2)(9 + 18k + 15k^2 + 6k^3 + k^4)]$$

$$= r^6[(4 + 12k + 13k^2 + 6k^3 + k^4) - \frac{4}{27}(27 + 54k + 45k^2$$
$$+ 18k^3 + 3k^4 + 27k + 54k^2 + 45k^3 + 18k^4 + 3k^5 + 9k^2 + 18k^3 + 15k^4 + 6k^5 + k^6)]$$

$$= r^6[(4 + 12k + 13k^2 + 6k^3 + k^4) - \frac{4}{27}(27 + 81k + 108k^2 + 81k^3 + 36k^4 + 9k^5 + k^6)]$$
$$= r^6[-3k^2 - 6k^3 - \frac{13}{3}k^4 - \frac{4}{3}k^5 - \frac{4}{27}k^6]$$

and examination of this last expression should convince us that the discriminant function for the cubic equation is always non-positive under the conditions given, for r was chosen to be positive, as was k, and since every term inside the brackets is negative, this function will assume its greatest value when $k = 0$. And upon reflection you will see that the exact same reasoning holds when equation (87) has two negative roots, for then we may choose r to be the negative root with the smaller absolute value, so that k and r^6 are still positive numbers. In other words, the only time Cardano's formula does not involve the cube root of a complex number is when at least two of the real roots of $P(y)$ are equal, or when $P(y)$ has just one real root

Well, dear reader, that's a pretty complete story on Cardano's formula for the general cubic equation. If you have questions about this little paper feel free to send me an e-mail message: davidbryant@att.net.