Parabola Volume 30, Issue 2 (1994)

HOW TO SOLVE CUBIC EQUATIONS BY FUNCTIONS ON THE CALCULATOR

David Tacon

Although there is a formula for solving cubic equations (see **Parabola**, 2nd issue 1989) it is rarely used in practice since it can be tricky to apply in even very simple cases. (Besides there are excellent numerical techniques which quickly find approximate solutions to any degree of accuracy.) Our purpose here is to describe another old technique, due to Francçis Viète (1540-1603), which gives exact solutions in terms of trig functions precisely when the cubic has three real roots. Thus, unfortunately, if the cubic has only one real root (and two complex conjugate roots) this method is unable to find it, a phenomenon which must have puzzled Viète greatly. Books which describe Viète's method rarely show how his technique can be modified to handle the other cases and this is what we wish to show here. To do so we need utilize two other functions on the calculator, one of which we "see" in our daily lives.

We begin with an observation which is even older than Viète: by a simple translation we can remove the "square term" from any cubic equation we wish to solve. Substitute $X = x + \frac{a}{3}$ or $x = X - \frac{a}{3}$ in the equation

$$x^3 + ax^2 + bx + c = 0; \qquad (1)$$

then

$$(X - \frac{a}{3})^3 + a(X - \frac{a}{3})^2 + b(X - \frac{a}{3}) + c = 0$$

which simplifies to

$$X^3 + (b - \frac{a^2}{3})X + (\frac{2a^3}{27} - \frac{ab}{3} + c) = 0. \qquad (2)$$

The point to appreciate is that any solution X to (2) gives rise to a solution $x = X - \frac{a}{3}$ to (1), and, of course, vice versa. It follows that we can solve all cubics if we can solve all cubic equations of the form

$$x^3 + ax + b = 0. \qquad (3)$$

Since the special cases where either $a = 0$ or $b = 0$ are easily solved we henceforth assume that $a \neq 0$ and $b \neq 0$. Following Viète our intention is to exploit the trig identity

$$\cos 3\theta = 4\cos^3\theta - 3\cos\theta.$$

This we rewrite as

$$\cos^3\theta - \frac{3}{4}\cos\theta - \frac{1}{4}\cos 3\theta = 0. \tag{4}$$

(To prove (4) start with the better known identity

$$\cos(A+B) = \cos A\cos B - \sin A\sin B$$

and remember that $\cos^2\theta + \sin^2\theta = 1$.)

We look for a solution to (3) of the form $x = r\cos\theta$ where $r > 0$. We want r and θ to satisfy

$$r^3\cos^3\theta + ar\cos\theta + b = 0$$

$$\text{or} \qquad \cos^3\theta + \frac{a}{r^2}\cos\theta + \frac{b}{r^3} = 0. \tag{5}$$

By comparison with (4) r and θ satisfy (5) if

$$\frac{a}{r^2} = -\frac{3}{4} \quad \text{and} \quad \frac{b}{r^3} = -\frac{1}{4}\cos 3\theta$$

