The Solution of the General Cubic Equation: A Personal Journey

For years I had known (without remembering many details) that the general cubic and quartic equations were solved in the Renaissance. But I had never learned or used the formulas for the solutions. In fact, it is generally agreed that numerical, graphical, or asymptotic (power series) methods for finding approximate solutions are much more useful in practice than those exact formulas. It was not until 2001 that I had an actual need for a formula for a solution of a cubic equation, a formula that depends on a number of variables or parameters and can be differentiated with respect to them. (If you want to skip the physics and go directly to the cubic equation, it is $(*)$.)

I became interested in the quantum-mechanical problem of a charged particle moving under a constant electric field. I will say nothing more about quantum mechanics except to say that to approximate the quantum solution one wants to know the solution of the corresponding *two-point boundary-value* classical problem: If the particle starts at position x at time 0 and arrives at position y at time t, what is its trajectory? (For simplicity we assume that the motion takes place entirely in one dimension.)

Classical mechanics problems are more often formulated as *initial-value* problems: If the particle starts at time 0 at position x and velocity v, where does it go? In elementary physics classes the constant-force problem is most often studied for a gravitational field instead of an electric one; this is the *falling body* problem, whose solution is

$$y(t) = x + vt + \tfrac{1}{2}gt^2$$

if there is no interference with the motion. (Notation is chosen so that the positive direction is downward.) I was interested in the case where the "ball" hits a "ceiling" at $y = 0$. For example, choose units so that $g = 1$ and consider the initial data $x = 8$, $v = -5$. The ball rises (y decreases) until $t = 2$, at which $y = 0$ and $y' = -3$. Then the ball bounces back, which means that its velocity suddenly changes sign. After that it falls freely again, so we need to solve the problem with initial time $t = 2$, initial position $x = 0$, and initial velocity $v = +3$: $y(t) = 3(t-2) + \frac{1}{2}(t-2)^2$. In all, then,

$$y(t) = \begin{cases} 8 - 5t + \frac{1}{2}t^2 & \text{for } 0 \le t \le 1, \\ -4 + t + \frac{1}{2}t^2 & \text{for } t \ge 1. \end{cases}$$

On the other hand, if v is positive, or negative but too small, then the ball will never hit the ceiling and the trajectory is just the usual single parabola.

Now turn to the boundary-value problem. I need to find all trajectories with $y(0) = x$ and $y(t) = y$, for given positive numbers x, y, and t. To avoid ambiguity we need to complicate the notation: Write τ and q for the time and space coordinates along the trajectory (in contrast to t and y, which are fixed). Then $q(\tau; t, x, y)$ is supposed to be a solution of

$$\frac{d^2q}{d\tau^2} = g = 1$$

satisfying

$$q(0;t,x,y) = x, \quad q(t;t,x,y) = y.$$

If there is no ceiling, the solution always exists and is unique:

$$q(\tau;t,x,y) = x + \left(\frac{y-x}{t} - \frac{t}{2}\right)\tau + \frac{1}{2}\tau^2.$$

If there is a ceiling, for certain values of (t,x,y) this solution is still valid because q is never negative. However, you can see that if t is too big (for a given x and y), there will be a problem: If we toss the ball up too softly, so that it never hits the ceiling, then it will have fallen to a depth greater than y long before time t. On the other hand, if we throw the ball up really hard, then it will bounce off the ceiling right away and come back fast, and again it will be gone long before t. It turns out, in fact, that if $t > \sqrt{2x} + \sqrt{2y}$, then *there is no solution* to the two-point boundary-value problem, whereas if $t < \sqrt{2x} + \sqrt{2y}$, there are *two* solutions, one with a bounce and one without! When $t = \sqrt{2x} + \sqrt{2y}$ there is only one solution, which just grazes the ceiling (at time $\sqrt{2x}$):

$$q(\tau;t,x,y) = x - \sqrt{2x}\,\tau + \tfrac{1}{2}\tau^2 = \tfrac{1}{2}(\tau - \sqrt{2x})^2.$$

Our primary interest here is in the construction of a solution with a bounce. Suppose that the bounce occurs at $\tau = s$. Then for $0 \le \tau \le s$ the solution must have the form

$$q = q_1(\tau;t,x,y) \equiv q(\tau;s,x,0) = x + \left(\frac{0-x}{s} - \frac{s}{2}\right)\tau + \frac{1}{2}\tau^2$$

(because the ball moves from x to 0 in time s, and we can use the bounceless q formula over this interval), and for $t^2 \le \tau \le t$ it must have the form

$$q = q_2(\tau;t,x,y) \equiv q(\tau - s; t - s, 0, y) = 0 + \left(\frac{y-0}{t-s} - \frac{t-s}{2}\right)(\tau - s) + \frac{1}{2}(\tau - s)^2$$

(because the ball moves from 0 to y in time $t - s$, with the clock starting at $\tau = s$). Also, the velocities at time s must be equal and opposite, since that is part of the definition of a bounce (elastic collision):

$$0 = q_1' + q_2' = \left(-\frac{x}{s} + \frac{s}{2}\right) + \left(\frac{y}{t-s} - \frac{t-s}{2}\right).$$

This condition simplifies to

$$s^3 - \frac{3t}{2}\,s^2 + \left(\frac{t^2}{2} - x - y\right)s + tx = 0. \qquad (*)$$

I need to know s as a function of (t,x,y) to complete the solution.

