Liverpool University Maths Club

Interesting Paper-Folding Diversions
(Solving the Cubic and Trisecting the Angle)

Mary Fortune & Jon Tims

September 29, 2007

1 Here're the Instructions

How to Solve a Cubic of the Form $x^3 + ax + b = 0$

1 Get a blank sheet of paper.

2 Fold it vertically and horizontally several times to create a grid.

3 It is helpful but optional to markone horizontal line and one vertical line as your axes and mark them with increments of 1.

4 With the axes you have made mark the line $y = -1$ and the point $(0, 1)$ and label them both **A**.

5 Also mark the line $x = -b$ and (b, a) and label them **B**.

6 Now find a way of folding the sheet that simultaneously folds the point **A** on to the line **A** and the point **B** on to the line **B**.

7 The gradient of the fold line you have found is a solution to the cubic equation. (If there are n solutions to the cubic equation then there are n folds that you can find.)

Two good examples to try are $x^3 + x - 2 = 0$ (because we already know the one solution that this has) and $x^3 - x = 0$ (because this has three distinct solutions).

How to Solve Cubics of the Form $x^3 + \alpha x^2 + \beta x + \gamma = 0$

You can solve equations of this form by using the identity

$$x^3 + \alpha x^2 + \beta x + \gamma = (x + \frac{\alpha}{3})^3 + (\beta - \frac{\alpha^2}{3})(x + \frac{\alpha}{3}) + (\gamma + \frac{2\alpha^3}{27} - \frac{\alpha\beta}{3})$$

To elaborate, given a cubic of form $x^3 + \alpha x^2 + \beta x + \gamma = 0$ a solution can be found by changing the variable to $y = x + \frac{\alpha}{3}$; the equation then becomes one of the form $y^3 + ay + b = 0$ which we already know how to solve; therefore, the y-value can be found using the above method and the x-value can be found by using $x = y - \frac{\alpha}{3}$ to convert the solutions of the equation back.

2 First thing's first...

A very important tool in analysing the geometry of folding is to look at the *envelope* of folds in a certain situation. You're not expected to know what an envelope is and we will describe it when it becomes important.

The crucial experiment is

1 Start off with a sheet of unmarked paper (as always!);

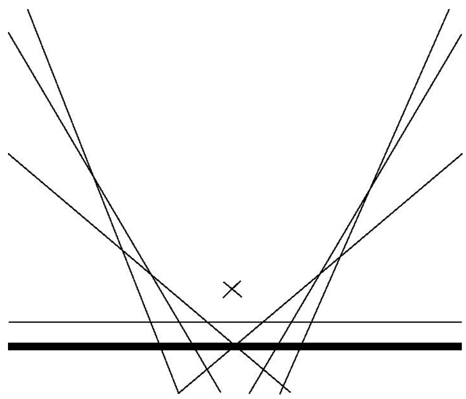
2 mark a line and point (not on the line) on the paper;

3 now fold the point on to the line.

That is, find a way of folding the paper so that in a single fold the point ends up on top of the line. Now, the crucial step in this experiment is

4 Repeat step 3.

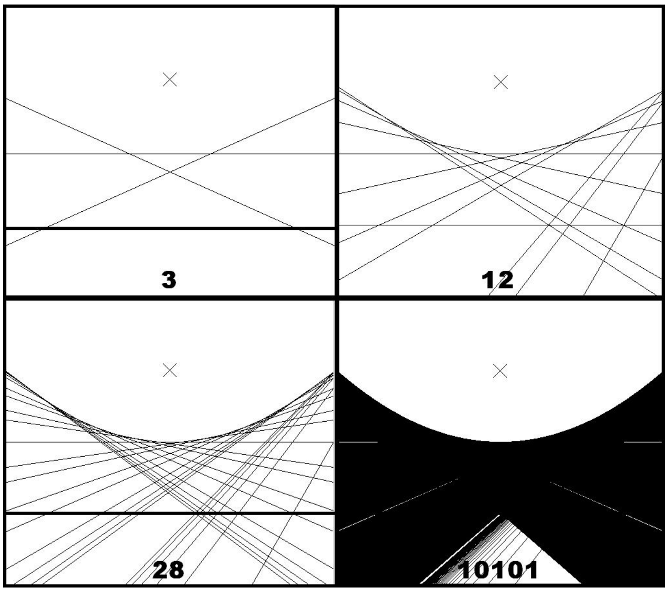
It is obvious that there is an infinite set of ways to fold the point on to the line. However, as you create more and more creases in the paper it becomes clear that the folds seem to trace out a curve. After some strategic folding you should have something that looks like the image below.