$$\text{i.e., if} \qquad r^2 = -\frac{4}{3}a \qquad 6(i)$$

$$\text{and} \qquad \cos 3\theta = -\frac{4b}{r^3}. \qquad 6(ii)$$

Let's test the method by solving

$$y^3 - 3y^2 - y + 3 = 0$$

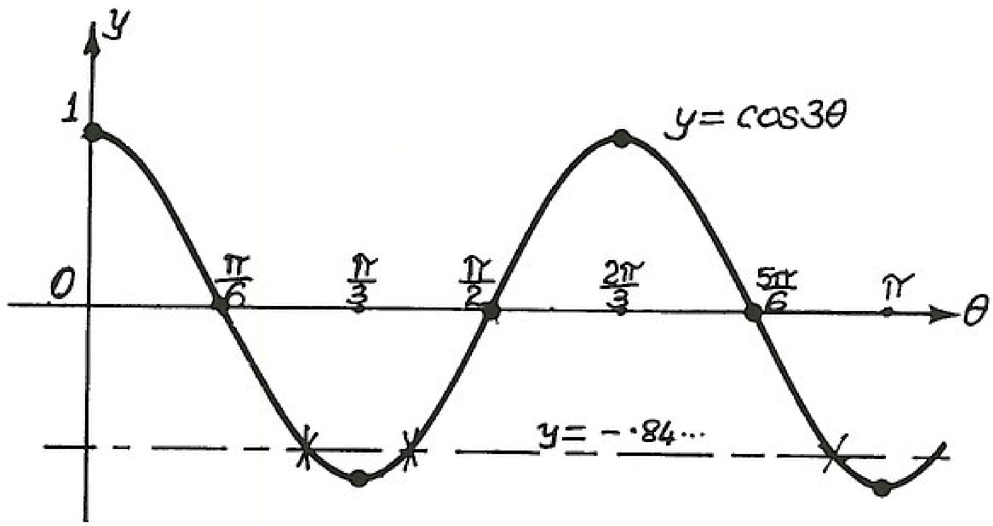
which we can see has solutions $y = -2, 2, 3$.

We make the substitution $y = x + 1$ so that the equation becomes

$$x^3 - 7x + 6 = 0$$

and we now want to "discover" the three roots $-3, 1, 2$. We know $x = r\cos\theta$ is a solution if $\qquad r^2 = -\frac{4}{3}.-7 = \frac{28}{3} \quad \text{and} \quad \cos 3\theta = -\frac{24}{4^3}$.

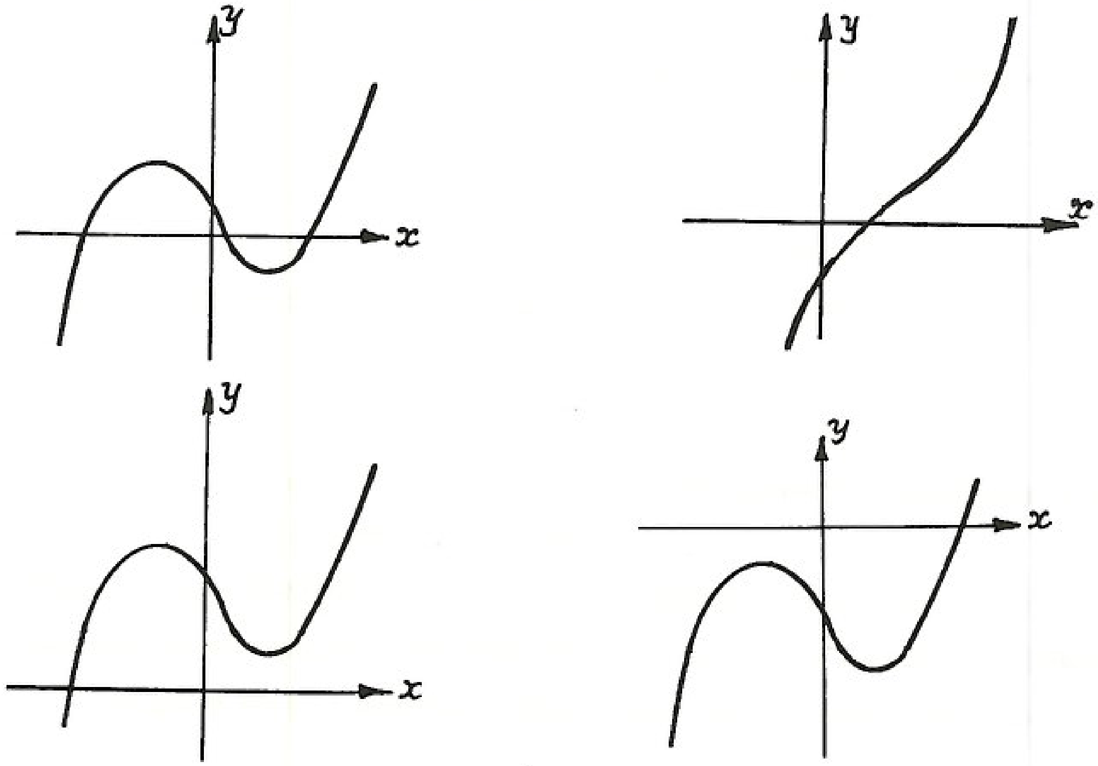
My calculator gives $r = 3.0550505$ so that $\cos 3\theta = -.8416975$ and $\theta = .8570719$ (or $49.106603°$). This gives $x = r\cos\theta = 2.0000001$. My calculator assumes its user knows that cosine is periodic. Let's graph $y = \cos 3\theta$ to see what is really going on.



