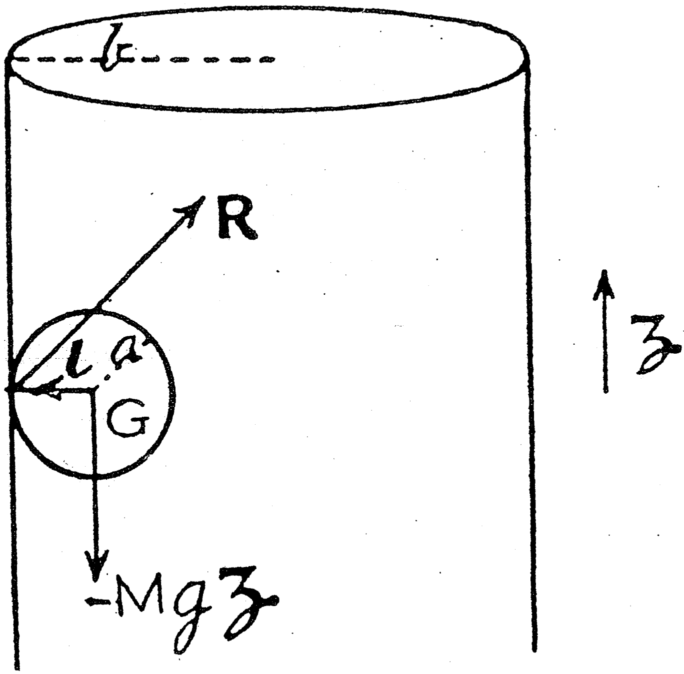
Function

Founder Editor G. B. Preston

Volume 15 Part 2 April 1991



A SCHOOL MATHEMATICS MAGAZINE

by Aust. Post Publ. No. VBH0171

FUNCTION is a mathematics magazine addressed principally to students in the upper forms of secondary schools.

It is a 'special interest' journal for those who are interested in mathematics. Windsurfers, chess-players and gardeners all have magazines that cater to their interests. FUNCTION is a counterpart of these.

Coverage is wide — pure mathematics, statistics, computer science and applications of mathematics are all included. Recent issues have carried articles on advances in mathematics, news items on mathematics and its applications, special interest matters, such as computer chess, problems and solutions, discussions, cover diagrams, even cartoons.

* * * * *

Articles, correspondence, problems (with or without solutions) and other material for publication are invited. Address them to:

The Editors,
FUNCTION,
Department of Mathematics,
Monash University,
Clayton, Victoria, 3168.

Alternatively correspondence may be addressed individually to any of the editors at the mathematics departments of the institutions listed on the inside back cover.

FUNCTION is published five times a year, appearing in February, April, June, August, October. Price for five issues (including postage): $16.00*; single issues $3.50. Payments should be sent to the Business Manager at the above address: cheques and money orders should be made payable to Monash University. Enquiries about advertising should be directed to the business manager.

*$8.00 for *bona fide* secondary or tertiary students.

* * * * *

FUNCTION

Volume 15 *Part 2*

(Founding editor: G.B. Preston)

CONTENTS

* * * * *

Published by the department of Mathematics, Monash University

Printed by MONASH UNIVERSITY PRINTING SERVICES
900 Dandenong Road Caulfield East, Vic. 3145

THE FRONT COVER

Michael A.B. Deakin, Monash University

The diagram reproduced on the front cover and opposite is taken from an older text of classical mechanics: E.A. Milne's *Vectorial Mechanics*.

It depicts a somewhat unusual situation: a sphere of radius a is rolling around inside a fixed vertical cylinder of radius b (where, of course, $b > a$).

Two forces act on the sphere. First there is its own weight which may be regarded as a vector $-Mg\mathbf{k}$ acting vertically down, and passing through the centre of the sphere. Second, there is the force exerted on the sphere by the cylindrical wall. This passes through the point of contact as indicated, and has two components: an inward one which acts to keep the sphere confined by the cylinder and a vertical component due to the action of friction at the contact (it is friction that causes the sphere to roll rather than to slide). The vector sum of these components is called $\mathbf{R}$. (An older notation $\mathbf{z}$ is used for $\mathbf{k}$.)

The other letters on the diagram represent other quantities: $\mathbf{i}$ is a unit vector from the centre of the sphere towards the point of contact, G is the centre of the sphere, M is its mass and g is a constant called the acceleration due to gravity.

The problem is to calculate, from Newton's laws of motion, what happens. This is a study in the field known as *Three-Dimensional Rigid Body Dynamics*, which can be quite complicated and now is not widely taught, although it was once very fashionable. (This precise problem was set for us as an exercise when I was in my second year at University.)

I will not give the analysis which takes about two pages of vector calculus, somewhat beyond the secondary syllabus and using equations not now familiar. However, the conclusions can be given and they are surprising.

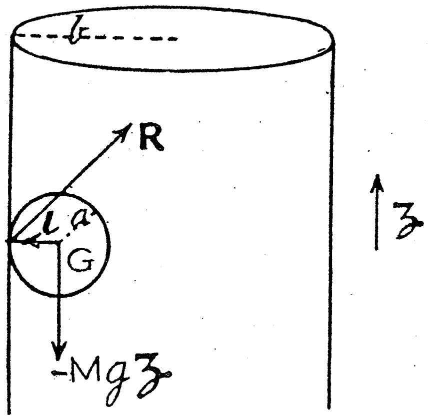
Concentrate on the motion of G. This turns out to travel around the axis of the cylinder at constant angular velocity. In the vertical direction, we might expect that the sphere would gradually spiral "downhill", so to speak. However, it doesn't, or to be more precise, in this slightly idealised model it doesn't. ζ, the height of G above its average level, varies sinusoidally with time. So the sphere moves round and round on the surface of the cylinder travelling up and down in a sine-wave as it goes!

In real life, the sphere would not be quite rigid and would suffer small energy-losing deformations that ultimately would cause the motion to run down, but the rigid approximation is actually quite a good one.

The answer seems so much at variance with "common sense" that many people, on first hearing it, query it.

However, it explains a real-life phenomenon. You may have seen it on TV. A golfer putts the ball toward the hole – it seems to go in, indeed it does go in, but it rolls back out. What has happened is that when the ball entered the hole, its motion was such that it commenced to roll around the side of the hole. Initially its vertical motion was downward, but then (because of the sinusoidal motion) it came back up. If when it comes back to the top, the lip of the hole is very slightly lower than it was where the ball went in, the ball rolls back out of the hole!

Three-Dimensional Rigid Body Dynamics is now not very widely studied and so we see cases where people who should know better have claimed that such violations of "common sense" revealed limits to Newton's laws of motion. Various cranks have made such claims in respect of gyroscopes, when there is a fully developed and perfectly adequate theory to explain all the phenomena they adduce. Yet again I saw a letter to *New Scientist* saying that snooker balls hit with "side" violated Newton's laws. They don't: they travel along parabolic arcs exactly as theory says they should!



* * * * *

A RADICAL PROBLEM

Peter Grossman, Monash University

A great deal of school mathematics is devoted to learning techniques for simplifying mathematical expression: expanding and factorising polynomials, finding common denominators, cancelling common factors, collecting terms, and carrying out various other manipulations, all with the aim of making an expression as simple as possible.

Radicals

This article is about expressions containing *radicals*: square roots, cube roots, and so on. No doubt you have learnt some techniques for simplifying expressions of this kind, and would have no trouble when presented with something like:

$$\frac{\sqrt{10}}{\sqrt{5}}$$

or:

$$\frac{\sqrt{54} + \sqrt{15}}{3\sqrt{3}}.$$

If you think this is all pretty basic, here's a more difficult one for you to try.

Express the following number in the simplest possible form:

$$\sqrt{3+2\sqrt{2}}. \qquad (1)$$

After looking at this expression for a while, you will probably conclude that it isn't the kind of problem where you could apply any of the techniques you have learnt, and you might suspect that it can't be expressed any more simply than it already is. (We could use a calculator to find a numerical approximation, of course, but we are not interested in doing that here.)

Surprisingly, Expression (1) *can* be simplified: it equals $1 + \sqrt{2}$. Knowing the answer, it's easy to check by squaring it that it's correct. However, we are still left with a problem: how could we have discovered the simplified form of Expression (1)?

There has been a lot of interest recently in the problem of finding algorithms for simplifying mathematical expressions. Such algorithms are used in computer algebra software: programs for manipulating *mathematical expressions* rather than (or in addition to) decimal numbers. Commercially available computer algebra systems can be used to carry out many of the techniques you have learnt to do with pen and paper, including doing exact arithmetic with arbitrarily large numbers, solving equations, differentiating and integrating, manipulating vectors, matrices and complex numbers, drawing graphs of functions, and many other more sophisticated mathematical techniques. A computer algebra system needs an armoury of techniques for simplifying expressions without approximating them if it is to work properly.

