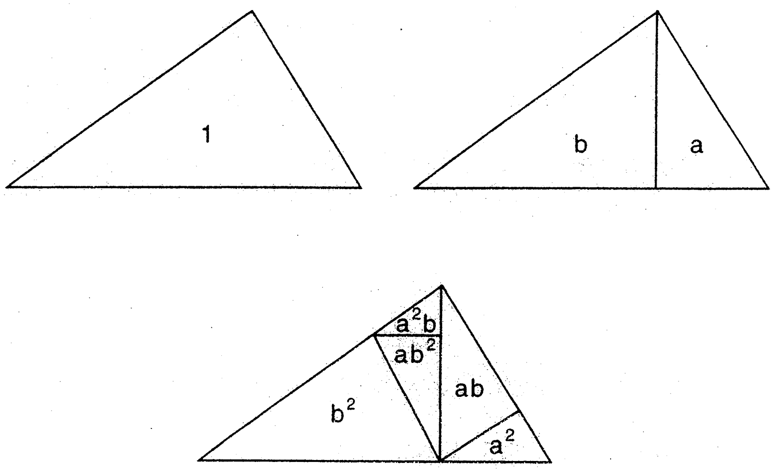
Function

A School Mathematics Magazine

Volume 20 Part 3 June 1996



Mathematics Department - Monash University

Reg. by Aust. Post Publ. No. PP338685/0015

Published by Monash University Mathematics Department

EDITORIAL

We welcome our readers to this issue of *Function.*

The figure on the front cover relates to a problem published in *Function* several years ago. It is about a process of subdividing a triangle into smaller similar triangles, and reveals a surprising connection between this problem and a certain type of series.

We include another article by our reader Benito Hernández-Bermejo. It explains an efficient procedure that the Babylonians used to approximate a square root, which shows the highly advanced mathematical development of this ancient people.

We are pleased to also include an article by one of our youngest readers. Mark Eid is a Year 9 student; he shows some very nice formulae for generating right-angled triangles with integer side lengths.

Michael Deakin recommends an interesting book in which the author, Ido Yavetz, presents the mathematical thinking of Oliver Heaviside, put in the context of the knowledge existing at that mathematician's time.

We have an area-related topic in the *History* column. It deals with the history of the question: *Given the side lengths of a polygon (but not its angles), what is the configuration that maximises the area?*

Have you ever been puzzled by the answers a mathematics program gives as solutions to polynomial equations? The *Computers and Computing* column will help you to understand how computer algebra systems solve these equations.

As usual, there are several new problems for your entertainment; for brighter minds we also include the problems set for this year's Asian Pacific Mathematics Olympiad. Solutions to the problems, as well as articles and comments of any kind, are always welcome from our readers.

We hope you enjoy this issue of *Function*!

* * * * *

THE FRONT COVER

Subdivision of Triangles

Peter Grossman, Monash University

The illustration on the front cover depicts an intriguing problem from the August 1992 issue of *Function* for which we have not previously published a solution. The problem was originally stated as follows:

> The figure below (Figure 1) shows two copies of the same right-angled triangle, whose area in one copy is divided by the altitude in the ratio $a : b$ $(a < b)$, the whole being held to have unit area. The three portions thus have areas $1, a, b$.

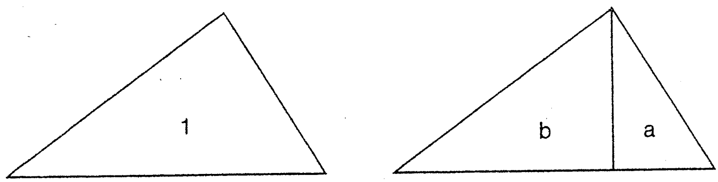


Figure 1

> Further copies of the triangle are to be divided into constituent right-angled triangles, by repeated insertions of altitudes in such a way that no two subdivisions have the same area, anywhere over the full set of copies. How many copies in all will this allow?

The front cover shows the two copies of the triangle given in Figure 1, together with a third copy which has been subdivided according to the rules given in the problem; the two smaller triangles introduced in the second copy have each been subdivided, and one of the resulting triangles has in turn been subdivided in order to avoid the occurrence of two subdivisions with the same area. It is easy to see that no two subdivisions have the same area, anywhere over the full set of copies (except for some special choices of a and b which we will ignore). Thus at least three copies in all can be produced. We will now show that no further copies are possible.

In order to do this, it is useful to define the *level* of each subdivision in the construction. We will say that the original triangle has level zero, and a triangle produced by subdividing a triangle at level j has level $j+1$. Thus the second of the three copies on the front cover contains two triangles at level 1, and the third copy contains three triangles at level 2 and two at level 3.

We can list the possible areas of triangles at any given level. At level 1, the possible areas are a and b. At level 2, they are a^2 and ab if the level 1 triangle with area a is subdivided, and ab and b^2 if the level 1 triangle with area b is subdivided. (You can easily check this by verifying in each case that the two areas are in the ratio $a : b$ and their sum is the area of the level 1 triangle from which they are produced by subdivision.)