It turns out that, as the number of lines tends to infinity, the border that I have emboldened in the above picture does indeed tend to a curve. This curve is called the *envelope* of the lines[1] and refers to any akin phenomenon. You may recognise the type of curve this particular envelope is.

To demonstrate this I have written a computer program in the F# programming language which shows how the folds on a theoretical indestructable sheet look. The pictures below were produced with it and the numbers below indicate the number of folds the computer program drew to produce it.

[1]presumably because the curve envelopes the set of lines – in a sense.



It can be proved (see below) that this curve is Quadratic, that is, if the line is horizontal (as above) then it can be written in the form

$$y = ax^2 + bx + c$$

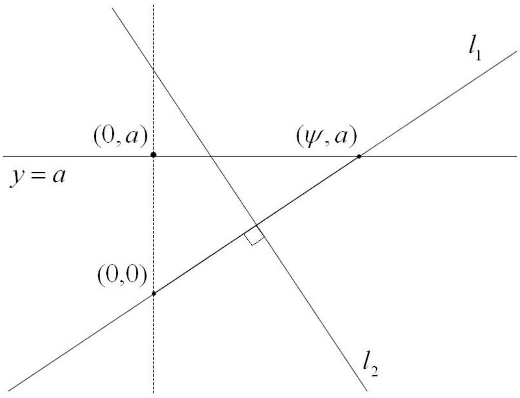
In fact, if the point you marked has coordinates $(0, 0)$ and the line has equation $y = a$ then the equation of the envelope is

$$y = \frac{-1}{2a}x^2 + \frac{1}{2}a$$

Formal

A-Level Module C1/2

Firstly, we note that the envelope of a set of lines is a curve that is *tangent* to every single line. This may sound like a jump but if you look at the pictures above and think about it, we knew this all along. We described the envelope as the curve outlined by the lines; for it to be outlined it must 'only touch' every line once, and therefore every line is a tangent to it (because the tangents are exactly those lines which 'only touch' the curve. So the kernal of this proof will prove exactly that every line is a tangent to the curve.



Now, if we begin with a point $(0,0)$ and a line $y = a$ then we can create a fold that folds the point $(0,0)$ on to any given point of the line. So let us pick a point to fold to and call it (ψ, a).

We name the segment between $(0,0)$ and (ψ, a), l_1. Observe that the fold line that folds $(0,0)$ to (ψ, a) is the perpendicular bisector of l_1 and name it l_2 (ie. l_2 cuts l_1 into two equal segments – and is therefore a bisector – and is at right angles to it – and is therefore perpendicular). As l_1 has gradient $\frac{a}{\psi}$ thence l_2 has gradient $\frac{-\psi}{a}$.

We now know l_2's gradient and we also know that it passes through the midpoint of l_1 (as it is a bisector) which is $(\frac{\psi}{2}, \frac{a}{2})$. hence the equation of the line is

$$\begin{aligned} y - \frac{1}{2}a &= \frac{-\psi}{a}\left(x - \frac{1}{2}\psi\right) \\ ay - \frac{1}{2}a^2 &= -\psi x + \frac{1}{2}\psi^2 \end{aligned}$$

Now, what we wish to prove is that this line is tangent to the curve $y = \frac{-1}{2a}x^2 + \frac{1}{2}a$ no matter what the value of ψ is. To do this we solve the formula for the line and curve as simultaneous equations and find that there is exactly one for every value of ψ. We do this by substituing the equation for the curve into that of the line:

$$\begin{aligned} a\left(\frac{-1}{2a}x^2 + \frac{1}{2}a\right) - \frac{1}{2}a^2 &= -\psi x + \frac{1}{2}\psi^2 \\ \frac{-1}{2}x^2 + \frac{1}{2}a^2 - \frac{1}{2}a^2 &= -\psi x + \frac{1}{2}\psi^2 \\ \frac{-1}{2}x^2 + \psi x - \frac{1}{2}\psi^2 &= 0 \\ x^2 - 2\psi x + \psi^2 &= 0 \end{aligned}$$

and because the descriminant is $(-2\psi)^2 - 4\psi^2 = 0$ we have just proved that the curve is tangent to every line[2].

