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

RWD Nickalls [2]

The Mathematical Gazette (1993); 77 (Nov, No 480), 354–359 (JSTOR)
http://www.nickalls.org/dick/papers/maths/cubic1993.pdf

1 Introduction

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 [3]. 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.*

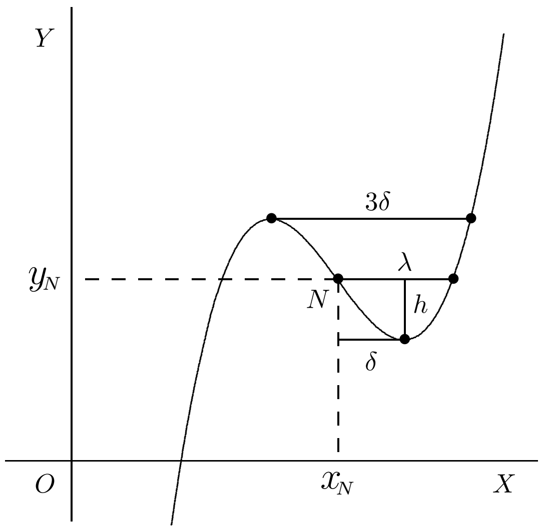

Figure 1:

[1] This minor revision of the original article corrects typographic errors and incorporates some explanatory footnotes and a minor improvement to Figure 2. The original published version is available from http://www.jstor.org/stable/3619777.

[2] Department of Anaesthesia, Nottingham University Hospitals, City Hospital Campus, Nottingham, UK. email: dick@nickalls.org

[3] Nickalls RWD (2009). The quartic equation: invariants and Euler's solution revealed. *The Mathematical Gazette*; 93 (March), 66-75. (http://www.nickalls.org/dick/papers/maths/quartic2009.pdf)

It is necessary to start with a definition. Let $N(x_N, y_N)$ be a point on a polynomial curve $f(x)$ of degree n such that moving the axes by putting $x = z + x_N$ 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

$$ax^n + bx^{n-1} + \ldots + k = 0$$

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

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 of δ namely [4, 5]

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

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 real 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 $x_1, x_2, \cdots, x_{n-1}$ (excluding the solution $x = 0$) is related to the product of the x-coordinates of the turning points $t_1 t_2 \cdots t_{n-1}$ by
>
> $$x_1 x_2 \cdots x_{n-1} = n t_1 t_2 \cdots t_{n-1},$$

a result whose proof readers can profitably set to their classes, and which parallels a related but more difficult result about the y-coordinates of the turning points which we have discovered [6].

[4] Unfortunately in the original printed version h was presented with a negative sign.

[5] Note that if m_N denotes the slope at the point of inflection N, then $m_N = -3a\delta^2$. See also Thomas Müller's interactive demonstration of how the parameters x_N, y_N, h, δ, m_N influence the shape and location of the cubic at http://demonstrations.wolfram.com/ParametersForPlottingACubicPolynomial/

[6] Nickalls RWD and Dye RH (1996). The geometry of the discriminant of a polynomial. *The Mathematical Gazette*, 80 (July), 279–285 (JSTOR). (http://www.nickalls.org/dick/papers/maths/discriminant1996.pdf)

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 (Burnside and Panton, 1886) [7], they reveal how the solution is related to the geometry of the cubic.

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

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

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

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

where

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

Subsequent development yields a discriminant of the form $G^2 + 4H^3$ where

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

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.

2.1 New approach

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 $x = z + x_N$ (see Figure 1). The equation will now have the form [8]

$$az^3 - 3a\delta^2z + y_N = 0, \tag{1}$$

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

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

[7] §§ 56–57 (pages 106–109).
[8] $y_N \equiv f(x_N) \equiv 2b^3/(27a^2) - bc/(3a) + d$

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 $h^2 = 4a^2\delta^6$, this becomes[9]

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

When this solution is viewed in the light of Figure 1, it is immediately clear that Equation 2 is particularly 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

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

However, Equation 2 can be rewritten as

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

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

$$y_N + h = y_{T_1} \quad \text{and} \quad y_N - h = y_{T_2}.$$

Our solution (Equation 2) 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 [geometric] discriminant[10, 11] of the cubic, we have

$$\Delta_3 = y_{T_1} y_{T_2} = y_N^2 - h^2.$$

[9]Providing $h \neq 0$ then this is equivalent to (see also footnote 13)

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

[10]Note that the product $y_{T_1} y_{T_2}$ of the y-coordinates of the turning points is known as the *geometric* discriminant Δ of the cubic; it is the geometric analogue of the algebraic discriminant (see Nickalls and Dye (1996)—for URL see footnote 6). The classical discriminant $G^2 + 4H^3$ has the same sign as the geometric discriminant since $G^2 + 4H^3 = a^4(y_N^2 - h^2) = a^4 y_{T_1} y_{T_2}$.