From the graph it is clear that another two solutions to $\cos 3\theta = -.8416975$ which give rise to different values of $\cos\theta$ are $\theta = 1.2373232$ (or $70.893397°$) and $\theta = 2.951467$ or ($169.106603°$). These give rise to x values of .9999998 and -3. Clearly we have found our three roots.

This is all very well but we can solve 6(i) and 6(ii) (if and) only if $a < 0$ and $\left|\frac{4b}{r^3}\right| \leq 1$.

What does this tell us about the cubic $y = x^3 + ax + b$? Given the periodicity of cosine we might expect that our method is linked to the existence of multiple roots. Consider the following sketches of the graph of a cubic.



It is only in the 1st situation (I) that the function has three zeros. This is the case when the function has a minimum below the x axis and a maximum above the x axis. In the other cases, where the function has no turning points (II) or where the turning points are on the same side of the x axis, as in (III) or (IV), there is only one zero.

The turning points occur where $\frac{dy}{dx} = 3x^2 + a$ vanishes. If $a > 0$ then we have no turning point and the graph is as in (II). Otherwise turning points occur at $x = \pm\sqrt{-\frac{a}{3}}$ so that cases (III) and (IV) occur if

$$y\left(\sqrt{-\frac{a}{3}}\right) \cdot y\left(-\sqrt{-\frac{a}{3}}\right) < 0.$$

$$\text{i.e., if} \qquad 27b^2 + 4a^3 > 0.$$

So let $\triangle = b^2 + \frac{4}{27}a^3$. A moment's reflection shows that we have proved:

(i) the cubic has three real roots if $\triangle < 0$,

(ii) the cubic has one real root if $\triangle > 0$.

($-\triangle$ is called the "discriminant" of the cubic.)

Let's check that Viète's attack works if $\triangle < 0$. First observe that if $\triangle < 0$ then $a < 0$. This implies we can solve 6(i) for r, obtaining $r = \sqrt{-\frac{4}{3}a}$. Equation 6(ii) then has a solution since $|4b| \leq r^3$ as $27b^2 < -4a^3$. This means that if the cubic has three real solutions then we can find (all of) them by using our trig identity. Our argument simultaneously shows that if there is only one real solution then this approach cannot find it.

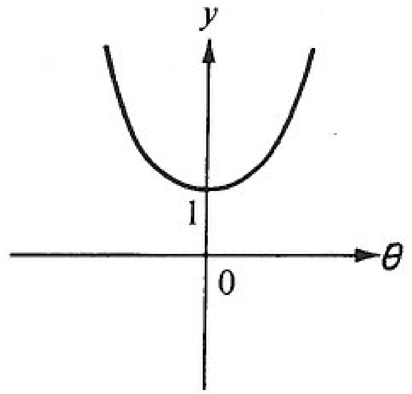
What to do? We introduce two new functions cosh and sinh (pronounced "shine") which are defined by

$$\cosh\theta = \frac{e^{\theta} + e^{-\theta}}{2}$$

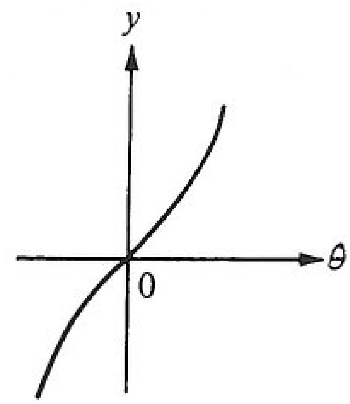
and

$$\sinh\theta = \frac{e^{\theta} - e^{-\theta}}{2} \quad \text{where } \theta \text{ is a real number.}$$

These are called "hyperbolic functions" (the others are tanh, cosech, sech and coth – no doubt you can imagine how they are defined), and they are evaluated on most calculators by hitting the "hyp" button before the appropriate "cos" or "sin" button. (If you are familiar with the complex variable formulae $\cos\theta = \dfrac{e^{i\theta}+e^{-i\theta}}{2}$ and $\sin\theta = \dfrac{e^{i\theta}-e^{-i\theta}}{2i}$ where $i=\sqrt{-1}$ and θ is a real number you will easily appreciate the notation.) These functions are not (real) periodic and their graphs are sketched below.



