A new approach to solving the cubic: Cardan's solution revealed[1]

R. W. D. Nickalls

The cubic holds a double fascination, since not only is it interesting in its own right, but its solution is also the key to solving quartics. This article describes five fundamental parameters of the cubic (δ, λ, h, x_N and y_N), and shows how they lead to a significant modification of the standard method of solving the cubic, generally known as *Cardan's solution.*

It is necessary to start with a definition. Let $\mathrm{N}(x_N, y_N)$ be a point on a polynomial curve $f(x)$ of degree n such that moving the x-axis by putting $z = x - x_N$ makes the sum of the roots of the new polynomial $f(z)$ equal to zero. It is easy to show that for the polynomial equation

$$ax^n + bx^{n-1} + ..k = 0,$$

$x_N = -b/(na)$. If $f(x)$ is a cubic polynomial then $f(z)$ is known as the reduced cubic, and N is the point of inflexion.

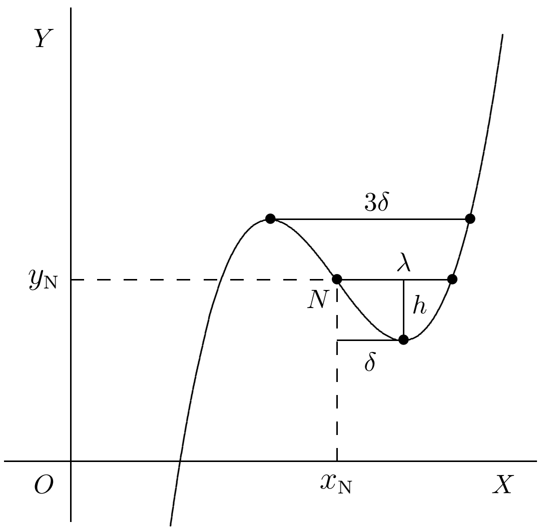

Figure 1:

Now consider the general cubic

$$y = ax^3 + bx^2 + cx + d.$$

Here x_N is $-b/(3a)$, and N the point of symmetry of the cubic. Let the parameters δ, λ, h, be defined as the distances shown in figure 1. It can be shown, and readers will easily do this, that λ and h are simple functions

[1]This revision of the original article incorporates some minor corrections, footnotes, and an improvement to figure 2.

of δ namely

$$\lambda^2 = 3\delta^2 \quad \text{and} \quad h = 2a\delta^3,$$

where

$$\delta^2 = \frac{b^2 - 3ac}{9a^2}$$

This result is found easily by locating the turning points. Thus the shape of the cubic is completely characterised by the parameter δ. Either the maxima and minima are distinct ($\delta^2 > 0$), or they coincide at N ($\delta^2 = 0$), or there are no turning points ($\delta^2 < 0$). Furthermore, the quantity $a\delta\lambda^2/h$ is constant for any cubic, as follows

$$\frac{a\delta\lambda^2}{h} = \frac{3}{2}$$

The relationship $\lambda^2 = 3\delta^2$ is a particular case of the general observation that

> If a polynomial curve passes through the origin, then the product of the roots $x_1, x_2, \ldots, x_{n-1}$ (excluding the solution $x = 0$) is related to the product of the x-coordinates of the turning points $t_1 t_2 \ldots t_{n-1}$ by
>
> $$x_1 x_2 \ldots x_{n-1} = n t_1 t_2 \ldots t_{n-1},$$

a result whose proof readers can profitably set to their classes, and which parallels a related but much more difficult result about the y-coordinates of the turning points which we have discovered[2].

Solution of the cubic

In addition to their value in curve tracing, I have found that the parameters δ, h, x_N and y_N, greatly clarify the standard method for solving the cubic since, unlike the Cardan approach [1] they reveal how the solution is related to the geometry of the cubic.

For example the Cardan solution, using the standard terminology, involves starting with an equation of the form

$$ax^3 + 3b_1x^2 + 3c_1x + d = 0$$

and then substituting $z = x + (b_1/a)$ to generate a reduced equation of the form

$$z^3 + \frac{3H}{a^2}z + \frac{G}{a^3} = 0$$

where

$$H = ac_1 - b_1^2 \quad \text{and} \quad G = a^2d - 3ab_1c_1 + 2b_1^3.$$

This obscures the fact that the reduced form of the cubic has the point N on the y-axis. Subsequent development yields a discriminant of the form $G^2 + 4H^3$, where

[2]Nickalls RWD and Dye R (1996). The geometry of the discriminant of a polynomial. *The Mathematical Gazette*, 80 (July), 279–285.

$$G^2 + 4H^3 = a^2(a^2d^2 - 6ab_1c_1d + 4ac_1^3 + 4b_1^3d - 3b_1^2c_1^2).$$

The problem is that it is not clear geometrically what the quantities G and H represent. However by using the parameters described earlier, not only is the solution just as simple, but the geometry is revealed.