[2]The eagle eyed reader may notice that this argument does not hold for $a = 0$ because this implies division-by-zero; however, adding the trivial proof that it holds for this case too rectifies the problem but is omitted here.

Finally, the proof may be completed by proving that the quadratic is the only such envelope (otherwise we have not proved that the curve we recognise in the images above is the quadratic we have found); however, this can be done by various methods (the best of which is probably to prove that there is only one point on each line exposed to the open region above that therefore border the curve); however, this level of rigor is too far to bother with for now.

3 Okay, so what next?

Now that we know that the envelope of fold lines from a point to a line is describable by a quadratic, we can look at folding geometry in a very different way; and the problem of solving cubics is only one such problem made easier by this.

To explore this it is necessary to note what the parabola (ie. the type of quadratic curve we are dealing with) means in this case: it is the curve that every fold is a tangent to – this is very important.

Axiom 6 of the Huzita system of Axioms is that

> Given two points P_1 and P_2 and two creases L_1 and L_2, a new crease can be folded such that it folds P_1 into L_1 and P_2 into L_2 if this is possible.

It is helpful to try this with a sheet of paper because until then the axiom itself is largely obscure. Viewing Axiom 6 in terms of fold envelopes we find that we have two fold envelopes (one for each paired point and line) and that if there is a fold as described in the axiom then then this fold must be a tangent to *both* envelopes. Then we can rephrase the axiom as follows.

> Given two parabolas (constructed as fold envelopes) we can construct a simultaneous tangent if it exists.

It is interesting to note at this point that work on Bezout's Theorem assures us that there can be no more than 3 simultaneous tangents; hence, 3 folds. This is one of the wonderful things about mathematics: when proving the theorem Bezout[3] was looking at the problem of conics completely unaware that years later his theorem would say something very important about folding paper!

4 Go on?

This section will be mostly algebra and requires basic A-Level mathematics (specifically Differentiation) but for those who haven't that I will now provide a schematic view of what is going on.

It turns out that when trying to solve the problem of finding simultaneous tangents to parabolas it all boils down to sovling a cubic equation in which the unknown is the gradient of the simultaneous tangent (which is all that is described below). Now, if we want to solve a cubic equation by paper folding what we do is work backwards to find the parabolas that we need to find a simultaneous tangent to; then we work backwards further to find the pair of

[3] I assume this was their name.

points and lines that generates these parabolas; and then find the simultaneous tangent then by simply folding as described above point-to-line and point-to-line simultaneously. Then, because the fold we have made solves the problem of folding it also solves the equivalent problem of solving the cubic. This is exactly why the method of solving a cubic works.

Formal

Basic A-Level. Differentiation.

The most difficult part of deriving the method is to pick the right parabolas to begin with. The best selection I have seen uses

$$y = \frac{1}{2}x^2 \quad \text{and} \quad (y - \frac{1}{2}a)^2 = 2bx \tag{1}$$

Then, we wish to find algebraically a simultaneous tangent meeting the first curve at (x_1, y_1) and the second at (x_2, y_2) and having gradient μ.

Differentiating the equations respectively gives

$$\frac{dy}{dx} = x \quad \text{and} \quad 2(y - \frac{1}{2}a)\frac{dy}{dx} = 2b \tag{2}$$

and substituting in the points above where $\frac{dy}{dx} = \mu$ gives

$$\mu = x_1 \quad \text{and} \quad (y_2 - \frac{1}{2}a)\mu = b \tag{3}$$

whereas substituting instead simply the points (x_1, y_1) and (x_2, y_2) into (1) gives

$$y_1 = \frac{1}{2}x_1^2 \quad \text{and} \quad (y_2 - \frac{1}{2}a)^2 = 2bx_2 \tag{4}$$

but also, because the tangent passes through (x_1, y_1) and (x_2, y_2),

$$\mu = \frac{y_2 - y_1}{x_2 - x_1} \tag{5}$$

and then to eliminate x_1, y_1, x_2, y_2 we substiute in (3) and (4) into (5) give