Returning to our problem, let

$$x = \sqrt{3+2\sqrt{2}}.$$

Then

$$x^2 = 3 + 2\sqrt{2} \qquad \text{(squaring both sides)}$$

so

$$x^2 - 3 = 2\sqrt{2}$$

and thus

$$(x^2 - 3)^2 = 8$$

which, upon expanding and rearranging, becomes:

$$x^4 - 6x^2 + 1 = 0. \qquad (2)$$

The key to simplifying (1) is the fact that Equation (2) can be factorised by completing the square on the first and last terms of the left-hand side as follows:

$$(x^2 - 1)^2 - 6x^2 + 2x^2 = 0.$$

Therefore

$$(x^2 - 1)^2 - (2x)^2 = 0$$

so

$$(x^2 - 2x - 1)(x^2 + 2x - 1) = 0$$

and hence

$$x^2 - 2x - 1 = 0 \quad \text{or} \quad x^2 + 2x - 1 = 0. \qquad (3)$$

Using the quadratic formula, Equations (3) are equivalent to

$$x = 1 \pm \sqrt{2} \quad \text{or} \quad x = -1 \pm \sqrt{2}.$$

Only one solution, $1 + \sqrt{2}$, is the required simplification of (1). The other three solutions occur because of the squaring (twice) of both sides of the equation.

This example suggests that the problem of simplifying expressions containing radicals (i.e. expressions involving n^{th} roots) is closely related to the problem of factorising polynomials, which in turn is related to the problem of solving polynomial equations. It will therefore pay us to take a closer look at polynomial equations, which we now do.

Polynomial Equations and (some of) their Solutions

The simplest example of a polynomial equation is a linear equation:

$$ax + b = 0. \tag{4}$$

The solution is of course:

$$x = -b/a. \tag{5}$$

Notice that we can solve the equation in terms of its coefficients a and b using just the operations of subtraction and division.

Next is the quadratic equation:

$$ax^2 + bx + c = 0 \tag{6}$$

for which we have the well-known solution:

$$x = \frac{-b \pm \sqrt{b^2-4ac}}{2a}. \tag{7}$$

In contrast to the linear equation, the operations of addition, subtraction, multiplication and division are not sufficient to allow us to obtain the solution, and we need to use a radical (in this case the extraction of a square root) as well.

Can we solve polynomial equations of higher degree by radicals? The expression "by radicals" means we want to be able to calculate the solution from the coefficients of the polynomial, using only the operations of addition, subtraction, multiplication, division and extraction of n^{th} roots, where n may be any positive integer.

This problem exercised the minds of a number of mathematicians in earlier times, and they succeeded in obtaining solutions for cubic (degree 3) and quartic (degree 4) equations. You have probably not seen the formula for the solution of a general cubic equation, and it is rather too complicated to give here. As you might imagine, it involves taking cube roots. (It also involves taking square roots.) The solution of a general quartic equation is even more complicated.

At this point, the well runs dry. Despite numerous attempts by mathematicians to solve a general quintic (degree 5) equation, no such solution was forthcoming. It was a triumph when the Norwegian mathematician Niels Abel (1802-1829) established that no solution by radicals exists for a general polynomial equation of degree 5 or more. (It should be emphasised that this does not mean that solutions of these equations do not exist, only that they cannot be expressed in this particular form.) A general theory for determining precisely which polynomial equations can be solved by radicals was subsequently developed by the French mathematician Evariste Galois (1811-1832 – yes, those dates are right! – see *Function, Vol.* 3, *Part* 2).

An Excursion into the Fields

In order to investigate these ideas more closely, we need to study mathematical objects called *fields*. (The definition that follows is not actually the most general definition of a field, but it is sufficient for our purposes.)

We shall call a set F of numbers a *field* if:

(i) 0 and 1 are elements of F;

(ii) addition, subtraction, multiplication and division of elements of F always yield a result in F (except that division by zero is not permitted).

The set R of real numbers and the set Q of rational numbers are examples of fields, while the set J of integers is not. The set C of complex numbers, which some readers will have studied, is also a field. (We should really include complex numbers in our discussion in order to deal with solutions of polynomial equations in full generality; however, we shall restrict out examples to real numbers for the benefit of readers who are unfamiliar with complex numbers.)

Another example of a field is $\{q + r\sqrt{2} : q,r \in Q\}$, which is denoted by $Q(\sqrt{2})$. Note that $3 - \sqrt{2} \in Q(\sqrt{2})$ (choosing $q = 3$, $r = -1$), $\frac{-5 + 3\sqrt{2}}{4} \in Q(\sqrt{2})$ (choosing $q = -\frac{5}{4}$, $r = \frac{3}{4}$), but $\sqrt{3} + 2\sqrt{2} \notin Q(\sqrt{2})$ since q and r must both be rational. $Q(\sqrt{2})$ contains Q as a subset (as a *subfield*, in fact), as can be seen by choosing $r = 0$ and letting q be any rational number.

It is not hard to check that the sum, difference and product of any two elements $q_1 + r_1\sqrt{2}$ and $q_2 + r_2\sqrt{2}$ of $Q(\sqrt{2})$ are also in $Q(\sqrt{2})$. Showing that the quotient is in $Q(\sqrt{2})$ (assuming that q_2 and r_2 are not both zero) is a little trickier: try multiplying the numerator and denominator of $\dfrac{q_1 + r_1\sqrt{2}}{q_2 + r_2\sqrt{2}}$ by $q_2 - r_2\sqrt{2}$ and then simplifying.

The field $Q(\sqrt{2})$ is fairly typical of the kinds of fields that arise when we investigate the solution of polynomial equations by radicals. Clearly, we could obtain other examples of fields by replacing 2 in this example by other numbers, giving a whole family of fields $Q(\sqrt{n}) = \{q + r\sqrt{n} : q,r \in Q\}$.

More complicated examples are possible, such as $\{q + r\sqrt{2} + s\sqrt{3} + t\sqrt{6} : q,r,s,t \in Q\}$, which is denoted by $Q(\sqrt{2},\sqrt{3})$. (If you are wondering why $\sqrt{6}$ appears here, just note that any field containing $\sqrt{2}$ and $\sqrt{3}$ must contain $\sqrt{6}$ also, since $\sqrt{6} = \sqrt{2} \times \sqrt{3}$.) There are also fields that are defined in a similar fashion but involving cube roots or higher: $\{p + 2^{1/3}q + 2^{2/3}r : p,q,r \in Q\}$ is an example.

In order to see how fields of this kind can be related to our problem, let's consider again the general quadratic equation, Equation (6), and assume that the coefficients a,b,c are all rational numbers. We can calculate the discriminant $b^2 - 4ac$ while remaining within the field Q, since the calculation involves only subtraction and multiplication. Unless $b^2 - 4ac$ happens to be a perfect square, however, we have to 'step up' to the larger field $Q(\sqrt{b^2-4ac})$ in order to take the square root. The remaining calculations (subtracting b and dividing by $2a$) take place within this larger field. The field $Q(\sqrt{b^2-4ac})$ is described as an *extension field* of Q; we have just shown that the solution of the quadratic equation $ax^2 + bx + x = 0$ is an element of this field.

This idea helps to shed some light on our original problem of simplifying $\sqrt{3+2\sqrt{2}}$. The introduction of $\sqrt{2}$ puts us in the field $Q(\sqrt{2})$, and we remain within that field when we consider $3 + 2\sqrt{2}$. It appears at first that we have to step up to a larger field when we take the square root of this number, but in fact it is not necessary to do so, since the answer, $1 + \sqrt{2}$, is an element of $Q(\sqrt{2})$. A slight change in one of the numbers in the original problem (replacing 3 by 2, for example) would mean that the second square root *would* involve stepping up to a larger field.

In general, expressions such as (1) which involve "nested radicals;" give rise to a sequence of fields, starting at Q, with each field containing the previous one. The evaluation of each radical in the expression usually corresponds to a step up to the next field in the sequence. Exceptionally, a step can sometimes be avoided because a radical can be evaluated within the current field, and whenever this happens there is a way of simplifying the original expression so that it contains fewer levels of nesting.

Quintic Equations Can't be Solved by Radicals!

With these ideas in mind, it is now possible to say a little more about why quintic equations can't be solved by radicals.

The solutions of a polynomial equation are elements of a field known as the *splitting field* of the polynomial, since it is related to the way the polynomial "splits" into factors. For example, the splitting field of $x^2 - 2$ is $Q(\sqrt{2})$, since this field contains the solutions $\sqrt{2}$ and $-\sqrt{2}$ of the equation $x^2 - 2 = 0$ (and is the smallest such field).

If we want to solve a polynomial equation by radicals, we need to look for a sequence of fields in which each field is an extension field of the previous one. The first field in the sequence is Q and the last is the splitting field of the polynomial, and each step in the sequence corresponds to the evaluation of a radical.

At this point the details become complicated, and we can do no more than give an outline of what happens. It turns out that associated with each of these fields is a certain set of functions which form a mathematical structure called a *group*. The group associated with the splitting field (the last field in the sequence) is called the *Galois group* of the polynomial, and this group must satisfy a certain technical condition if the required sequence of fields is to exist. It can be shown that the Galois group of a general polynomial with degree 5 or more does not satisfy this condition. Therefore any attempt to find a general formula for solving quintic equations by radicals is destined to fail.

How can Expressions be Simplified?