$y = \cosh\theta$



$y = \sinh\theta$

Notice that cosh is an even function (so that its graph is symmetric about the y axis) and that sinh is an odd function (so that its graph is symmetric about the origin). Note too that $\cosh\theta \geq 1$ for all θ whilst $\sinh\theta$ takes all real values. Fortunately for us the hyperbolic functions also satisfy cubic identities. On the one hand

$$\begin{aligned}\cosh^3\theta = \left(\frac{e^{\theta}+e^{-\theta}}{2}\right)^3 &= \frac{1}{8}(e^{3\theta}+3e^{\theta}+3e^{-\theta}+e^{-3\theta})\\ &= \frac{1}{4}\cosh 3\theta + \frac{3}{4}\cosh\theta\end{aligned}$$

so that

$$\cosh^3\theta - \tfrac{3}{4}\cosh\theta - \tfrac{1}{4}\cosh 3\theta = 0 \tag{7}$$

which has the same form as (4).

On the other hand

$$\begin{aligned}\sinh^3\theta = \left(\frac{e^{\theta}-e^{-\theta}}{2}\right)^3 &= \frac{1}{8}(e^{3\theta}-3e^{\theta}+3e^{-\theta}-e^{-3\theta})\\ &= \frac{1}{4}\sinh 3\theta - \frac{3}{4}\sinh\theta\end{aligned}$$

so that

$$\sinh^3\theta + \tfrac{3}{4}\sinh\theta - \tfrac{1}{4}\sinh 3\theta = 0. \tag{8}$$

When $a > 0$, which means the graph of cubic is of type (II), we look for a solution of the form $x = r\sinh\theta$. This we can always find since we can solve

$$r^2 = \frac{4}{3}a \quad \text{and} \quad \sinh 3\theta = -\frac{4b}{r^3}.$$

If $a < 0$ yet $\Delta > 0$ (as is the case if the graphs are of types (III) or (IV)) we substitute $x = r\cosh\theta$, where we no longer assume that $r > 0$, and then solve $r^2 = -\frac{4}{3}a$ and $\cosh 3\theta = -\frac{4b}{r^3}$. (If $b > 0$ we would do well to choose $r = -\sqrt{-\frac{4}{3}a}$. Can you see why?)

Here are a few equations to try.

$$\text{(a)} \quad x^3 - 6x + 4 = 0$$

$$\text{(b)} \quad x^3 - 2x - 4 = 0$$

$$\text{(c)} \quad x^3 - 2x + 4 = 0$$

$$\text{(d)} \quad x^3 + 2x - 3 = 0.$$

There are a number of interesting points which arose in our discussion which deserve further investigation.

Firstly did you notice that our proof of the identities (7) and (8) did not use any property of $e(= 2.718\ldots)$? For our purposes we could have just as easily have defined a new function "cosk" by

$$\text{cosk } \theta = \frac{10^\theta + 10^{-\theta}}{2} \quad \text{for real values of } \theta.$$

(This function is not very useful and you won't see the "cosk" function defined in any book!) We still have an identity:

$$\text{cosk }^3\theta - \frac{3}{4}\text{cosk } \theta - \frac{1}{4}\text{cosk } 3\theta = 0.$$

Now suppose we wish to solve equation (c) above using the cosk function. We look for a solution $x = r\text{cosk } \theta$ so we need r and θ to satisfy

$$\text{cosk }^3\theta - \frac{2}{r^2}\text{cosk } \theta + \frac{4}{r^3} = 0.$$

We want $\frac{2}{r^2} = \frac{3}{4}$ and $\frac{4}{r^3} = -\frac{1}{4}\text{cosk } 3\theta$ so we set $r = -2\sqrt{\frac{2}{3}}$ and $\text{cosk } 3\theta = \frac{3\sqrt{3}}{\sqrt{2}} = \sqrt{13.5}$. Here is a serious problem for there is no cosk $^{-1}$ button on the calculator. But we don't need it! Suppose we wish to solve the equation

$$\text{cosk } \ t = y$$

for t in terms of y (we need to assume $y \geq 1$).