$$\begin{aligned}
\mu &= \frac{\frac{b}{\mu} + \frac{1}{2}a - \frac{1}{2}x_1^2}{\frac{(y_2 - \frac{1}{2}a)^2}{2b} - \mu} \\
&= \frac{\frac{b}{\mu} + \frac{1}{2}a - \frac{1}{2}\mu^2}{\frac{(\frac{b}{\mu} + \frac{1}{2}a - \frac{1}{2}a)^2}{2b} - \mu} \\
&= \frac{\frac{b}{\mu} + \frac{1}{2}a - \frac{1}{2}\mu^2}{\frac{b}{2\mu^2} - \mu}
\end{aligned}$$

therefore

$$\begin{aligned}
\mu(\frac{b}{2\mu^2} - \mu) &= \frac{b}{\mu} + \frac{1}{2}a - \frac{1}{2}\mu^2 \\
b - 2\mu^3 &= 2b + a\mu - \mu^3
\end{aligned}$$

and so the cubic we are solving by finding the simultaneous tangent to the parabolas written above is

$$\mu^3 + a\mu + b = 0$$

and that is that[4].

Well, almost.

In fact, this shows us what parabolas we need to find a simultaneous tangent to to solve the cubic $x^3 + ax + b = 0$ but does not show us pair of lines and points we need to fold to find the simultaneous tangents!

To do this, simply translate and reflect the formulae we found in our discussion of envelopes to show that $y = \frac{1}{2}x^2$ is generated by folding $(0, \frac{1}{2})$ on to $y = -\frac{1}{2}$; and $(y - \frac{1}{2}a)^2 = 2bx$ is generated by folding $(\frac{1}{2}b, \frac{1}{2}a)$ on to $x = -\frac{1}{2}b$.

The final step is then to simply drop all the fractions: this isn't as silly as it first sounds because if we multiply all the coordinates by 2 (and so kill every $\frac{1}{2}$) then the gradient of the line will be exactly the same.

5 Even Further!?

One of the other things that this leads to is the problem of trisecting the angle. This is based on the following identity

$$\tan 3\theta = \frac{3\tan\theta - \tan^3\theta}{1 - 3\tan^2\theta}$$

which we can prove using the double angle formula

$$\tan(\theta + \phi) = \frac{\tan\theta + \tan\phi}{1 - \tan\theta\,\tan\phi}$$

by writing

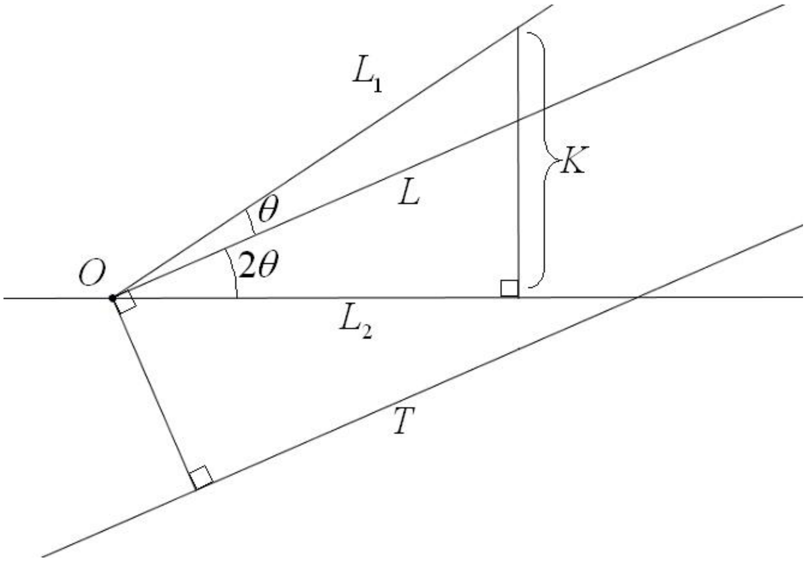
$$\begin{aligned}
\tan 3\theta &= \tan(2\theta + \theta) \\
&= \frac{\tan 2\theta + \tan\theta}{1 - \tan 2\theta\,\tan\theta} \\
&= \frac{\frac{2\tan\theta}{1-\tan^2\theta} + \tan\theta}{1 - \frac{2\tan\theta}{1-\tan^2\theta}\,\tan\theta} \\
&= \frac{2\tan\theta + (1-\tan^2\theta)\tan\theta}{(1-\tan^2\theta) - 2\tan^2\theta} \\
&= \frac{3\tan\theta - \tan^3\theta}{1 - 3\tan^2\theta}
\end{aligned}$$