If we are given an expression involving nested radicals, such as $\sqrt{3+2\sqrt{2}}$, then we saw earlier how we could find a polynomial equation with rational coefficients of which the expression is a solution. In this case, the Galois group of the polynomial will always satisfy the condition referred to in the previous paragraph, but this doesn't tell us anything useful, since we already know that the equation can be solved by radicals – one of the solutions is precisely the expression we started with!

Enter Susan Landau, a computer scientist at Wesleyan University in the United States. In an article in the journal *Science* (15 September 1989), it was reported that Landau had developed an algorithm for "denesting" nested radicals into their simplest form. Earlier researchers had found algorithms that worked in special cases, but Landau's algorithm is the only one that can be applied in general. The algorithm systematically searches inside the Galois group for the shortest possible sequence of subgroups satisfying certain conditions. This sequence of subgroups corresponds in a particular way to the sequence of fields for the denested expression being sought, and hence to the denested expression itself. The search algorithm can be guaranteed to stop after a finite time because Galois groups of polynomials always contain only a finite number of elements. (The Galois group for $x^2 - 2$, for example, has just 2 elements, while the Galois group for the general quintic polynomial has 120 elements.)

Unfortunately, as a subsequent article in *Science* (24 November 1989) revealed, there was a mistake in Landau's original proof that the algorithm always produces the least nested form of the expression. Landau has since modified the proof to show that the result is always either the least nested form or a form with at most one extra level of nesting. She has not actually found any examples of expressions that don't reduce to the least nested form when run through the algorithm, so her original claim might still turn out to be correct.

Even if Landau's algorithm is eventually shown to produce the least nested form in all cases, it will not be the last word on the subject. The algorithm requires a large amount of computation: inserting just one extra radical sign in the original expression can more than double the computation time. The search will no doubt be on among mathematicians and computer scientists for ways of improving the algorithm's efficiency. It would be interesting to know how long it would take a computer running Landau's algorithm to simplify the following expression:

$$\sqrt[3]{\sqrt[5]{32/5} - \sqrt[5]{27/5}} - \sqrt[5]{1/25} - \sqrt[5]{3/25} + \sqrt[5]{9/25}. \qquad (8)$$

Early this century, well before the advent of the computer age, the Indian mathematical genius Srinivasa Ramanujan correctly stated that Expression (8) is equal to zero! At a time when it is easy to be overwhelmed by the blazing computational power of the latest computers, it is worth reflecting on the ability of the human mind to tackle problems that stretch the limits of our fastest machines.

Perhaps the most interesting theme to arise out of these ideas is the increasingly complex relationship that continues to develop between computing and mathematics. It is not just a matter of using computers to perform calculations; the computer is profoundly influencing the directions that mathematics itself is taking. We have seen in this article how the "pure" mathematics of Abel, Galois and others is now an important tool in the design of computer software, which in its turn is driving mathematical research in directions undreamt of by those earlier mathematicians. And a simple problem with square roots has led us into an area of current research where there are many questions for which no-one has yet found satisfactory answers.

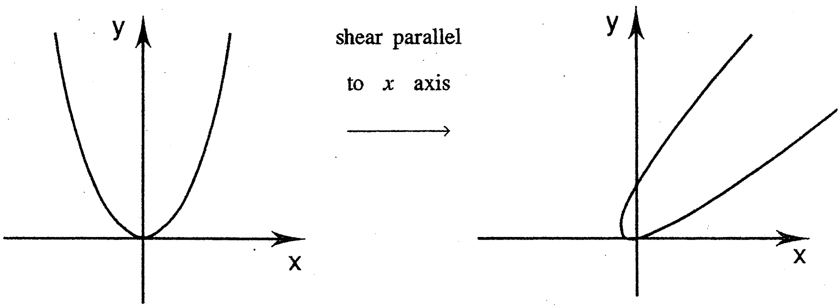
* * * * *

FACTS AND FORMULAS ABOUT SHAPES AND TRANSFORMATIONS OF CONIC SECTIONS

D.F. Charles, Pascoe Vale Girls High School

If you ask a Maths teacher or a student who has studied conics "What is the eccentricity (e) of a parabola?" they will have no difficulty in telling you that $e = 1$. However, teachers do also describe parabolas as 'wider', 'narrower', or 'steeper' depending upon the constant k in the equation $y = kx^2$. In fact, the textbooks do the same thing and introduce the concept of dilation of graphs by changing the constant in front of the function. Unfortunately, most students then retain the erroneous idea that there is a 'family' of parabolas all roughly similar in shape, but some 'wider' and some 'narrower' than the 'normal' parabola; even those students who go on to become Maths teachers!

The problem lies in the fact that the parabola is not a good example with which to introduce the concept of dilation because no matter how you try to dilate it or indeed transform it with any (non-degenerate) linear transformation, it retains exactly the same shape. Even trying to shear the parabola has no effect on its shape.



The only effect that the linear transformation can have is to alter the parabola's position on the axes or to scale it up or down. This can be demonstrated by an accurate plot of $y = x^2$ and $y = 2x^2$, using the photocopier to diminish $y = x^2$ and seeing that it fits exactly over $y = 2x^2$. The same thing can be demonstrated by enlarging the parabola $y = x^2$ by a factor of 2 using the matrix

$$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

then

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad \therefore \quad \begin{matrix} x' = 2x \\ y' = 2y \end{matrix} \iff \begin{matrix} x = \frac{1}{2}x' \\ y = \frac{1}{2}y' \end{matrix}$$

substituting in the original equation and dropping dashes gives:

$$y = \tfrac{1}{2}x^2.$$

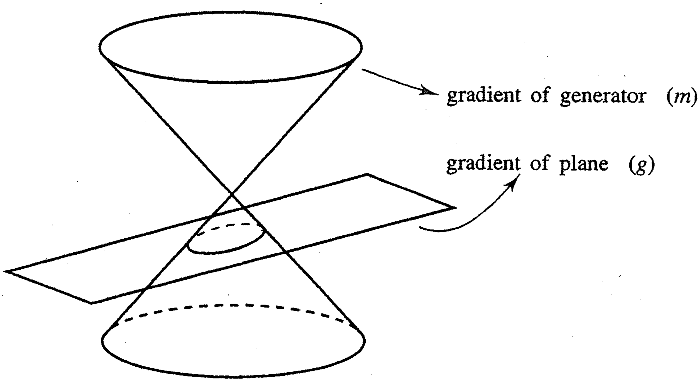
It is interesting to compare the invariance of the parabolic shape with other conic sections. This can be done from the mathematical definition of a conic (as the locus of a point whose distance from a focus and a directrix are in the ratio e).

One fact which emerges is that e determines the *exact* shape of a conic. E.g. all conics with $e = 0.7$ will be exactly the same shape and differ only in scale or position.

This explains why all parabolas are the same shape because for all of them $e = 1$. There are no 'thinner' or 'wider' parabolas.

To compare shapes of conics therefore, we need only find e. The following formulas give quick ways to find e for two of the other definitions of a conic section.

Plane – Circular Cone definition



By a suitable orientation of the cone on the axes, it can be shown that:

$$e^2 = \frac{g^2(1+1/m^2)}{1+g^2}.$$

One can see that any translation of the plane does not affect e and will only change the scale. (This fact is obvious for the circle or ellipse, but not as obvious for the parabola or hyperbola.)

Polynomial definition

If the conic is looked at as the locus of points given by the ordered pairs (x,y) related by the general 2nd degree equation in x & y:

$$Ax^2 + By^2 + Cxy + Dx + Ey + F = 0$$

then e can be found from the formula

$$e^2 = \frac{2\left[(A-B)^2+C^2\right]^{1/2}}{|A+B|+\left[(A-B)^2+C^2\right]^{1/2}}$$

This formula is convenient in that it eliminates the necessity to rotate and translate the conic into a more manageable form in order to find e.

Lastly we can look at the limitations (if any) of a non-degenerate linear transformation in shape alteration and also ask the question "Can a linear transformation change one type of conic into another?"

The answer to this question is "No". Certainly it can change the e-value of a hyperbola or ellipse and hence change its shape but only within the range $0 \leq e < 1$ for an ellipse and $e > 1$ for a hyperbola. The point where $e = 1$ is a stumbling block for a linear transformation. The reason for this is that a linear transformation has the effect of multiplying the ratio $\frac{b^2}{a^2}$ where:

$$e^2 = 1 \pm b^2/a^2.$$

If b^2/a^2 is already zero (as in the parabola), then e doesn't change. We can prove the above statements more rigorously as follows:

Proof

Any ellipse, by a suitable translation and rotation, can be transformed into the form:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1. \qquad \ldots\ldots \qquad (E)$$

Let us transform this with a linear transformation T represented by the matrix

$$T = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \text{ whose inverse is } T^{-1} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

then:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix} \iff \begin{cases} x = Ax' + By' \\ y = Cx' + Dy'. \end{cases}$$

Substituting into Equation (E) gives (after dropping dashes):

$$\left[\frac{A^2}{a^2}\frac{C^2}{b^2}\right]x^2 + \left[\frac{B^2}{a^2}\frac{D^2}{b^2}\right]y^2 + 2\left[\frac{AB}{a^2} = \frac{CD}{b^2}\right]xy = 1.$$

If we let

$$M = \text{coefficient of } x^2$$
$$N = \text{coefficient of } y^2$$
$$2H = \text{coefficient of } xy,$$

then the condition for the equation to represent each type of conic is:

$$H^2 - MN > 0 \text{ gives a Hyperbola}$$
$$H^2 - MN = 0 \text{ gives a Parabola}$$
$$H^2 - MN < 0 \text{ gives an Ellipse.}$$

Evaluating $H^2 - MN$ for our equation gives:

$$H^2 - MN = -\frac{(AD\text{-}BC)^2}{a^2b^2} = -\frac{(\det T^{-1})^2}{a^2b^2}.$$

This is clearly always negative, indicating that the transformed graph must also be an ellipse. (Note that $\det T^{-1}$ cannot be zero because if it were, T^{-1} would be singular having no inverse and so T would not exist.)