Start with the usual form of the cubic equation

$$f(x) = ax^3 + bx^2 + cx + d = 0$$

having roots α, β, γ, and obtain the reduced form by the substitution $z = x - x_N$ (see figure 1). The equation will now have the form

$$az^3 - 3a\delta^2 z + y_N = 0$$

and have roots $\alpha - x_N, \beta - x_N, \gamma - x_N$; a form which allows the use of the usual identity

$$(p+q)^3 - 3pq(p+q) - (p^3 + q^3) = 0.$$

Thus $z = p + q$ is a solution where

$$pq = \delta^2 \quad \text{and} \quad p^3 + q^3 = -y_N/a.$$

Solving these equations as usual by cubing the first, substituting for q in the second, and solving the resulting quadratic in p^3 gives

$$p^3 = \frac{1}{2a}\left\{-y_N \pm \sqrt{y_N^2 - 4a^2\delta^6}\right\}$$

and since $h = 2a\delta^3$, this becomes

$$p^3 = \frac{1}{2a}\left\{-y_N \pm \sqrt{y_N^2 - h^2}\right\} \tag{1}$$

When this solution is viewed in the light of figure 1, it is immediately clear that equation 1 is *only* useful when there is a single real root, that is when

$${y_N}^2 > h^2.$$

Contrast this with the standard Cardan approach which gives

$$p^3 = \frac{1}{2a^3}\left\{-G \pm \sqrt{G^2 + 4H^3}\right\}$$

which completely obscures this fact. The values of G, H, and $G^2 + 4H^3$ are therefore found to be

$$G = a^2 y_N, \qquad H = -a^2\delta^2 \quad \text{and} \quad G^2 + 4H^3 = a^4(y_N^2 - h^2).$$

However, since the sign of h depends on that of δ, letting $h = h_1 = -h_2$ allows equation 1 to be rewritten as

$$p^3 = \frac{1}{2a}\left\{-y_N \pm \sqrt{(y_N + h_1)(y_N + h_2)}\right\}.$$

If the y-coordinate of a turning point is y_t then let

$$y_N + h_1 = y_{t_1} \quad \text{and} \quad y_N + h_2 = y_{t_2}.$$

Our solution can therefore be written as

$$p^3 = \frac{1}{2a}\left\{-y_N \pm \sqrt{(y_{t_1})(y_{t_2})}\right\}.$$

Using the symbol Δ_3 for the discriminant[3] of the cubic, we have

$$\Delta_3 = (y_{t_1})(y_{t_2}) = y_N^2 - h^2$$

Returning to the geometrical viewpoint, figure 1 shows that the rest of the solution depends on the sign of the discriminant[4] as follows:

$$y_N^2 > h^2 \qquad \text{1 real root}$$

$$y_N^2 = h^2 \qquad \text{3 real roots (two or three equal roots)}$$

$$y_N^2 < h^2 \qquad \text{3 distinct real roots.}$$

These are now dealt with in order.

(i) $y_N^2 > h^2$ i.e. $(y_{t_1})(y_{t_2}) > 0$, or Cardan's $G^2 + 4H^3 > 0$

Clearly, there can only be 1 real root under these circumstances (see figure 1). As the discriminant is positive the value of the real root α is easily obtained as

$$\alpha = x_N + \sqrt[3]{\frac{1}{2a}\left(-y_N + \sqrt{y_N^2 - h^2}\right)} + \sqrt[3]{\frac{1}{2a}\left(-y_N - \sqrt{y_N^2 - h^2}\right)}$$

(ii) $y_N^2 = h^2$ i.e. $(y_{t_1})(y_{t_2}) = 0$, or Cardan's $G^2 + 4H^3 = 0$

Providing $h \neq 0$ this condition yields two equal roots, the roots being $z = \delta$, δ and -2δ. The true roots are then $x = x_N + \delta$, $x_N + \delta$ and $x_N - 2\delta$. Since there are two double root conditions the sign of δ is critical, and depends on the sign of y_N, and so in these circumstances δ has to be determined from

$$\delta = \sqrt[3]{\frac{y_N}{2a}}.$$

If $y_N = h = 0$ then $\delta = 0$, in which case there are three equal roots at $x = x_N$.