[11]The *algebraic* discriminant D of the cubic is defined as the product of squared differences of the roots $D = (\alpha - \beta)^2(\alpha - \gamma)^2(\beta - \gamma)^2$ and hence $a^2 D = -27(y_N^2 - h^2) = -27 y_{T_1} y_{T_2}$.

Returning to the geometrical viewpoint, Figure 1 shows that the rest of the solution depends on the sign of the discriminant [12] as follows:

$y_N^2 > h^2$	1 real root,
$y_N^2 = h^2$	3 real roots (two or three equal roots),
$y_N^2 < h^2$	3 distinct real roots.

These are now dealt with in order.

2.2 $y_N^2 > h^2$ 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 [13, 14]

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

2.3 $y_N^2 = h^2$ 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 $x = x_N$.

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

[13] The remaining two complex roots are given by

$$\beta, \gamma = x_N - \frac{(\alpha - x_N)}{2} \pm j\frac{\sqrt{3}}{2}\sqrt{(\alpha - x_N)^2 - 4\delta^2}$$

where $j^2 = -1$ (for derivation see: Nickalls RWD (2009). Feedback: 93.35: *The Mathematical Gazette*; 93 (Mar), 154–156. http://www.nickalls.org/dick/papers/maths/cubictables2009.pdf).

[14] Note that multiplying top and bottom by h ($h \neq 0$) gives

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

and since $h = 2a\delta^3$ then $h/2a = \delta^3$, and hence

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

which highlights the significance of the ratio y_N/h. Thus when $y_N/h = -1$ we obtain $\alpha = x_N + 2\delta$.

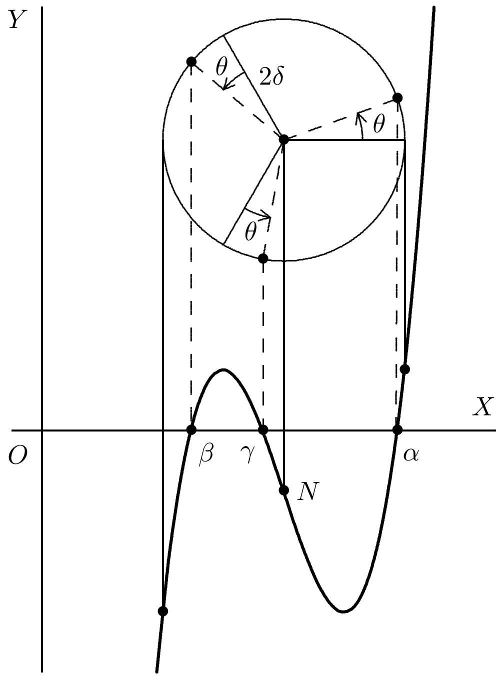

Figure 2:

2.4 $y_N^2 < h^2$ i.e. $y_{T_1} y_{T_2} < 0$, or Cardan's $G^2 + 4H^3 < 0$

From Figures 1 and 2 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 form using the substitution[15] $z = 2\delta\cos\theta$ in Equation 1. 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

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

[15]Note that if we use instead the substitution $z = 2\delta\sin\phi$, this leads to $\sin 3\phi = y_N/h$, for which the condition $y_N = 0$ is associated with $\phi = 0$. Use of this form is described in Nickalls, RWD (2006). Viète, Descartes and the cubic equation. *The Mathematical Gazette*; 90 (July), 203–208. (http://www.nickalls.org/dick/papers/maths/descartes2006.pdf).

The three roots α, β and γ are therefore given by [16, 17]

$$\left\{ \begin{array}{l} \alpha = x_N + 2\delta\cos\theta, \\ \beta = x_N + 2\delta\cos(2\pi/3 + \theta), \\ \gamma = x_N + 2\delta\cos(4\pi/3 + \theta). \end{array} \right. \tag{4}$$

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

[16] Assuming $\delta \neq 0$, since otherwise $h = 0$ and hence $\cos 3\theta = -y_N/0$ etc. In practice, this situation does not arise since if $\delta = 0$ then the reduced cubic (equation 1) reduces to the trivial form $ax^3 + y_N = 0$ which is easily solved.

[17] Extending footnote 13 to embrace the three real-root condition $-1 \leq y_N/h \leq +1$ (see Figures 1 and 2) we can write

$$\alpha = x_N + \delta\sqrt[3]{\frac{-y_N}{h} - j\sqrt{1 - \frac{y_N^2}{h^2}}} \; + \; \delta\sqrt[3]{\frac{-y_N}{h} + j\sqrt{1 - \frac{y_N^2}{h^2}}},$$

and so the substitution $-y_N/h = \cos 3\theta$ then gives

$$\alpha = x_N + \delta\sqrt[3]{\cos 3\theta - j\sin 3\theta} \; + \; \delta\sqrt[3]{\cos 3\theta + j\sin 3\theta},$$

and hence

$$\alpha = x_N + \delta(\cos\theta - j\sin\theta) \; + \; \delta(\cos\theta + j\sin\theta),$$

i.e.

$$\alpha = x_N + 2\delta\cos\theta.$$

Note that it is sometimes convenient to express the three roots in the following succinct form: $\alpha,\ \beta,\ \gamma = x_N + 2\delta\cos(2k\pi/3 + \theta)\,, \quad (k = 0, 1, 2)$.