Similar reasoning applied to a hyperbola

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1 \text{ gives } H^2 - MN = \frac{(\det T^{-1})^2}{a^2b^2}$$

again ensuring that the transformed graph is also a hyperbola.

Thus if no transformation from hyperbola or ellipse to parabola exists, then the inverse transformation from parabola to other conic cannot exist either.

Students of Physics will be interested to know that the above fact is reflected in nature. The path of a body around the sun is a conic section. If the body is captured in orbit then the path is an ellipse. If the body has just enough energy to escape then the path is a parabola, more energy, then the path is a hyperbola.

Special Relativity deals with different views of a system in different inertial frames of reference. The transformation of coordinates from one frame to another is linear. If a linear transformation could change the type of conic, then it would be possible for an observer in one frame of reference to state that a body is captured in orbit since its path is elliptical, while another observer in a different frame could say that the body is not captured since its path is parabolic or hyperbolic. This is not possible because the body is either captured or not.

Question: Are there any curves other than the parabola which are immune to a shape change by any linear transformation?

* * * * *

PI

Karl Spiteri, Student, University of Melbourne†

If we take a unit circle we may inscribe a square in it as shown in Figure 1a. The perimeter of the square will then be $4\sqrt{2}$ and the circumference of the circle will be 2π. Figure 1b shows how an octagon may be constructed from the square. We would expect the perimeter of the octagon to give a better approximation to 2π. (After all, $4\sqrt{2} = 5{\cdot}656$ ... while $2\pi = 6{\cdot}283$ ..., so the first approximation is not very good.)

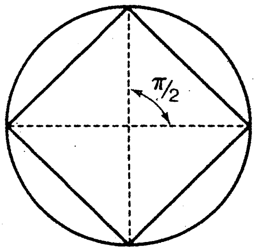


Figure 1a

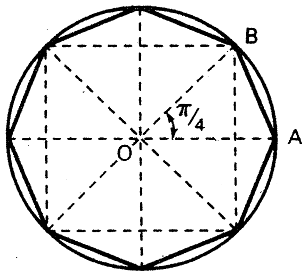


Figure 1b

Indeed, we would expect to get better and better approximations as we go from a square to an octagon to a 16-gon to a 32-gon and so on.††

Let x_n be the length of the side in a 2^n-gon. First take the case $n = 3$. See Figure 2.

We can use the cosine rule to find x_3 :

$$x_3^2 = 1^2 + 1^2 - 2 \times 1 \times 1 \cos \frac{\pi}{4},$$

i.e.

$$x_3^2 = 2 - \sqrt{2}.$$

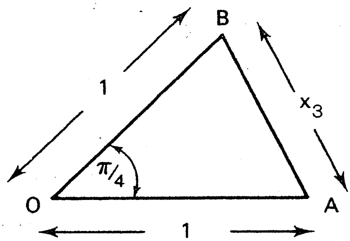


Figure 2

So it follows that

$$x_3 = \sqrt{2-\sqrt{2}}.$$

† This is adapted from a prize-winning project written in 1988 when the author was in Year 11.

†† For another approach to this, see *Function, Vol.* 4, *Part* 1.

The perimeter of the octagon is therefore

$$2^3 \times \sqrt{2-\sqrt{2}} = 6{\cdot}1229 \ldots$$

which gives the approximation

$$\pi \approx 3{\cdot}0614 \ldots .$$

To continue we need to evaluate $\cos(\pi/2^{n-1})$. This can be done using the trigonometric identity

$$\cos^2\theta = \tfrac{1}{2}(1 + \cos 2\theta).$$

This gives

$$\cos(\pi/2^{n-1}) = \tfrac{1}{2}\sqrt{2+2\cos(\pi/2^{n-2})}.$$

Using this result we may find, from the value of $\cos(\pi/4)$, $\cos(\pi/8) = \frac{1}{2}\sqrt{2+\sqrt{2}}$, and then use this to find $\cos(\pi/16) = \frac{1}{2}\sqrt{2+\sqrt{2+\sqrt{2}}}$, and so on. In general

$$\cos(\pi/2^n) = \tfrac{1}{2}\sqrt{2+\sqrt{2+\sqrt{2+\ldots}}}$$

where there are $n-1$ square root signs.

Now that we have this we may use the cosine rule as we did earlier in the case $n = 3$. We find, for example, that

$$x_4 = \sqrt{2-\sqrt{2+\sqrt{2}}}.$$

This gives a new approximation $\pi = 3{\cdot}1214 \ldots$. Similarly

$$x_5 = \sqrt{2-\sqrt{2+\sqrt{2+\sqrt{2}}}}$$

from which $\pi \approx 3{\cdot}1365 \ldots$.

In general $\pi \approx P_n$ where

$$P_n = 2^{n-1}\sqrt{2-\sqrt{2+\sqrt{2+\ldots}}},$$

where there are $n-1$ square root signs, one minus sign and $n-3$ plus signs. I have calculated the following table.

n	P_n
2	2·82843
3	3·06147
4	3·12144
5	3·13655
6	3·14033
7	3·14128
8	3·14151

If we put

$$Q_n = \sqrt{2+\sqrt{2+\sqrt{2+\ldots}}}$$

where there are $n-2$ square root signs, then

$$\lim_{n\to\infty} Q_n = 2.$$

To prove this square both sides

$$Q_n^2 = 2 + Q_{n-1}.$$

If $Q_n \to Q$ as $n \to \infty$, then Q will satisfy

$$Q^2 = 2 + Q$$

and since $Q > 0$, the only solution is $Q = 2$. We then have

$$P_n = 2^{n-1}\sqrt{2-Q_n}$$

and the interesting point is that as $n \to \infty$, $2^{n-1} \to \infty$ and $\sqrt{2-Q_n} \to 0$ but their product tends to π.

I have been able to extend these results, starting with a hexagon. This gives similar results, with rather better approximations to π, as one might expect. I have also used circumscribing polygons. These give, again as we might expect, approximations which converge to π from above. If we start from an inscribed decagon, the side is $1/\phi$, where $\phi = 1{\cdot}618$ ... is the golden ratio. This leads to further approximations to π, each involving multiple square roots based on ϕ. I have also considered starting with a square and filling in the gaps with rectangles constructed by bisecting the sides of previous rectangles, either approaching the circle from inside or from outside. This leads to further approximations to π, but in this case the rate of convergence is very slow.

* * * * *

LETTERS TO THE EDITOR

Computers and Euclidean Theorems

I was interested to read the article about the new theorem in geometry produced by Wu Wenjun's program.† It states that "I have checked many such claims", and found them wanting.

However, Wu's work has now changed the situation considerably. Shang-Ching Chou has implemented Wu's program on computers at the University of Texas at Austin, and that has proved hundreds of significant theorems in Euclidean geometry. Chou's report indicates the algebraic method used for proving theorems, and a selection of the theorems which have been proved (Shang-Ching Chou, "A collection of geometry theorems proved mechanically", July 1986, *Technical Report* **50**, Institute for Computing Science, The University of Texas at Austin).

Wu's algebraic method of proof is markedly different from the traditional proofs based on geometric axioms; and the actual proofs are clearly "unfit for human consumption". Chou explains that 348 theorems were proved within a *total* period of 40 to 50 minutes, but 18 attempted proofs were stopped when they produced polynomials with more than 5000 terms.

Chou's report does not make clear how much work is required for presenting the theorems in a form acceptable to the program, taking care about non-degeneracy. And Chou seems not to claim that any new theorems have been produced by the program.

However, Wu's program is clearly very effective in proving non-trivial theorems in Euclidean geometry.

Garry J. Tee,
University of Auckland

* * * * *

And More on the Same

In the last edition of *Function*, Michael A.B. Deakin published an interesting article called "A Computer-Generated Theorem in Elementary Geometry" which deals with a problem of a plane intersecting a square pyramid. The question asked is: "Can the plane intersect the pyramid in such a way that a regular pentagon is formed?" There is one obvious way in which this is possible. The article goes on to describe a second way, where the pentagon lies outside the pyramid. The author is surprised by the result, and is pleased by the fact that the proof was found by using a computer; he believes that this is the first such geometrical result to be found in this way.

Much of the article gives a history of the proof and the author's personal account of re-creating the proof; this is very interesting to read. However, the account of an attempt to prove the result using a computer language called *Mathematica* does not provide

---

† See *Function*, *Vol.* 15, *Part* 1, p.8.

good reading; likewise the author's implied suggestion (as I read it!) that he could solve the equations faster and better than the computer provides a challenge.