Formulas for the solution of the cubic can be found in handbooks, such as the National Bureau of Standards *Handbook of Mathematical Functions*. Unlike Cardano (Allen p. 17), the handbook indicates unambiguously how to obtain all three complex roots. However, that's far from the end of the story.

Recall that a cubic with real coefficients can have either one real root and two complex conjugate roots (example: $x^3 + x = 0$) or three real roots (example: $x^3 - x = 0$). (This is obvious after sketching a few graphs.) There are also special cases where two real roots coincide. Which case occurs depends on the sign of the quantity inside the square root (which in turn is inside the cube roots) in the Tartaglia–Cardano formula. When it is positive, there is only one real root and the formula gives it without incident; Tartaglia et al. were blissfully unaware of the two complex roots, just as they blew off $x^2 + 1 = 0$ as "a quadratic equation with no solutions". But if the radicand is negative, the formula gives real roots as differences of two *complex* numbers whose imaginary parts happen to cancel, and considerable formalistic handwaving was needed to extract the three real roots, even when they were known beforehand (see Bombelli, Allen p. 18).

It turns out that for positive values of t, x, and y, equation $(*)$ always falls into the class with three real roots, exactly one of which satisfies the physically necessary condition $0 < s < t$. Although the complex numbers in the formula are no longer a conceptual problem for us, they are still a practical nuisance. Asking *Maple* to plot the solution as a function of one of the parameters with the other two fixed leads to disaster, for instance. *Maple* calculates the difference of the two cube roots numerically, and often the imaginary part does not turn out to be zero, because of roundoff error. The plotting routine in *Maple* doesn't know what to do with imaginary numbers, so it just leaves gaps in the graph. Furthermore, dealing with the nested radicals analytically is highly unpleasant algebra.

It turns out that for the case with three real roots there is a better way of treating the cubic, which is attributed to Viète (sometime in the late 1500s) but apparently not published until 1629, by Girard. In other words, it took over 50 years after Cardano for this next step. (The most complete discussion I have found of this part of the history is in Boyer's book, and he is vague about details and dates. Boyer says, "In many respects the work of Viète is greatly undervalued ... There is no generally accessible edition of [his] works, nor even a good general account in English of his life and work.") Viète's greatest contribution was modern algebraic notation (writing letters like x instead of words like "cosa"), but his second greatest was a heap of trigonometric identities, including the third-degree one relevant to the trisection problem and to the cubic equation (see Allen p. 19; this is still in "Renaissance" although Viète's biography is in "Transition").

Boyer (p. 341) describes Viète's trick thus: The general cubic can be put into the form

$$y^3 + 3m^2py + m^3q = 0,$$

where m can be chosen arbitrarily. Viète's trig identity is

$$\cos^3\theta - \tfrac{3}{4}\cos\theta - \tfrac{1}{4}\cos(3\theta) = 0.$$

So, substitute $y = \cos\theta$ and get

$$3m^2p = -\tfrac{3}{4}, \quad -\tfrac{1}{4}\cos(3\theta) = m^3q.$$

Solve the first equation for m, then the second equation for θ (using a table or approximate algorithm for inverse cosines; there is no free lunch). Computing the cosine, we get y. (The triple nature of the root is related to the multivaluedness of the inverse cosine.)

The long and short of it is that the most useful formula for the physically relevant solution of $(*)$ is

$$s = \frac{t}{2} + \frac{1}{\sqrt{3}}\sqrt{t^2 + 4x + 4y}\,\sin\left[\frac{1}{3}\sin^{-1}\frac{6\sqrt{3}\,t(x-y)}{(t^2+4x+4y)^{3/2}}\right].$$

(I owe the final simplification of the expression (using sines instead of cosines) to B.-G. Englert, a visitor to the physics and math departments in 2001–2002.) To check this, define z by $s = \frac{t}{2} + z$, and note that $(*)$ then becomes ("completing the cube")

$$4z^3 - (t^2 + 4x + 4y)z + 2t(x-y) = 0.$$

Then use

$$\sin^3 t = -\tfrac{1}{4}\sin(3t) + \tfrac{3}{4}\sin t$$

to verify the answer.

Finally, one can recover the condition $t < \sqrt{2x} + \sqrt{2y}$, obtained earlier by geometric reasoning (which I didn't present in detail). A bounce solution must have $0 < s < t$ and also $q_1'(s) < 0$, $q_2'(s) > 0$. (Solutions violating these last inequalities describe unphysical motions that pass through the ceiling and strike it from the wrong side.) From the equation above $(*)$, these inequalities are equivalent to

$$-2x + s^2 < 0, \quad 2y - (t-s)^2 > 0,$$

which imply

$$t = s + (t-s) < \sqrt{2x} + \sqrt{2y}.$$

This shows that the t inequality is necessary for existence of a bounce trajectory. Showing that it is sufficient is harder.