A similar argument in the general case shows that the possible areas of level j triangles are:

$$a^j, a^{j-1}b, a^{j-2}b^2, \ldots, b^j.$$

In other words, the area of a level j triangle must be $a^{j-k}b^k$ for some value of k between 0 and j. This follows from the fact that a triangle with area $a^{j-1-k}b^k$ from level $j-1$ can be subdivided into two triangles with areas $a^{j-k}b^k$ and $a^{j-1-k}b^{k+1}$ at level j. A formal proof proceeds by induction on j; we leave it to you to supply the details.

Now, suppose we have a set of copies of the original triangle which have been subdivided according to the rules given in the problem. Let m_{ij} denote the number of subdivisions of copy i at level j. (In the situation depicted on the front cover, for example, $m_{1,0} = 1$, $m_{2,1} = 2$, $m_{3,2} = 3$, $m_{3,3} = 2$, and $m_{ij} = 0$ for all other values of i and j.) Let d denote the maximum level of all the triangles. Then, for any given copy, i, the m_{ij}'s must satisfy the equation

$$\sum_{j=0}^{d} 2^{-j} m_{ij} = 1. \tag{1}$$

To prove equation (1), notice firstly that for $i = 1$, when the original triangle has not been divided, we have $m_{1,0} = 1$ and $d = 0$, so the sum is just a single term equal to 1. Thus equation (1) is true in this case. We now proceed inductively, noting that as each successive subdivision is made, say in copy i by subdividing a triangle at level j, m_{ij} decreases by 1 and $m_{i,j+1}$ increases by 2, so the value of the sum is unchanged.

We also observe that for each level j the m_{ij}'s satisfy the inequality

$$\sum_{i=1}^{n} m_{ij} \leq j+1 \tag{2}$$

where n is the number of copies of the original triangle. Inequality (2) follows from the fact that only $j+1$ different areas are possible at level j, and each subdivision must have a different area.

(1) and (2) impose quite severe restrictions on the possible values of the m_{ij}'s, and hence ultimately on the possible values of n. On the one hand, equation (1) tells us that the m_{ij}'s must grow rapidly as j increases, in order to compensate for the rapid rate of decrease of 2^{-j}. On the other hand, inequality (2) prevents the m_{ij}'s from growing very fast. In order to make the argument precise, we need to investigate the double sum

$$S = \sum_{i=1}^{n} \sum_{j=0}^{d} 2^{-j} m_{ij}.$$

Using equation (1), we obtain

$$S = \sum_{i=1}^{n} 1 = n.$$

If we reverse the order of the summations in the definition of S and use inequality (2), we obtain

$$\begin{aligned} S &= \sum_{j=0}^{d} \sum_{i=1}^{n} 2^{-j} m_{ij} \\ &= \sum_{j=0}^{d} 2^{-j} \sum_{i=1}^{n} m_{ij} \\ &\leq \sum_{j=0}^{d} 2^{-j}(j+1). \end{aligned}$$

The sum in the last line may not be familiar to you; it is an example of an *arithmetico-geometric series*. The form taken by its general term suggests that its value can be found by differentiating the formula for a suitably chosen geometric series with respect to its common ratio:

$$\sum_{j=0}^{d} r^{j+1} = \frac{r(1-r^{d+1})}{1-r} = \frac{r-r^{d+2}}{1-r}.$$

Differentiating with respect to r:

$$\begin{aligned} \sum_{j=0}^{d} (j+1) r^j &= \frac{(1-r)[1-(d+2)r^{d+1}] - (r-r^{d+2})(-1)}{(1-r)^2} \\ &= \frac{(d+1)r^{d+2} - (d+2)r^{d+1} + 1}{(1-r)^2}. \end{aligned}$$

Putting $r = 2^{-1}$, we obtain

$$\begin{aligned}\sum_{j=0}^{d}(j+1)\times 2^{-j} &= \frac{(d+1)\times 2^{-d-2}-(d+2)\times 2^{-d-1}+1}{\frac{1}{4}} \\ &= 4\{2^{-d-2}[d+1-(d+2)\times 2]+1\} \\ &= 4-2^{-d}(d+3).\end{aligned}$$

Thus $\sum_{j=0}^{d} 2^{-j}(j+1) < 4$, so $S < 4$ and hence $n < 4$. We conclude that three is the maximum possible number of copies of the triangle that can be subdivided in the manner specified.

* * * * *

All horses are of the same colour

This can be proved by induction as follows.

Let n be the number of horses.

1. Base step

If $n = 1$, there is only one horse, so the statement is true.

2. Inductive step

Assume the statement is true for $n = k$. We need to prove that it is then also true for $n = k + 1$.

If $n = k + 1$, we take away one of the horses; then by the inductive hypothesis, the remaining k horses must be of the same colour. Now swap the horse that was taken away with one of the k horses in the group. The k remaining horses are, again by the inductive hypothesis, of the same colour. But the horse left out must also be of the same colour.

Thus the statement is true for all n.

Conclusion: All horses are of the same colour!

Can you see what is wrong with this proof?