In less than two hours (with several interruptions) I solved the equations, using a language called REDUCE, finding several solutions in the process. The speed says nothing about my abilities – anyone adept in using a Computer Algebra package such as *Mathematica*, *Maple*, or REDUCE could have done the same. I do not know what the people using *Mathematica* in Kyoto (as described by Michael Deakin) were doing; they were obviously expecting to crack a nut using a sledgehammer when they did not even know how to lift the sledgehammer!

The key to using Computer Algebra (or Symbolic Computation) in a problem like this is to use it interactively. This was done here.

There are four equations to be solved:

$$z1 = a^2(s-t)^2 + s^2 + t^2 - 4(v-1)^2$$

$$z2 = a^2(t-1)^2 + v^2 + (1-v-t)^2 - 4(v-1)^2$$

$$z3 = a^2(t-1)(s-t) + sv + t(1-v-t) - 2(1-v-t)(v-1)$$

$$z4 = a^2(s-t)^2 + s^2 - t^2 + 2(1-v-t)(v-1)$$

together with a fifth equation

$$z5 = s(1-t+v) - tv$$

where $z1 = \dots = z5 = 0$. I was told that it was most likely true that the first four equations implied the fifth, but no proof had been found. It turns out that this is not true: a solution to the first four equations exists which does not satisfy the fifth (see Case (c) below).

The procedure used is as follows: Type in the statements for $z1$ to $z5$. Keep track in your mind that each zi is zero. Put (after playing around with the equations for a while) $v = t + y$ and $t = 1 + m$. Then $z2 - z1 - z4 = 0$ gives $a^2 = 1 - 2y/m^2$ for $m \neq 0$. Check that $m = 0$ leads to a contradiction. Then s is found to be –(the coefficient in $z3$ of s^0) / (the coefficient in $z3$ of s^1). $z2$ gives an expression for m^2 in terms of m and powers of y. Make a substitution for m^2 (the computer does not realise that this says anything about m). Continue in this vein. Soon three possibilities emerge: (i) $y = 0$, (ii) $y = 1$ and (iii) m is some horrible ratio of third order polynomials in y. Now investigate these three cases in detail. The third case teaches us something about the power of Computer Algebra, because the expressions are ones which would probably make us want to give up on if we had to do the calculations by hand. However, the computer does such calculations with ease.

We find:

(a) $y = 0$, $m = -1 - \phi$ where $\phi^2 - \phi - 1 = 0$, $a^2 = 1$, $s = 1 + \phi$, $t = v = -\phi$,

(b) $y^2 - y - 1 = 0$, $a = \infty$, $m = 0$, $t = 1$, $s = t$, $v =$ arbitrary,

(c) $y^4 - 8y^3 + 19y^2 - 12y + 1 = 0$, $a = 0$.

Case (a) gives the two solutions discussed by Michael Deakin. The reader is left to think whether Cases (b) and (c) lead to any geometrically relevant solutions. In Case (c), s also is the ratio of two third order polynomials in y. However, $z5$ is not zero and this shows that the equations $z1 = 0$ to $z4 = 0$ do not imply that $z5 = 0$.

Colin B.G. McIntosh,
Monash University

* * * * *

It ain't necessarily so!

Recently I had a further letter from my friend Dai Fwls ap Rhyll, the little-known Welsh eccentric whose discoveries are quite revolutionary and have not yet received the attention they deserve.

Dr. Fwls has discovered that, in addition sums, the order of addition is important and not, as is widely taught, immaterial. He bases his work on "long and cross-tots" – those tabular arrays used by accountants and others, where two sums are made to balance and equality of the two is in fact used as a check on the accuracy of the calculation.

However, he has found a case where, quite clearly, this does not hold. Here it is.

$$
\begin{array}{l}
0 + 1 + 0 + 0 + 0 + 0 + \dots \\
-1 + 0 + 1 + 0 + 0 + 0 + \dots \\
+0 - 1 + 0 + 1 + 0 + 0 + \dots \\
+0 + 0 - 1 + 0 + 1 + 0 + \dots \\
+ \dots\dots\dots\dots \\
\dots\dots\dots\dots\dots
\end{array}
$$

If first we sum each row and then add up all the sub-totals so generated, the sum is clearly $+1$. However, if we first add by columns, the sum, equally clearly, is -1. To compound the paradox, if we add by diagonals in this direction $\nearrow$, each sub-total is zero and so the sum is zero.

The question is: which is correct?

Kim Dean,
Union College,
Windsor

* * * * *

HISTORY OF MATHEMATICS SECTION

EDITOR: M.A.B. DEAKIN

Facing Mecca

I travel to Indonesia quite frequently and on one such trip, an Australian who was travelling with me noticed that on the ceiling of each of the hotel rooms he had occupied there was an arrow. When he pointed this out to me I noticed such an arrow on the ceiling of my hotel room also. My friend suggested that the arrow might be an indication of the direction of Mecca: so that Muslim guests could face in the appropriate direction when they prayed.

This turned out to be so. Climbing onto a chair to investigate matters more closely, I read the words 'Arah Kiblat', which my dictionary told me mean 'direction' and 'towards Mecca' respectively.

It occurred to me that the determination of how to orient this arrow must be a matter of some mathematical complexity. It does in fact need spherical trigonometry for its solution and so the requirements of Islamic devotion might well, it seemed to me, have stimulated mathematical progress in this area. Such is indeed the case.

The three great monotheistic religions (Judaism, Christianity and Islam) each spring from a Middle Eastern tradition attributed by all three to the prophet Abraham (or Ibrahim). The events leading to the first codification of that tradition (the Jewish) are recorded in the book commonly referred to as the Old Testament. The events leading to a break-away movement within that tradition and the founding of Christianity are recorded in the book Christians call the New Testament. The ancestral tradition was not codified until much later, when the prophet Muḥammad (*c.* 570-632) founded Islam, whose sacred book is the Qur'an.

It seems that from the earliest times of the monotheistic tradition, the practice was adopted of praying with one's face toward some particular direction. In Arabic, this direction is known as the qibla (or ḳibla) – this term being the origin of the Indonesian word 'kiblat'.†

Originally the qibla seems to have been either geographic due east, or else (what is not quite the same) the direction, from day to day, of sunrise. It seems that it was King Solomon who first changed its direction.

† The letter *Q* in Arabic transliterations approximates our sound '*k*', not as we might think '*kw*'. In fact, even in English, this is so: "QUIT", pronounced *kwit*, has $Q \to k$, $U \to w$, $I \to i$, $T \to t$. Perfectly logical. In getting this wrong, we Australians are the world's worst. For the name of our national carrier QANTAS is pronounced with an initial *kw* sound when there is no *U* present!

"Then Solomon stood up before the altar of the Lord in the presence of the whole assembly of Israel and spread forth his hands to the heavens, and said 'O Lord ... if thy people go out to battle against their enemy, by whatever way thou shalt send them, and they pray to the Lord in the direction of the city which thou hast chosen and the house which I have built for thy name, then hear thou in the heavens their prayer and their supplication, and uphold their cause'."

1 Kings viii 22-23, 44-45.

Thus the Jews, and in all probability other monotheists, came to pray facing Jerusalem. Later, probably in our year† 624, Muḥammad altered the direction of the qibla, which came to be directed toward Mecca. The change is recorded in the Qur'an (Chapter ii, Verses 143-148, though these numbers may vary between editions).

So from the earliest times of Islam, the qibla has been towards Mecca. As Islam spread, the determination of the qibla became a problem complicated by the need to take account of the curvature of the earth's surface. Look at Figure 1. M is the position of Mecca and X is another point on the earth's surface. It is shown to the north and to

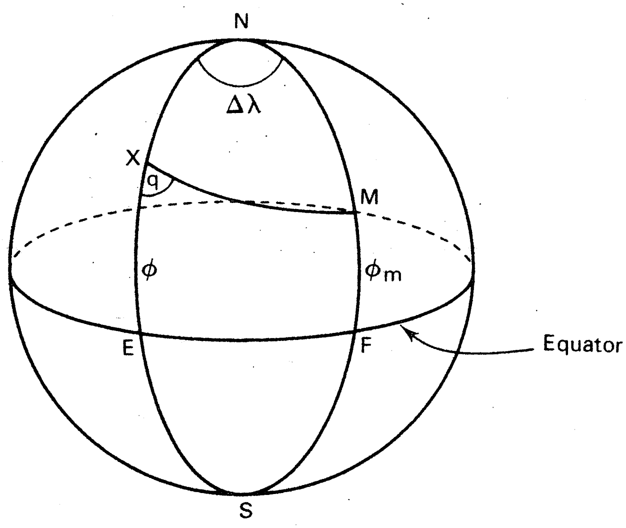


Figure 1

† The Muslim calendar dates from the *Hidjra*, or migration of the prophet from Mecca to Medina, on 16/7/622. The qibla was altered some 17 months after this event.

the west of Mecca because these were the directons in which Islam spread in its early years.

The arc XE of the meridian through X subtends an angle ϕ at the centre of the earth. ϕ is the latitude of X. Similarly, ϕ_m, the angular measure of the arc MF, is the latitude of Mecca. The difference in longitude between the two places is $\Delta\lambda$, the angle XNM on the diagram, where N is the north pole. The arc XM is the shortest distance from X to M. It lies along a "great circle" – a circle whose radius is that of the earth.