It is clear from Equation 3 that trigonometry [using real θ][18] can only be used to solve the reduced cubic when

$$-1 \le \frac{y_N}{h} \le +1,$$

a point which is completely obscured by the corresponding classical form

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

[18]However, if we extend the range of θ to include complex angles ($\theta \in \mathbb{C}$), then *all* cubics can be solved using this approach. Consider the cubic $x^3 - 2x + 4 = 0$ (roots: -2, $1 \pm j$) for which $x_N = 0$, $\delta^2 = 2/3$, $h = 2a\delta^3 \approx 1{\cdot}08$, and $y_N/h \approx 3{\cdot}67$. Using (4) the roots are therefore given by

$$r_k = 2\delta\cos(\theta + \frac{2k\pi}{3}), \qquad (k = 0, 1, 2)$$

where $\cos 3\theta = -y_N/h \approx -3{\cdot}67$. Working in degrees for convenience, it follows that $3\theta = (180^\circ + j113{\cdot}18^\circ)$. Thus $\theta = (60^\circ + j37{\cdot}7^\circ)$, and hence the three roots r_k are given by

$$r_k = 2\delta\cos(60^\circ + k(120^\circ) + j37{\cdot}7^\circ), \qquad (k = 0, 1, 2).$$

For example, $k = 1$ generates the single real root r_1, as follows:

$$r_1 = 2\delta\cos(180^\circ + j37{\cdot}7^\circ).$$

Using the identity $\cos(A + B) = \cos A\cos B - \sin A\sin B$ this becomes

$$r_1 = 2\delta\{\cos(180^\circ)\cos(j37{\cdot}7^\circ) - \sin(180^\circ)\sin(j37{\cdot}7^\circ)\},$$

which reduces to

$$r_1 = -2\delta\cos(j37{\cdot}7^\circ).$$

Using the power series $\cos x \equiv 1 - x^2/2! + x^4/4! + \ldots$, and letting $x = 37{\cdot}7\pi/180$ (since x is in radians) we obtain (since $j^2 = -1$)

$$r_1 = -2\delta(1 + x^2/2! + x^4/4! + \ldots) \equiv -2\delta\cosh(x) = -2\delta(1{\cdot}2247),$$

and since $\delta^2 = 2/3$, this gives

$$r_1 = -2\sqrt{2/3}\,(1{\cdot}2247) = -2,$$

as required.

3 Example

Solve the equation [19]

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

The parameters are

$$x_N = 7/3, \quad y_N = f(x_N) \approx -0{\cdot}7407, \quad \delta^2 = 7/9 \quad \text{and} \quad h = 2a\delta^3 \approx 1{\cdot}3718.$$

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

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

where

$$\cos 3\theta = \frac{-y_N}{h} \approx \frac{0{\cdot}7407}{1{\cdot}3718} \approx 0{\cdot}5399.$$

So $\theta \approx 19{\cdot}1066^\circ$, and the three roots are

$$\begin{cases} \alpha = \dfrac{7}{3} + 2\sqrt{\dfrac{7}{9}}\cos 19{\cdot}1066^\circ \; = 4, \\ \beta = \dfrac{7}{3} + 2\sqrt{\dfrac{7}{9}}\cos 139{\cdot}1066^\circ = 1, \\ \gamma = \dfrac{7}{3} + 2\sqrt{\dfrac{7}{9}}\cos 259{\cdot}1066^\circ = 2. \end{cases}$$

4 Conclusion

Finally, 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.

5 References

- Burnside W. S. and Panton A. W. (1886) [20]. The theory of equations: with an introduction to the theory of binary algebraic forms. 2nd ed. Longmans, Green and Co., London [reprinted: Dover, 1960; 2 vols (7th ed., 1928)].

◇ ◇ ◇

---

[19] For another example see: Nickalls RWD (1996). A note on solving cubics. *The Mathematical Gazette*; 80 (November), 576–577 (JSTOR). (http://www.nickalls.org/dick/papers/maths/cubefink.pdf)

[20] The 1st ed (1881) is available on the web at http://quod.lib.umich.edu/cgi/t/text/text-idx?c=umhistmath;idno=ACA7397
See §§ 56–57 (pp. 104–107), and the excellent historical notes at back of the book, especially Note A *Algebraic solution of equations*, pp. 433–436, and Note B *Solution of numerical equations*, pp. 437–440.