(iii) $y_N^2 < h^2$ i.e. $(y_{t_1})(y_{t_2}) < 0$, or Cardan's $G^2 + 4H^3 < 0$

From figure 1 it is clear that there are three distinct real roots in this case. However, our solution requires that we find the cube root of a complex number, so it is easier to use trigonometry to solve the reduced

[3] Note that the product $y_{t_1}y_{t_2}$ of the y-coordinates of the turning points is the geometric discriminant of the cubic (see article in footnote 2).

[4] Since the sign reflects the relative magnitude of y_N^2 and h^2.

form using the substitution $z = 2\delta \cos\theta$. This gives

$$2a\delta^3 \left(4\cos^3\theta - 3\cos\theta\right) + y_N = 0,$$

and since $2a\delta^3 = h$, this becomes

$$\cos 3\theta = \frac{-y_N}{h}. \tag{2}$$

The three roots α, β and γ are therefore given by

$$\begin{aligned}
\alpha &= x_N + 2\delta\cos\theta \\
\beta &= x_N + 2\delta\cos(2\pi/3 + \theta) \\
\gamma &= x_N + 2\delta\cos(4\pi/3 + \theta).
\end{aligned}$$

These are shown in figure 2 in relation to a circle, radius 2δ, centered above N. Note that the maximum between roots β and γ corresponds to the angle $2\pi/3$.

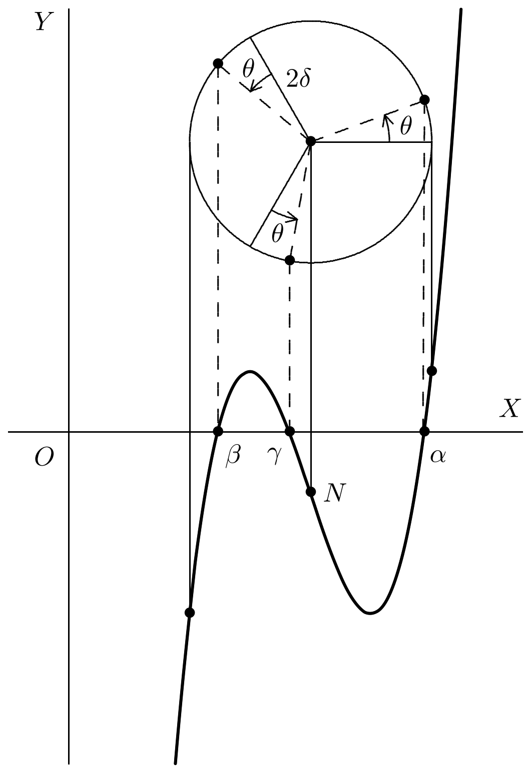

Figure 2:

It is clear from equation 2 that trigonometry can only be used to solve the reduced cubic when

$$-1 \leq \frac{y_N}{h} \leq +1$$

a point which is completely obscured by the corresponding Cardan equation

$$\cos 3\theta = \frac{-G}{2\,(-H)^{\frac{3}{2}}}.$$

Example

Solve the equation

$$x^3 - 7x^2 + 14x - 8 = 0$$

The parameters are

$$x_N = 7/3, \quad y_N = f(x_N) = -0.7407, \quad \delta^2 = 7/9 \quad \text{and} \quad h = 1.3718.$$

Since $y_N^2 < h^2$, it follows (see figure 1) that there are three distinct real roots, which are given by

$$x = x_N + 2\delta\cos\theta$$

where

$$\cos 3\theta = \frac{-y_N}{h} = \frac{0.7407}{1.3718} = 0.5399$$

So $\theta = 19.1066°$, and the three roots are

$$\begin{aligned}
\alpha &= \tfrac{7}{3} + 2\sqrt{\tfrac{7}{9}}\cos 19.1066° &= 4\\
\beta &= \tfrac{7}{3} + 2\sqrt{\tfrac{7}{9}}\cos 139.1066° &= 1\\
\gamma &= \tfrac{7}{3} + 2\sqrt{\tfrac{7}{9}}\cos 259.1066° &= 2
\end{aligned}$$

In summary, I would like to suggest that the usual Cardan-type terminology for cubics and quartics, though it has been used for hundreds of years, be abandoned in favour of the parameters δ, h, x_N, y_N which reveal to such advantage how the algebraic solution is related to the geometry of the cubic.

Reference

1. Burnside W. S. and Panton A. W. The theory of equations: with an introduction to the theory of binary algebraic forms. 2nd ed. Longmans, Green and Co., London (1886).

RWD Nickalls
Department of Anaesthesia, City Hospital, Nottingham, UK.
email: *dicknickalls@compuserve.com*