[At first I wasn't sure that the direction was necessarily measured along such a great circle. Lines of latitude, for example, are *not* (apart from the equator) great circles. Thus if one travels (say) due east of Mecca (along a line of latitude), should one face back west, or somewhat north of west, along the great circle? (There is a point in the eastern Pacific Ocean about a third or so of the way from Hawaii to the Mexican coast where the great circle to Mecca passes through the north pole.) I asked a number of Muslim mathematicians about this and eventually received the answer that the qibla is indeed taken along great circular arcs.†]

The angle q, called in Arabic *inḥiraf al-qibla*, is then to be determined from a knowledge of ϕ, ϕ_m and $\Delta\lambda$. This is a problem in spherical trigonometry. Spherical trigonometry was described in *Function, Vol.* 6, *Part* 5, pp. 8-12, but there is a lot more to the subject than could be mentioned in that brief article. From one of the formulae of spherical trigonometry, not actually given in that article but deducible (with work) from others that are, we may derive the result:

$$\cot q = \{\sin\phi\cos\Delta\lambda - \cos\phi\tan\phi_m\}/\sin\Delta\lambda \qquad (1)$$

which enables the qibla to be determined.

The question thus arises as to when this formula or some equivalent of it was first derived. In the west, the person who most studied such questions was the late Karl Schoy whose works unfortunately are mostly in German and in books and journals to which I don't have ready access. However, he summarised much of this work in the article he wrote (in English) on *Ḳibla* in the *Encyclopaedia of Islam.*

The first name he mentions is that of al-Battānī, who lived in the late ninth and early tenth centuries in what is now Iraq. al-Battānī, as mentioned in *Function, Vol.* 6, *Part* 5, p. 12, is credited by some authors with the discovery of the cosine rule of spherical trigonometry. The cosine rule is the basic rule of spherical trigonometry from which all the others may be derived – including Equation (1) above. However, it seems that al-Battānī did not go on to derive Equation (1).

He did discuss the determination of the qibla but gave only an approximate construction. Here it is. Look at Figure 2; it represents a horizontal circle on which the four cardinal points are indicated.

† That not all Islamic mathematicians know this need not surprise us. How many Christian mathematicians know how the date of Easter is determined? [See *Function, Vol. 9, Part* 3, pp. 10-12.]

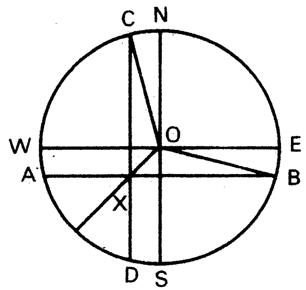


Figure 2

First mark off $\angle EOB$ equal to $\phi - \phi_m$, next mark off $\angle NOC$ equal to $\Delta\lambda$, both as shown. Now draw AB parallel to WE and CD parallel to NS. Suppose these lines meet at X. Then

$$q \simeq \angle SOX. \tag{2}$$

As an aid to further analysis, write $\Delta\phi = \phi - \phi_m$. It may then readily be shown that al-Battānī's approximation is

$$\cot q \simeq \frac{\sin \Delta\phi}{\sin \Delta\lambda}. \tag{3}$$

Exact is Equation (1), which may be approximated for small $\Delta\phi$, $\Delta\lambda$ by

$$\cot q \simeq \left[\frac{\sin \Delta\phi}{\sin \Delta\lambda}\right] \sec \phi_m. \tag{4}$$

ϕ_m, the latitude of Mecca, is 21°27′, so $\sec \phi_m \simeq 1{\cdot}07$. Thus al-Battānī's approximation is a reasonably good one. In fact, it can work out to give quite good agreement in certain localities. Schoy demonstrated for example that it is extremely accurate for Cairo.

A somewhat better approximation was produced by the tenth century Egyptian astronomer Ibn Yunus, who is also credited by some with the discovery of both the sine and cosine laws of spherical trigonometry. Exact solutions, however, seem initially to have involved various graphical or mechanical devices. Schoy mentions two early ones, due respectively to Abu 'l-Wafā' (who died in 998) and to al-Faḍl b. Hatim al-Nairīzī, who died in 922 or 923. However, the latter's calculation of q for Baghdad was very badly in error.

Schoy goes on to describe a geometrical construction produced by Ḥasan b. al-Ḥusain b. al-Haiṯham, a somewhat later astronomer, who died in 1039. It is rather involved and so will not be given here, but it is exact. More recently, details of an earlier exact construction have come to light. They form indeed the subject of the very first paper in the very first issue of the journal *Historia Mathematica.* This method is due to the ninth century Baghdad astronomer Habash al-Ḥāsib, though we know it through the writings of the somewhat later Central Asian mathematician Abū al-Rayḥān Muḥammad b. Aḥmed al-Bīrūnī (born 973). Again its details are complicated and will be omitted.

Later, these complicated constructions gave way to tables. One such was produced around 1365, by Shams al-Dīn Muḥammad ibn Muḥammad al-Khalīlī, a timekeeper of Damascus. A lengthy account of this is given by David King in Volume 34 of the *Journal of Near Eastern Studies*.

Tables are indeed what modern Muslims use. In Islamic countries, the direction of the qibla is often indicated, as in the hotel rooms in Indonesia, and mosques are normally aligned correctly.† However, when travelling, a devout Muslim will normally carry a small magnetic compass and a card incorporating a brief qibla table. I owe this information to an Iranian mathematician whom I met at a recent international conference.

It would seem, therefore, that it was the Islamic mathematicians who for many hundreds of years led the world in the study of spherical trigonometry. In the West, there is little to be found between the days of Ptolemy (second century) and those of Regiomontanus (fifteenth century). And one of the main incentives to the Arab tradition was the determination of the qibla.

Finally, we may note that there is one place on earth, other than Mecca itself, where the qibla is undefined. It is the point precisely antipodal to Mecca, very near Vanavana atoll in the Tuamoto Archipelago, but quite close to the more notorious Mururoa. As an exercise, show that for this point Equation (1) yields $\cot q = 0/0$. Can you show from Equation (1) that this is the only such point?

* * * * *

Snow on Mathematics

> Finally, we should agree that mathematical creation is one of the great triumphs of the species since we climbed our way out of the caves, and one of the reasons, perhaps the chief reason, why in the last 400 years, we have climbed so quickly. It does not need stating that mathematics is the basis of our scientific civilization. But I would also state that mathematics is a token and a symbol of the human achievement in its own right. There are many tests you can apply to a society's education. But one of them is – does it bring out a number of creative mathematicians who, by world standards, can hold their own? This is, of course, not the only test. But it is a harsh, objective and essential one. If a society's education does not reach it, then there is something wrong with the society or with the education. A society which does not choose to encourage excellence, and this particular kind of excellence, won't be a decent society for long.
>
> C.P. Snow

† The last vestige of the qibla in Christian worship would seem to be the choice of the east as the direction of the naves of cathedrals.

COMPUTER SECTION

EDITOR: R.T. WORLEY

Some Computing Methods

The sequence 0, 1, 1, 2, 3, 5, 8, 11,... of Fibonacci numbers F_n, in which each number apart from the first two is the sum of the previous two numbers, is useful for demonstrating some techniques. The Fibonacci numbers are formally defined by

$$F_0 = 0, \quad F_1 = 1$$
$$F_n = F_{n-1} + F_{n-2} \qquad \text{for } n \geq 2, \tag{1}$$

and are known to have many properties. Some of these are mentioned in the article by Garnet J. Greenbury in *Function, Vol.* 14, *No.* 4, while methods of computing the Fibonacci numbers are discussed in *Function, Vol.* 14, *No. 5*. I have recently become aware of better methods of computing the Fibonacci numbers.

The simplest method of calculating F_n is to calculate in turn $F_2, F_3, \ldots F_{n-1}, F_n$ using the defining relation (1). However it is certainly possible to calculate F_n by evaluating only some of $F_2,\ldots,F_{n-1}$ before F_n. For example, consider the formula

$$F_{m+k} = F_{m-1}F_k + F_mF_{k+1} \tag{2}$$

of Greenbury's article. Using this formula we have, for example,

$$F_{27} = F_{13}F_{13} + F_{14}F_{14}. \tag{3}$$

Hence we clearly only need to evaluate $F_2,\ldots,F_{14}$ before we can calculate F_{27} using (3) – we certainly don't need to calculate any of $F_{15},\ldots,F_{26}$. We could be even more economical, using (2) again to observe that

$$F_{13} = F_6F_6 + F_7F_7, \tag{4}$$

$$F_{14} = F_6F_7 + F_7F_8, \tag{5}$$

and so we could calculate F_{27} by calculating only $F_2,F_3,\ldots,F_8,F_{13}$ and F_{14}.

Based on these ideas we can write a program to evaluate F_n using (2) in the way illustrated, where $k = m$-1 or m. Using these values of k, formula (2) becomes

$$F_{2m-1} = F^2_{m-1} + F^2_m \tag{6}$$
$$F_{2m} = F_m(F_{m-1} + F_{m+1}) = F_m(2F_{m-1} + F_m). \tag{7}$$

We start by setting up an array $F(0)$, $F(1)$,...$F(\text{n})$ with $F(0) = 0$, $F(1) = 1$, and $F(2) = \ldots = F(\text{n}) = -1$. The '-1' indicates that the true value has not yet been worked out. Then we have a function Fib(i) which does the following.