We rewrite as $10^t + 10^{-t} = 2y$

which becomes $$(10^t)^2 - 2y.10^t + 1 = 0.$$

This is a quadratic in 10^t: it has solution

$$\begin{aligned} 10^t &= \frac{2y \pm \sqrt{4y^2-4}}{2} = y \pm \sqrt{y^2-1} \\ &= y + \sqrt{y^2-1} \text{ or } \frac{1}{y+\sqrt{y^2-1}} \\ \therefore t &= \pm \log_{10}(y + \sqrt{y^2-1}) \end{aligned}$$

and we define the inverse function by

$$t = \text{cosk}^{-1} y = \log_{10}(y + \sqrt{y^2-1})$$

Clearly this argument also proves that

$$\cosh^{-1} y = \ell n(y + \sqrt{y^2-1})$$

where $\ell n = \log_e$. There is no similar real valued formula for $\cos^{-1}$ or $\sin^{-1}$!

Now back to our original problem which was to solve cosk $3\theta = \sqrt{13.5}$. We have

$$3\theta = \log_{10}\left(\sqrt{13.5} + \sqrt{12.5}\right)$$

so that $\theta = .2859737\cdots$.

But then $$\text{cosk } \theta = \frac{1.9318\cdots + .5176\cdots}{2} = 1.2247449$$
so that $x = r\text{cosk } \theta = -1.63\cdots \times 1.22\cdots = -2$.

Do you realise why we find the cosh function on our calculators but not the similarly defined cosk function? The reason hinges on the fact that $y = e^x$ is the most important function in advanced mathematics. Why is it so important? Simply because $\frac{de^x}{dx} = e^x$; it is the only function which is not altered by differentiation. Consequently

$$\frac{d}{dx}\cosh x = \sinh x, \quad \frac{d^2}{dx^2}\cosh x = \cosh x \text{ and so forth.}$$

Finally it's apparent, after a little reflection, that we could have worked with the function

$$f(t) = \frac{1}{2}(t + \frac{1}{t})$$

rather than cosh t. This is because we could have exploited the identity

$$f(t^3) = 4(f(t))^3 - 3f(t)$$

by seeking solutions of the form $x = rf(t)$ for appropriate r and t.

Indeed we can exploit this identity in a slightly different way. Let's suppose we wish to solve the equation

$$t^6 + at^5 + bt^4 + ct^3 + bt^2 + at + 1 = 0. \tag{9}$$

It can be rewritten as

$$(t^3 + \frac{1}{t^3}) + a(t^2 + \frac{1}{t^2}) + b(t + \frac{1}{t}) + c = 0$$

i.e.,

$$((t + \frac{1}{t})^3 - 3(t + \frac{1}{t})) + a((t + \frac{1}{t})^2 - 2) + b(t + \frac{1}{t}) + c = 0$$

or

$$x^3 + ax^2 + (b - 3)x + c - 2a = 0 \tag{10}$$

after substituting $x = t + \frac{1}{t}$. In other words we have reduced an equation of degree 6 (admittedly a rather special equation) to one of degree 3. Clearly we can thus solve (9) piecemeal: first solving (10) for x and then finding t from $x = t + \frac{1}{t}$. If you can muster the energy you might find the real solutions of

$$t^6 + t^5 - 9t^4 + 2t^3 - 9t^2 + t + 1 = 0$$

And second last, what of our comment that we see one of the hyperbolic functions in our everyday lives. The curve of suspension of a flexible rope or wire or chain (catena) is called a **catenary**. If you are aware that the trajectory of a projectile is a parabola you probably think that a catenary is a parabola. Even Galileo thought this but he was wrong too for a parabola only approximates the catenary. The catenary is actually a cosh curve.