* * * * *

THE BABYLONIAN ALGORITHM FOR SQUARE ROOTS

Benito Hernández-Bermejo
Universidad Nacional de Educación a Distancia, Madrid

The ancient city of Babylon (whose civilisation flourished from about 2000 BC to 550 BC) achieved a surprisingly rich development in mathematics. In fact, the Babylonian achievements surpassed those of many other (later) civilisations, such as the Egyptian. Some of their discoveries were both efficient and simple. An example of this is their algorithm (that is to say, computational procedure) for square root. Here is how it went.

Suppose we want to find $\sqrt{a}$, and suppose the correct value is x, that is to say, $x^2 = a$. It is always possible to make a simple guess at the value of x. Suppose for the sake of illustration that we guess a number l_0, somewhat smaller than the correct value. In that case $\frac{a}{l_0} > x$. And so now put $u_0 = \frac{a}{l_0}$. If on the other hand we had made an initial guess u_0 which was too large, we would have $l_0 = \frac{a}{u_0} < x$. Either way, we have two estimates of $\sqrt{a}$: l_0, which is too small, and u_0, which is too large.[1] Their arithmetic mean $m = (l_0 + u_0)/2$ will then be a better approximation than either l_0 or u_0.

The number m is the *arithmetic* mean of l_0 and u_0. The number x we require is the *geometric* mean of these same numbers because, by the definition of l_0 or u_0, we have $x = \sqrt{a} = \sqrt{l_0 u_0}$ and this is the definition of the geometric mean.[2] Now it is a known result that the arithmetic mean of two positive numbers is always greater than or equal to the geometric, with equality if and only if the numbers are equal.[3] Thus m (unless l_0, and thus u_0, happens to be exactly correct) is an *overestimate* of x. Call it u_1.

Similarly $\frac{a}{m}$ will also be a better approximation than either of l_0 or u_0. But as m $(= u_1)$ is an *overestimate*, $\frac{a}{m}$ $(= l_1$, say) will be an *underestimate*, and hence we have two new approximations l_1 and u_1 such that

$$l_0 < l_1 < x < u_1 < u_0.$$

We may now repeat the algorithm, but using l_1 and u_1 instead of l_0 and u_0, and indeed we may continue in this way, finding better and better approximations to x.

---

[1] It *could* just happen that the initial guess was spot on, but we will ignore this possibility except for some occasional remarks.

[2] See *Function*, *Vol* 15, *Part* 4, p. 98.

[3] A proof is given below. For a more general case, see *Function*, *Vol* 8, *Part* 1, p. 15.

As an example, work out successive approximations to $\sqrt{2}$, starting with $l_0 = 1$. We find successively $l_0 = 1$, $u_0 = 2$, $u_1 = 1.5$, $l_1 = 1.333...$, $u_2 = 1.41666...$, $l_2 = 1.41176470...$, $u_3 = 1.41421568...$, $l_3 = 1.41421143...$, $u_4 \approx l_4 = 1.41421356...$. So the fourth calculation (technically *iteration*) produces a result that is correct to 8 (in fact 9) decimal places. The method is very accurate and also quite simple and straightforward.

As a matter of detail, the Babylonians expressed their fractions in base sixty.[4] If we write this in a more modern notation, their value comes out as $(1; 24, 51, 10, 7, 46)$ which is to be interpreted to mean

$$1 + \frac{24}{60} + \frac{51}{60^2} + \frac{10}{60^3} + \frac{7}{60^4} + \frac{46}{60^5}.$$

This figure is accurate to 9 decimal places in our way of seeing things. However, this takes matters a little further than their scribes actually went.

They seem to have begun with an initial guess of 1.5 (which is u_1 in our notation above).[5] They called this (as we write it in modern notation) $(1; 30)$, which is to say $1\frac{30}{60}$. This then gave them $l_1 = (1; 20)$ (compare above), and then they had $u_2 = \frac{1}{2}\{(1; 30) + (1; 20)\}$, which is $(1; 25)$ and again there is agreement with our earlier calculation. From u_2, they calculated l_2, which we can write as $2/(1; 25)$ and this is equal to $(1; 24, 42, 21, 10, 35, ...)$. It seems, however, that they rounded this to $(1; 24, 42, 21)$, and so the next approximation was $\frac{1}{2}\{(1; 25)+(1; 24, 42, 21)\}$, which works out to be $(1; 24, 51, 10)$. In our notation this works out to be 1.414212963, a figure that is out by about 6 in the 7th decimal place. Had they performed the calculation a little more accurately, they would have got the value given in the paragraph above.

Figure 1 shows some of the evidence[6] for the calculation. The purpose would seem to be to find the diagonal of a square of side 30. You can make out the Babylonian numeral for 30 in the upper left; you can also make out the number $(1; 24, 51, 10)$, written in Babylonian cuneiform of course,

[4]For more on Babylonian mathematics and for this story in particular, see either Carl Boyer's *A History of Mathematics*, p. 31 or (with much more detail) *Mathematical Cuneiform Texts* by O Neugebauer and A Sachs, pp. 42-43.

[5]The calculation is preserved on a clay tablet known as YBC 7289. (YBC stands for "Yale Babylonian Collection" and it is No. 7289 in their catalogue.)