(i) If $F(i) \geq 0$, Fib(i) returns the value $F(i)$.

(ii) If $F(i) = -1$, then we use the correct one of equations (6), (7) to calculate the i-th Fibonacci number, store it in $F(i)$, and return the value $F(i)$.

This descent method was programmed, and the actual number of $F_2, \ldots, F_{n-1}$ evaluated in calculating F_n are given in Table 1. In addition, the total number of times values of F_n were used was counted, and these values are also included in the table.

Because of the way the Fibonacci numbers grow so rapidly, they quickly exceed the bound on integers that a computer normally handles completely accurately. This means that in most programming languages on most computers one cannot calculate F_n for $n > 46$. One can go a little further with Turbo Pascal on the IBM PC, because it has a 'comp' integer variable type which copes with up to $n = 92$ (although only values up to $n = 87$ can be printed out accurately). In some languages specialised for long integers, one has almost no limit. I used UBASIC, a public domain version of BASIC for the IBM PC which handles Fibonacci numbers up to $n = 12492$.

The above method is similar to calculating a^{27} by

$$\begin{aligned} a^{27} &= a^{13}a^{14} \\ &= (a^6a^7)(a^7a^7) \\ &= ((a^3a^3)(a^3a^4))((a^3a^4)(a^3a^4)) \qquad \text{(etc).} \end{aligned}$$

However, powers are usually calculated by the method known as 'binary powering', which writes

$$a^{27} = a^{16+8+2+1}$$

i.e.,

$$a^{27} = a^{16}a^8a^2a^1.$$

The powers on the right are easily calculated by repeated squaring – we actually calculate a^4 as well, because it is needed for a^8, although we make no other use of it.

$$\begin{aligned} a^2 &= (a^1)^2 \\ a^4 &= (a^2)^2 \\ a^8 &= (a^4)^2 \\ a^{16} &= (a^8)^2. \end{aligned}$$

The binary powering method builds up from a^1 to a^{27}, whereas the first method worked down from a^{27} to a^1.

Because binary powering is normally used in place of the descent method for powering, this approach should be investigated. It turns out that we can find an analogue of the binary powering method to calculate the Fibonacci numbers. However there are a couple of modifications. If we look at the formula (7) above we see that to calculate F_{2m} we need F_{m-1} as well as F_m. Thus, for example, to calculate F_8 we need F_3 as well as F_4. Thus we need both formulas (6) and (7), using F_1 and F_2 to calculate F_3 and F_4. We then use these to calculate F_7 and F_8, and use these to calculate F_{15} and F_{16}, (etc.). It now remains to see how we can use these terms to build up to F_{27}. To do this we make use of the formulae

$$F_{m+k} = F_{m-1}F_k + F_mF_k$$

$$F_{m+k+1} = F_mF_k + F_{m+1}F_{k+1} = F_mF_k + (F_{m-1} + F_m)F_{k+1},$$

which express F_{m+k} and F_{m+k+1} in terms of F_k and F_{k+1}. Starting with k=0, m=1 we generate F_1 and F_2. Now taking k=1, m=2 we then generate F_3 and F_4. Now we take k=3, m=8 to get F_{11} and F_{12}, and finally taking k=11, m=16 we generate F_{27} and F_{28}. (The values for k are the sums 0, 0+1, 0+1+2, 0+1+2+8 and the corresponding value of m is the power of 2 required to get the next value of k.)

A program was written to calculate the Fibonacci numbers using this method. For comparison with the descent method, the table gives the number of different values of the Fibonacci numbers calculated. Very little can be deduced from the table, apart from an obvious formula or two. Although it appears that the descent method uses less different values of the Fibonacci numbers, this is not always true and the descent method can use nearly 50% more different values for very large n of the form $2^k + 2$.

Although you may think there is little point in calculating large Fibonacci numbers, that is not what this article is really about. It is the two methods used that are of importance. The first method, which I called the descent method, is an example of what in computing is called the 'divide and conquer' method. The problem of calculating F_n is divided into two smaller problems (calculating F_m and F_{m-1} where m is approximately $n/2$) which are easier to solve. These problems themselves are divided, and so on. This technique is the basis of a number of algorithms, including the fast sorting procedure known as quicksort. Likewise the method based on binary powering is an important technique.

	descent method		**binary method**
n	# distinct F_i	# total F_i	# F_i
130	15	31	18
258	18	37	20
514	21	43	22
1026	24	49	24
2050	27	55	26
123	13	27	24
456	18	37	24
789	23	47	28
901	23	47	28
1234	24	49	30
4567	28	57	40
7890	32	65	40
9012	32	65	38
12345	35	71	38

Table 1

The following is a sample program using Turbo Pascal to calculate the Fibonacci numbers by the descent method. Some of the details are specific to Turbo Pascal. For example, the type 'comp'. If using another version of Pascal it may be better to use the 'real' type, in which case the constant maxfib will need to be altered to a smaller number to cope with the accuracy available with that type .

```
(* program to calculate fibonacci numbers by divide and conquer *)
(* R.T. Worley *)
{$n+}
const maxfib=87;
var n,i:integer;
var f:array [0..maxfib] of comp;

function fib(m:integer):comp;
var k:integer;
begin
if (f[m] >=0) then fib := f[m] else begin
      if odd(m) then begin
            k:=m div 2;
            f[m] := sqr(fib(k)) + sqr(fib(k+1));
            fib := f[m]
            end
      else begin
            k := m div 2;
            f[m] := fib(k) *(2*fib(k-1) + fib(k));
            fib := f[m]
            end
      end
end;

begin (* main *)
      for i:=2 to maxfib do f[i]:=-1;
      f[0]:=0; f[1]:=1;
      while (true) do begin
            write('which fibonacci number? (0<=n<=',maxfib:2,') :');
            read (n);
            if (n<0) or (n>maxfib) then writeln('illegal input value') else
            writeln ('fib(',n,') = ',fib(n):26);
            end
```

* * * * *

More from C.P. Snow

Mathematical excellence is difficult to handle, administratively and politically, simply because very few people possess it. Life would be simpler and tidier if no one possessed it. Then these tiresome arguments about some special arrangement could not enter at all. Yes, life would be simpler and tidier, it would have also lost intellectual glories. Concern for this kind of excellence is not the only value in education. Quantitatively it does not count, by the side of the task of educating as well and as justly as we can millions of children. At the most, this concerns only a few thousands. It is not the only value. Yet I am certain that if we neglect it, we shall show, whatever our motives, that we do not know what educational values are.

C.P. Snow

PROBLEMS AND SOLUTIONS

EDITOR: H. LAUSCH

Solutions

Solutions to Problem 14.5.3 have been received from John Barton (North Carlton, Victoria) and from Francisco Bellot-Rosado (I.B. "Emilio Ferrari", Valladolid, Spain). The latter also offers a variation of John Barton's solution to Problem 14.5.6 given in the last issue.

Problem 14.5.3. Solve the simultaneous equations in the unknowns x,y,z:

$$x^2 - yz = a^2$$

$$y^2 - zx = b^2$$

$$z^2 - xy = c^2.$$

Solution (Francisco Bellot-Rosado). After multiplying the first equation by y, the second by z, and the third by x, and adding the resulting equations, we obtain

$$c^2x + a^2y + b^2z = 0. \tag{1}$$

Similarly, after multiplying the first equation by z, the second by x and the third by y, and adding, we obtain

$$b^2x + c^2y + a^2z = 0. \tag{2}$$

Considering (1) and (2) together as a system of two linear equations in the three unknowns x,y,z and treating it by any method available, one ends up with

$$\frac{x}{a^4-b^2c^2} = \frac{y}{b^4-a^2c^2} = \frac{z}{c^4-a^2b^2} = \lambda, \text{ say,}$$

hence

$$x = (a^4-b^2c^2)\lambda,\ y = (b^4-a^2c^2)\lambda,\ z = (c^4-a^2b^2)\lambda. \tag{3}$$

Substitution of these values into the third equation of the problem gives

$\lambda^2[(c^4-a^2b^2)^2 - (a^4-b^2c^2)(b^4-c^2a^2)] = c^2$, from which it follows that

$$\lambda^2 = \frac{1}{a^6+b^6+c^6-3a^2b^2c^2}.$$

Going back with this into (3) provides the values of x,y,z.

John Barton's solution is, in principle, along similar lines, but uses the powerful resultant or (Sylvester) eliminant method that can be found in classical higher algebra texts.