This can be rewritten as

$$\tan^3\theta - 3\tan 3\theta\,\tan^2\theta - 3\tan\theta + \tan 3\theta = 0$$

which becomes more useful for our purposes if we write $t = \tan\theta$ and $u = \tan 3\theta$; hence, $t^3 - 3ut^2 - 3t + u = 0$, which is a quadratic in t which we can solve for any given u, but the question is "why would we want to?"

[4]As before there are quite a few division by zero problems here and there but the best way to solve that problem is to ignore it.



The answer is elaborated by the following reasoning, illustrated by the picture. We wish to trisect an angle between lines L_1 and L_2 (ie. the angle between L_1 and L_2 is 3θ and we want to find the line L such that the angle between L and L_2 is θ). If we create a new line K at any point which is perpendiculuar (at right angles) to L_2 then we create a right-angled triangle[5].

Now, we can call the length of the L_2 side of the triangle L length. Then the length of the K side is $u = \tan 3\theta$. We then know that if we attempt to solve the above for $t = \tan \theta$ that in the process we will create a fold-line of gradient t, which is therefore parallel to the line L we want!

We can then create a line parallel to this which passes through the point O (the intersection of L_1 and L_2), ie. the line L itself, using the following method: call the line with gradient t we have generated T; find the fold that passes through O and is perpendicular to T; then find the fold that passes through O and is perpendicular to the fold we have just created. This final fold is exactly L and both of the two intermediate folds can be created under Axiom 4 of the Huzita system of axioms.

6 Comparison with CS-Geometry

CS (Compass and Straightedge)-Geometry is exactly the original geometry of Euclid. Where in PF (Paper Folding)-Geometry anything is possible that can be folded (loosely speaking), in CS geometry anything is possible that can be created with a compass and completely unmarked straightedge.

The two operations described above (solving the general cubic and trisecting the general angle) are both impossible in CS-Geometry[6] and the obvious question is "does there exist a construction possible in CSG but impossible in PFG?"

[5]This is made possible by axiom 4 of the Huzita system of axioms

[6]The proof of this wonderful and well-known fact is unknown to the author who thinks it may be based on Galois Theory but is sure the interested reader will enjoy researching it further.

The answer is "No"; that is, every single thing possible in CSG is possible in PFG; that is, PFG is a *stronger* Geometry.

It is interesting to observe that the Greeks (who Axiomised CSG over the course of many lifetimes – a work that reached rigor with Euclid's Elements) picked a weak Geometry and that mathematics may have been very different had they picked the stronger PFG geometry. Possibly mathematics would not be so advanced now: the PFG Axioms contain no explicit formulation of the 6^{th} Euclidean Axiom, the discomfort over which provoked the investigation of Hyperbolic and Elliptical Geometry; and the fact that certain unexpected are impossible gave Galois Theory something to sink its teeth into in the first instance, and possibly provided some motivation for its development; but this is idle speculation truthfully.

7 Further Reading

This handout gets you off to a very good start but there is a lot of material on the internet to explore. This section is to give you a few clues on how to go about that.

The mathematics of Paper Folding (or Origami) was advanced when Huzita (whom I have mentioned once previously in passing) broke down every possible way of folding paper into just six unique operations that he thought could be combined to produce every possible fold. In fact, he was wrong, but not far off: he missed just one fold, which makes seven, and this was discovered recently by Hatori. Together the seven folds are called the Huzita-Hatori System of Axioms. There is more about this on www.cut-the-knot.org[7] (search for Origami and Paper Folding). However, it is only fair to say that the Huzita-Hatori Axioms are just the most popular of several systems describing Paper Folding.

Huzita lived an interesting life aside from his mathematics and a biography of his life (Google it!) is worth a read.

Finally, if you wish to know more about CS-Geometry then there is a wealth of information on the internet that is only a Google away with some simple and many difficult proofs that trisecting the angle is impossible with CS.

[7] www.cuttheknot.com is also a mathematical site: it helps people divide by 2.