[6]The tablet depicted here uses a standard value of $\sqrt{2}$ and does not say how it was calculated. However, there is modern research to show that the inspired guesswork of Neugebauer and Sachs (on which this account is based) is substantially correct. This modern work depends on later discoveries; it was publicised recently in an Internet posting by Eleanor Robson of Oxford University.

along the main diagonal. Underneath this is the answer $(42; 25, 35)$ if we make allowance for a missing symbol where the tablet has been damaged.

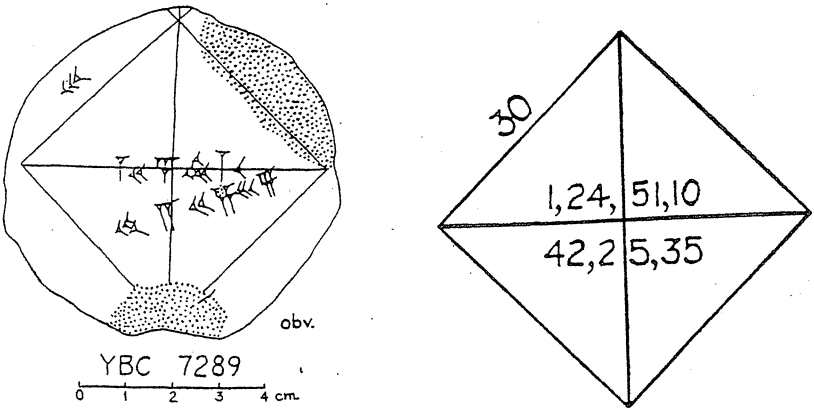


Figure 1

A drawing of the Babylonian clay tablet YBC 7289, showing the diagonal of a square of side 30. To the left is the front (obverse) of the tablet, to the right a stylised explanation. The back (reverse) of the tablet seems to depict the calculation of a rectangular diagonal, but it is too badly damaged to be deciphered, nor is it shown here. (From Neugebauer and Sachs.)

The above account explains all the basic concepts and demonstrates how accurate and efficient the method really is. However, it does contain some assertions which, while they are plausible, are not actually proved. For those who want more detail, the appendix below will give the technicalities.

Appendix

To show that the method always works, we prove some technical but not very difficult results. These will be expressed as four lemmas, or "mini-theorems". The first shows that the arithmetic mean of l_0 and u_0 does in fact lie between them.

Lemma 1. If $l_0 < u_0$, then $l_0 < \frac{l_0+u_0}{2} < u_0$.

Proof: Because $l_0 < u_0$, $l_0+l_0 < l_0+u_0$ and also $l_0+u_0 < u_0+u_0$. Dividing these inequalities by 2 produces the required result.

Next, we shall prove that m, that is to say u_1, is in fact an overestimate.

Lemma 2. $\frac{l_0+u_0}{2} \geq \sqrt{l_0 u_0}$.

Proof: $\frac{l_0+u_0}{2} - \sqrt{l_0 u_0} = \frac{1}{2}(l_0 + u_0 - 2\sqrt{l_0 u_0}) = \frac{1}{2}(\sqrt{u_0} - \sqrt{l_0})^2 \geq 0$.

Note that equality holds only in the unusual case for which $l_0 = u_0$ (and thus both are equal to x). Otherwise the arithmetic mean is larger than the true value of the square root, which is the geometric mean. In what follows, we ignore this special case.

Lemma 3. $l_0 < \frac{a}{m} < x$.

Proof: Since $m > x$ (because of the inequality of the arithmetic and the geometric mean), $1 > \frac{x}{m}$. Multiply by x throughout to find $x > \frac{x^2}{m} = \frac{a}{m}$. Also, by Lemma 1, $m < u_0$, and so $\frac{u_0}{m} > 1$. Multiply both sides of this inequality by l_0 to find $\frac{l_0 u_0}{m} > l_0$. However, by definition $l_0 u_0 = a$, and this completes the proof of the lemma. For convenience we will define $l_1 = \frac{a}{m}$.

The purpose of these lemmas is to show that m $(= u_1)$ and $\frac{a}{m}$ $(= l_1)$ are both better approximations to x than are l_0 and u_0. We now have

$$l_0 < l_1 < x < u_1 < u_0.$$

Furthermore, we may now use l_1 in place of l_0 and u_1 in place of u_0 in lemmas 1, 2, 3 to get even better approximations l_2 and u_2; and so on. In this way we can know that

$$l_0 < l_1 < l_2 < \ldots < x < \ldots < u_2 < u_1 < u_0.$$

We now know that each of the l's lies closer to the true value of x than the previous ones and the same is true of the u's. It remains as a theoretical possibility, however, that the values might home in not on the true value of x, but rather on some other nearby value. However, we shall now see that this cannot be. Our strategy will be to show that the difference between each l and its corresponding u becomes ever smaller and can be made as small as we like. Thus the true value is "squeezed" into an ever-diminishing space and we can confine it as closely as we please (although we will never attain the theoretically exact value).

Lemma 4. $\frac{u_1-l_1}{u_0-l_0} < \frac{1}{2}$.