Problem 14.5.6. Let x,y,z be three non-zero real numbers which are distinct from each other. If $x + y + z = 0$, what is the product

$$\left\{ \frac{y-z}{x} + \frac{z-x}{y} + \frac{x-y}{z} \right\} \cdot \left\{ \frac{x}{y-z} + \frac{y}{z-x} + \frac{z}{x-y} \right\} ?$$

Solution. If we set $a = \frac{y-z}{x}$, $b = \frac{z-x}{y}$, $c = \frac{x-y}{z}$, then the product in question is

$$(a+b+c)\left[\frac{1}{a} + \frac{1}{b} + \frac{1}{c} \right] = 3 + \frac{b+c}{a} + \frac{c+a}{b} + \frac{a+b}{c} .$$

Now consider $\frac{b+c}{a} = \frac{z}{xy}(-x-y-z+2z) = \frac{2z^2}{xy}$, because $x + y + z = 0$. Similar expressions are obtained for $\frac{c+a}{b}$ and $\frac{a+b}{b}$. Hence

$$\frac{b+c}{a} + \frac{c+a}{b} + \frac{a+b}{c} = \frac{2z^2}{xy} + \frac{2x^2}{yz} + \frac{2y^2}{zx} = \frac{2}{xyz}(x^3+y^3+z^3).$$

Again we use the data that $x + y + z = 0$, this time for concluding that $x^3 + y^3 + z^3 = 3xyz$, and therefore

$$\frac{b+c}{a} + \frac{c+a}{b} + \frac{a+b}{c} = 6,$$

so that the answer is $3 + 6 = 9$.

Francisco Bellot-Rosado remarks that this problem is included, for example, in the (not so old as 1878) book: Krechmar, A problem book in algebra, *Mir 1974 (Ex. 2.18).*

Problems

Juan Bosco Romero Márquez (I.B. "Isabel de Castilla"), Avila, in Old Castile, Spain, offers this problem to our readers. ¡Muchas gracias!

Problem 15.2.1. Let $f(n) = 1 + 2 + 3 + \dots + n$.

Evaluate:

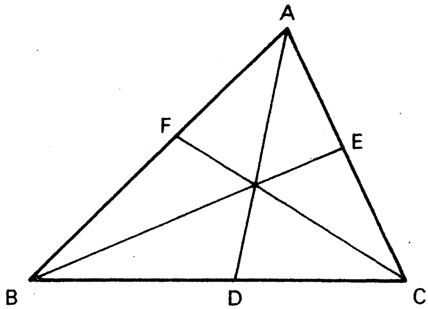
(a) $(-1)^{f(n)}$,

(b) $i^{f(n)}$, where i denotes as usual a square root of -1,

(c) $\omega_p^{f(n)}$, where ω_p is a p-th root of unity, i.e. a complex number whose p-th power equals 1.

[Note that (a) and (b) are the special cases of (c) in which $p = 2, 4$ respectively.]

Problem 15.2.2 (K.R.S. Sastry, Addis Ababa). AD, BE, CF are the angle bisectors of triangle ABC (see figure). Determine all triangles ABC in which DF bisects angle BFC. What are the angles of the triangle ABC if $\angle AFC = 2 \cdot \angle BAC$?



Problem 15.2.3 (K.R.S. Sastry, Addis Ababa). A triangle is called *self-median* if its sides are proportional to the medians. Let AD, BE be medians and G the centroid of triangle ABC. Prove that this triangle is self-median if $\angle DGC = \angle BAC$ and $\angle CGE = \angle ABC$.

Year Twelve International

Here are a few more problems from a British sixth form course.

Problem 15.2.4. Someone writes n letters and writes the corresponding addresses on n envelopes. If U_n denotes the number of ways of placing all the letters in the wrong envelopes, obtain a relation between U_n, U_{n-1} and U_{n-2}. Compute U_4 and find U_n.

Problem 15.2.5. Take n points on a circle and join them up in all possible ways. Into how many regions is the circle divided by the joins?

Problem 15.2.6. A point lies inside an equilateral triangle and has distances x,y,z from the three sides respectively; h is the altitude of the triangle. Prove that $x + y + z = h$. Can this result be generalized in any way?

Problem 15.2.7. A triangle T is drawn on a square grid (graph paper) with vertices at points of the grid, but no other grid-points is in or on it. Show that the area of T is $\frac{1}{2}$.

Problem 15.2.8. Let S be a set of n points in the plane such that any two of them are at most 1 unit apart. Find the radius of the smallest circular disc which will cover S.

Mathematical Olympiads

The Australian Mathematical Olympiad, competed in by 100 secondary-school students from across Australia, took place on February 12th and 13th this year. On each day competitors had a four-hour paper, each consisting of four questions, to tackle. Here are the papers. Can you do the questions? Send me your comments and solutions.

Paper 1

1. Let $ABCD$ be a convex quadrilateral. Prove that if g is the greatest and h is the least of the distances AB, AC, AD, BC, BD, CD, then $g \geq h\sqrt{2}$.

2. Let M_n be the least common multiple of the numbers $1, 2, 3, \ldots, n$. E.g., $M_1 = 1$, $M_2 = 2$, $M_3 = 6$, $M_4 = 12$, $M_5 = 60$, $M_6 = 60$. For which positive integers n does $M_{n-1} = M_n$ hold? Prove your claim.

3. Let A, B, C be three points in the x-y-plane and X, Y, Z the midpoints of the line segments AB, BC, AC, respectively. Furthermore, let P be a point on the line BC so that $\angle CPZ = \angle YXZ$. Prove that AP and BC intersect in a right angle.

4. Show that there is precisely one function f that is defined for all non-zero reals, satisfying:

(i) $f(x) = xf\left(\frac{1}{x}\right)$, for all non-zero reals x;

(ii) $f(x) + f(y) = 1 + f(x+y)$ for each pair (x,y) of non-zero reals where $x \neq -y$.

Paper 2

5. Let $P_1, P_2, \ldots, P_n$ be n different points in a given plane such that each triangle $P_iP_jP_k$ $(i \neq j \neq k \neq i)$ has an area not greater than 1. Prove that there exists a triangle Δ in this plane such that

a) Δ has an area not greater than 4; and

b) each of the points $P_1, P_2, \ldots, P_n$ lies in the interior or on the boundary of Δ.

6. For each positive integer n, let

$$f(n) = \frac{1}{\sqrt[3]{n^2+2n+1} + \sqrt[3]{n^2-1} + \sqrt[3]{n^2-2n+1}}.$$

Determine the value of $f(1) + f(3) + f(5) + \ldots + f(999997) + f(999999)$.

7. In triangle ABC, let M be the midpoint of BC, and let P and R be points on AB and AC, respectively. Let Q be the intersection of AM and PR.
Prove: if Q is the midpoint of PR, then PR is parallel to BC.

8. Find a sequence $a_0, a_1, a_2, \ldots$ whose elements are positive and such that $a_0 = 1$ and $a_n - a_{n+1} = a_{n+2}$ for $n = 0,1,2,\ldots$. Show that there is only one such sequence.

On the basis of their performances, 27 Australian students were selected to participate in this year's Asian Pacific Mathematical Olympiad, the third since its inception in 1989. They will have to compete with students from twelve other countries on the Pacific Rim.

They are (their grade or year appears in parentheses):

Geoffrey Brent	(10),	Canberra Grammar School,	ACT
Adrian Banner	(11),	Sydney Grammar School,	NSW
Kendall Bein	(12),	James Ruse High School,	NSW
Andrew Berthelson	(12),	James Ruse High School,	NSW
Tom Brennan	(12),	Knox Grammar School,	NSW
Peter Cotton	(12),	Newington College,	NSW
Anthony Douglas	(12),	Knox Grammar School,	NSW
Avery Fung	(12),	Randwick Boys High School,	NSW
Jonathan Heaney	(12),	Knox Grammar School,	NSW
Anthony Henderson	(10),	Sydney Grammar School,	NSW
Luke Kameron	(12),	Knox Grammar School,	NSW
Andrew Usher	(12),	Sydney Grammar School,	NSW
Eric Willigers	(12),	Colo High School,	NSW
Weiben Yuan	(12),	Cabramatta High School,	NSW
Robert McCahill	(12),	Ignatius Park School,	Qld
Meng Tan	(12),	Brisbane Grammar School,	Qld
Brian Ng	(12),	Prince Alfred College,	SA
Justin Sawon	(12),	Heathfield High School,	SA
Kingsley Storer	(12),	Prince Alfred College	SA
Martin Roberts	(12),	Rosny College,	Tas
Angelo Di Pasquale	(12),	Eltham College,	Vic
Lawrence Ip	(11),	MCEGS,	Vic
Joanna Masel	(12),	Methodist Ladies' College,	Vic
Brett Pearce	(11),	St. Michael's Grammar School,	Vic
Sam Watkins	(9),	MCEGS,	Vic
Stuart Sellner	(12),	Rossmoyne Senior High School,	WA
Robert Yuncken	(12),	Christchurch Grammar School,	WA

* * * * *

Abstraction and Mathematics

All the sciences which have for their end investigations concerning order and measure, are related to mathematics, it being of small importance whether this measure be sought in numbers, forms, stars, sounds or any other objects; that accordingly there ought to exist a general science which should explain all that can be known about order and measure, considered independently of any application to a particular subject, and that indeed this science has its own proper name, mathematics. And a proof that it far surpasses in facility and importance the sciences which depend upon it is that it embraces at once all the objects to which these are devoted and a great many more besides.

R. Descartes

Registered for posting as a periodical – "Category B"
ISSN 0313 – 6825

* * * * *

Published by Monash University Mathematics Department