Proof: We have:

$$u_1 - l_1 = \frac{l_0 + u_0}{2} - \frac{2a}{l_0 + u_0} = \frac{l_0^2 + u_0^2 + 2l_0u_0 - 4l_0u_0}{2(l_0 + u_0)},$$

where use has been made of the equation $a = l_0u_0$. But now this gives

$$u_1 - l_1 = \frac{(u_0 - l_0)^2}{2(u_0 + l_0)},$$

and so $\frac{u_1-l_1}{u_0-l_0} = \frac{u_0-l_0}{2(u_0+l_0)} < \frac{1}{2}$, as claimed.

Thus each pair of estimates offers a range less than half that of the previous pair, and the interval into which the true value of x is confined is progressively reduced and may be made as small as we like merely by continuing the process for long enough.

* * * * *

Three golfers named Tom, Dick, and Harry are walking to the clubhouse: Tom, the best golfer of the three, always tells the truth. Dick sometimes tells the truth, while Harry, the worst golfer, never does.

Figure out who is who.

* * * * *

MARK EID'S THEOREM

Mark Eid, Year 9, Donvale Secondary College

Generally speaking, we need two sides of a right-angled triangle in order to work out the third, but here are two ways to find right-angled triangles with integral sides when only *one* integer is given.

Part 1. Odd Numbers

Take any odd number and then:

1. Square it.
2. Divide the result by two.
3. Take the integers immediately above and below the answer to this division.

Example 1. Let 3 be the number; its square is 9. $\frac{9}{2} = 4.5$. The numbers found by the above process are therefore 4, 5. We now see that $3^2 + 4^2 = 5^2$.

Example 2. Let 55 be the number; its square is 3025. $\frac{3025}{2} = 1512.5$, giving 1512, 1513. We now see that $55^2 + 1512^2 = 1513^2$.

Example 3. Let 133 be the number; its square is 17689. $\frac{17689}{2} = 8844.5$, giving 8844, 8845. We now see that $133^2 + 8844^2 = 8845^2$.

This process works for any odd number.

Part 2. Even Numbers

Take any even number and then:

1. Square it.
2. Divide the result by four.
3. Take the integers immediately above and below the answer to this division.

Example 1. Let 8 be the number; its square is 64. $\frac{64}{4} = 16$. The numbers found by the above process are therefore 15, 17. We now see that $8^2 + 15^2 = 17^2$.

Example 2. Let 84 be the number; its square is 7056. $\frac{7056}{4} = 1764$, giving 1763, 1765. We now see that $84^2 + 1763^2 = 1765^2$.

Example 3. Let 124 be the number; its square is 15376. $\frac{15376}{4} = 3844$, giving 3843, 3845. We now see that $124^2 + 3843^2 = 3845^2$.

This process works for any even number.

[Triples of integers (a, b, c) such that $a^2 + b^2 = c^2$ are called *Pythagorean triples.* They are classed as primitive if a, b, c have no common factors. There are infinitely many primitive Pythagorean triples and every Pythagorean triple is either itself primitive or else a simple multiple of a primitive triple. For example, $(3, 4, 5)$ is primitive and so are $(5, 12, 13)$ and $(8, 15, 17)$. But $(6, 8, 10)$ and $(9, 12, 15)$ are not primitive, because they are multiples of $(3, 4, 5)$.

The rather complicated formula that gives all the Pythagorean triples has been known for quite a long time – possibly as long as 4000 years! It goes like this. Let u, v be positive integers with $u > v$. Now put $a = 2uv$, $b = u^2 - v^2$ and $c = u^2 + v^2$. It is a simple matter to check that $a^2 + b^2 = c^2$, in other words that (a, b, c) is a Pythagorean triple.

Not every Pythagorean triple can be generated by this process. For example $(9, 12, 15)$ cannot be expressed in this way. However, this process does give all the primitive triples, as well as some others. If we restrict u, v to be (i) relatively prime, and (ii) one even, the other odd, then we get exactly the set of primitive triples.

As an exercise, show that if we take $u = n + 1$, $v = n$, then we get Mark's first set of triples. You should also be able to show that these are all primitive. For a second exercise, show that if $u = n$, $v = 1$, we get Mark's second set. You should also be able to show that these triples are primitive if n is even, not otherwise.

Mark's formulae are very nice and well observed.

For more on Pythagorean triples, you could consult almost any standard text on number theory. A nice treatment is to be found on pp. 86-89 of B M Stewart's *Theory of Numbers* (Macmillan, 1952). See also *Function, Vol* 6, *Part* 3, p. 20 and *Vol* 15, *Part* 3, p. 85.

Eds.]

* * * * *

BOOK REVIEW

Ido Yavetz, *From Obscurity to Enigma: The Work of Oliver Heaviside, 1872-1889*
Birkhäuser, Basel, 1995. (Science Networks, Historical Studies, No. 16)

Reviewed by Michael A B Deakin, Monash University

Oliver Heaviside lived from 1850 to 1925. He was, first and foremost, a pioneer electrical engineer who contributed greatly to the early understanding of electromagnetic phenomena. It was he who first showed that telegraphy was possible over long distances (e.g. across the Atlantic). But he can also be regarded as a physicist (for the insights he contributed into the nature of electric and magnetic phenomena) and as a mathematician (as he made much use of mathematics in his work, and indeed had some novel ideas in this field).[1]

He hardly attended school at all, and his very considerable knowledge was almost all self-acquired. This had the effect of placing him well outside "the establishment", and so there was considerable resistance to his ideas. He was also eccentric and somewhat pugnacious by nature and so made powerful enemies, which did not help matters. In 1889, however, his work did receive recognition from acknowledged experts in his field, but at almost exactly this time, he withdrew to the relative obscurity of Paignton, a seaside resort in southwest England, and continued his researches in isolation.

So just when he emerged from obscurity, he retired to become an enigma. This gives the title of Yavetz's book, which also argues that his best work was already complete by the time that he gained recognition, and that his subsequent research was less important. Thus the concentration of attention on the period 1872-1889.

This is a wonderful and brilliant book, because it takes the state of knowledge *at the time of Heaviside's work*, and so puts us in the shoes of one of his contemporaries. By following all the twists and turns of his thought, we appreciate the effort and the genius that went into the final ideas that emerged.

[1]For a fuller biography of Heaviside, see *Function*, *Vol* 6, *Part* 2, p. 3.

We sometimes imagine that, because science has advanced so greatly since some earlier period, the science of that period must be somehow simpler and easier to follow. This, however, is a big mistake. Now that we have the insights so hard won by the early workers like Heaviside, we have a path through the subject, bypassing all the side alleys and blind turns that lurk for the unwary. But when Heaviside and others wrote, there were *no* textbooks; these came later and used the insights developed by pioneers such as him.

So if you want really to appreciate the achievements of such people, to eavesdrop on their thought so to speak, you need to be put back into their world and to be shown how to see things with their eyes. This is what Yavetz succeeds in doing, and doing most brilliantly, in the case of Heaviside. It is also the challenge that such a book presents. For we must discard our late 20th-century viewpoint and project ourselves back in time, and this takes some doing.

I will give one example: not only is it apt to illustrate the point I am making, but it shows how cleverly Yavetz makes his case. Electrical energy may be propagated through a vacuum in the form of a wave. In Heaviside's day, it was believed that this implied that a vacuum could not thus be completely empty, but rather contained a basic substance called the "ether".[2] The presence of a wave, so the argument went, necessarily implied that there was something that could "wave about". Nowadays, since Einstein, we have given up on the ether; the entire concept is seen as redundant and we do without it. But this was not so back then.

In order to help us understand the older view, Yavetz uses a beautiful illustration (in two different senses of the word!). He reminds us of an Escher drawing that will be familiar to many readers.

> "[The picture] is a very simple depiction of what appears to be the sun or the moon peeking behind some upside down tree branches. Both the tree branches and the sun or moon are rippled in a manner that suggests most forcefully that one observes their reflection in a water surface, perturbed by the ripples of two pebbles just thrown into it. Indeed the English, German, and French translations of the picture's title read respectively: 'Rippled Surface,' 'Gekräuselte Wasserfläche,' and 'Cercles dans L'eau.'

[2]Not to be confused with the organic compound and anaesthetic, of course!

> "Note, however, how cunningly Escher entitled his picture[3] by the single word 'Rimpeling,' which denotes the act of being rippled but avoids the specification of what precisely is being rippled. A closer examination of the picture quickly reveals that nowhere is the water surface, or any other surface for that matter, explicitly drawn. There are only correlated ripples in the tree branches and moon. The observer's mind adds the water."

The ether, Yavetz suggests, is rather like Escher's "water". By means of such well-chosen examples, he shows us how to think ourselves back into the context in which Heaviside worked and thought.

But this is a difficult work. Not only does it demand (and deserve) close attention from its reader, but it does presuppose a level of mathematical competence that few *Function* readers will have. A high school student would find much of interest in this book (if space permitted, I would quote a lot more), but would need to skip over much important material.

As an example, take the case of long-distance telegraphy. The point was to send signals along a cable, signals that could encode voice messages and so allow communications between (importantly) Europe and America. As signals went at that time (and still to a large extent go today), they travelled as pulses of electricity – waves of various frequencies. The hitch was that each frequency travelled at a different speed, and so after travelling for any appreciable distance, the signal became utterly garbled.

There was much debate on how to attack this problem, and it was Heaviside who discovered the answer to it. There is a unique particular set of circumstances under which all the frequencies may be made to travel together; the signal weakens, but it does not distort.

It needs quite advanced mathematics and a knowledge of electrical circuit theory to see the force of Heaviside's solution, and high school students will not be able to follow this. Other aspects of the story will, however, be accessible to them. For example, the politics of his debates with the British Post Office (whose engineers had a complete misunderstanding of the matter and thus were going in quite the wrong direction) and the disappointment Heaviside must have felt when the patent went to the American Pupin.[4]

---

[3] In Dutch.

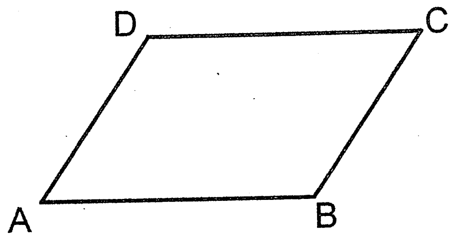
[4] Pupin became immensely rich as a result; Heaviside lived out his life in the most abject poverty.

So *Function* readers will find this a difficult and challenging work. But there is much of interest in it and a lot of this *will* be accessible, in Heaviside's own words, "with work".

Perhaps you can get your school librarian to order it. It is likely to become a classic of its kind.

* * * * *

Parallelogram fallacy



Consider the parallelogram with vertices $A = (x_1, y_1)$, $B = (x_2, y_2)$, $C = (x_3, y_3)$ and $D = (x_4, y_4)$. Then

$$\frac{y_1 - y_2}{x_1 - x_2} = \frac{y_3 - y_4}{x_3 - x_4} \text{ and } \frac{y_1 - y_4}{x_1 - x_4} = \frac{y_2 - y_3}{x_2 - x_3}$$

Therefore

$$\frac{y_1 - y_2}{y_3 - y_4} = \frac{x_1 - x_2}{x_3 - x_4} \text{ and } \frac{y_1 - y_4}{y_2 - y_3} = \frac{x_1 - x_4}{x_2 - x_3}.$$

Then,

$$\frac{y_1 - y_2}{y_1 - y_2 + y_3 - y_4} = \frac{x_1 - x_2}{x_1 - x_2 + x_3 - x_4},$$

and

$$\frac{y_1 - y_4}{y_1 - y_4 - y_2 + y_3} = \frac{x_1 - x_4}{x_1 - x_4 - x_2 + x_3},$$

which gives us

$$\frac{y_1 - y_2}{x_1 - x_2} = \frac{y_1 - y_4}{x_1 - x_4}$$

That is, the two adjacent sides $\overline{AD}$ and $\overline{AB}$ are parallel!

Sent by Garnet J Greenbury, Brisbane

* * * * *

HISTORY OF MATHEMATICS

The Greatest Area

Michael A B Deakin

Last year a colleague of mine received a query from one of the larger and more prestigious schools in the Melbourne area. They wanted information on a result known as "Fasbender's Theorem" and suggested (wrongly, as it turned out) that my colleague might be able to find something relevant on the Internet. Eventually we learned from them that the question related to a potential CAT[1], one that never in fact saw the light of day. Two (very vague) references were supplied and both turned out to be wrong in any event.

However, there are powerful and still quite convenient methods for finding out such things – without any need for the Internet. And that is how we learned of the story to follow. It is an interesting one, even if there are still quite a few loose ends we would like to tie off. The main outlines of the story are quite clear, and involve some nice mathematics.

To start the account, begin with a triangle. If the three sides of a triangle are given, then the triangle is completely specified. All its angles and also its area may be calculated from the given side-lengths. In other words, triangles are rigid objects; their angles cannot be made to alter without stretching or squashing one or more of the sides.[2]

When we come to quadrilaterals, however, matters are otherwise. The four sides of a quadrilateral do not uniquely fix all its other components, the angles and the area. If a quadrilateral were to be constructed from four rigid rods attached by pivots at the four vertices, the resultant structure would be floppy. We have considerable freedom to rearrange the shape, and of the various configurations that can be set up, some will enclose a larger area than others.

[1]Common Assessment Task – part of the senior years' assessment process in Victorian schools.

[2]This is important in engineering and is the principle underlying the cross-bracing of buildings. It is the second of the four "basic principles of Civil Engineering", which are: (1) Water flows downhill, (2) Buildings are made of triangles, (3) Use beams on their edges, not on the flat, (4) You can't push a string. I learned these principles from Sir Louis Matheson, an eminent civil engineer and Monash's first Vice-Chancellor.

This leads us to ask if there is not some such configuration that maximises the enclosed area. This is a question asked by the Greek mathematicians of antiquity, and indeed (at least in part) answered by them. That answer takes the following form.

> *Given the four side lengths of a quadrilateral, the area of the quadrilateral is maximised when its vertices are so arranged as to lie on a circle.*

It seems not to have occurred to them to ask whether this is always possible; in fact it is, but this is a matter requiring proof. But let us postpone the question for the moment and assume that such arrangement can in fact always be achieved.

This problem is one of a class of so-called "isoperimetric problems" in which the perimeter of a shape is determined and it is required to maximise the area.[3] The classic isoperimetric problem supposes a length of flexible but inextensible string (let us say). We form this into a loop and wish to arrange this loop on a flat surface in such a way as to maximise the enclosed area. It was known to the Greeks of antiquity that the answer was supplied by arranging the loop so as to form a circle.

(Just as an illustration, suppose the loop to have length L. If we arranged it in the form of a square, let us say, each side would have length $L/4$, and so the total area would be $L^2/16$. However, if we made it into a circle, the radius would be $L/2\pi$ and the area would be $L^2/4\pi$. As $\pi < 4$, this second area is the larger. In fact, if A is the area enclosed by a loop of length L, then we may write $A/L^2 \leq 1/4\pi$.)

Now in the case of the quadrilateral, the perimeter was the sum of the four side lengths and these were given. Thus we have a related problem, but we are unable to achieve the circular (and maximal) configuration, because the sides are rigid. Only at the joints may the quadrilateral "flop about". But it does the best it can and gets as close to the circular form as possible – having its vertices lie on a circle when the area is maximised.

This much and indeed much more was clearly known by about 400 AD, but I want now to move on to questions not asked back then and only raised in comparatively modern times. The triangle (any triangle) is rigid and furthermore its vertices necessarily lie on a uniquely defined circle. The

[3] The word "isoperimetric" comes from the Greek and means "equal perimeter".

quadrilateral will have maximal area if its vertices are so arranged as to lie on a circle. What then about higher polygons, pentagons, hexagons, and the like?

This is the question no-one seems to have asked for over a thousand years (except in very special cases such as regular polygons whose vertices necessarily lie on a circle). However, in 1843, Hermann Umpfenbach, a rather minor mathematician[4], stated the theorem

> *If the sides of any n-gon are given, but its angles are not specified, then its area may be maximised by so arranging the vertices that they lie on a circle.*

He gave a proof for the case $n = 5$ and that only.[5] This was, I imagine you will agree, an unsatisfactory state of affairs. However, it turned out to be short-lived. For in that same year[6], Eduard Fasbender[7] gave a general proof. That proof comprised two parts. The first was a detailed (and rather clumsy) demonstration of the case $n = 4$. (Fasbender seems to have been unaware of the fact that this was known from antiquity, and that much better proofs were already available. A quick proof will be given here in Appendix A.) The second was a short and very elegant demonstration that the case $n = 4$ in fact implied all subsequent cases. This too will be given below (Appendix B).

The theorem thus was first proved by Fasbender and it would seem that his name has been applied to it in recognition of this fact. However, this is not a completely watertight conclusion. Of my own resources, I was able to find only one other reference to the theorem: Section 12.5 of Ivan Niven's

[4]1798-1862.

[5]His proof was published in Volume 25 of *Crelle's Journal.* The official title of this journal is *Journal für die reine und angewandte Mathematik* (Journal of Pure and Applied Mathematics); it is often affectionately called after its founder and first editor. There are now hundreds (many hundreds) of periodicals devoted to the publication of mathematics. However, Crelle's was the first and it remains one of the foremost. It would seem to be a lapse from their usually very high standards to allow the publication of a theorem whose proof was incomplete. Incidentally, the proof Umpfenbach gave uses as standard theorems concerned with the properties of quadrilaterals (and akin to the sine and cosine rules for triangles) that seem now to be utterly forgotten. I had never heard of them nor could I find them in the reference books I consulted. It took some effort for me to prove these results for myself. The problems at the end of this article invite the reader to expend the same effort.

[6]In the next volume of *Crelle.*

[7]1816-1892. Another minor mathematician.

Maxima and Minima without Calculus. Niven's proof is different, and also very elegant. It is given below (Appendix C).[8]

This was the state of my knowledge when I *did* resort to the Internet, posing three questions.

1. Was there anywhere a published reference naming the theorem as Fasbender's?
2. Could anyone supply a *simple and elementary* proof that the vertices could in fact be so arranged as to lie on a circle?
3. If the classic isoperimetric problem (the one with the loop of string which ultimately is disposed as a circle) is attacked by a modern technique called "the finite element method", the answer reached is that achieved by maximising the area of a polygon with prescribed sides. Thus Fasbender's Theorem says that the solutions to the classical isoperimetric problem produced by the finite element method are the best possible. Was the theorem known in this form?

In reply, I got no answers to question 3, and no satisfactory answers to question 1. But I did get several on question 2. One in particular was most elegant and I reproduce it here as Appendix D. It came from Professor Chih-Han Sah of the Stonybrook campus of the State University of New York and attracted much favourable attention when he first posted it. Sadly, however, it turned out that the proof had already and independently been discovered by someone else, so Professor Sah was unable to claim priority for it.[9]

So there is more to be discovered but enough already to follow the main lines of the story.[10]

Following the technical details in the appendices are two problems you might like to try.

---

[8]Niven in fact gives two proofs. The first, that of Appendix C, would seem to be his own; the second is Fasbender's. Many of Niven's results are accompanied by notes which supply some of the history. Sadly, this section has no notes.

[9]The first adequate account would seem to be that in Z A Melzak's (1983) *Introduction to Geometry* (pp. 8-10), but this may depend on an earlier discussion, D S McNab's in *The Mathematical Gazette*, *Vol.* 65 (1981), pp. 22-28. McNab's discussion is, however, incomplete and thus his proof is flawed.

[10]The above discussion ignores the order in which the sides are placed. If the sides can be reconnected (but leaving their lengths intact), there are a few minor complications but of a fairly simple sort! These are left for the reader to explore.

Appendix A: Proof for the case $n = 4$

Start with Figure 1 depicting a quadrilateral with sides (in order) a, b, c, d and angles A, B, C, D as shown. Let its area be S. Since S is the sum of the areas of the triangles $\triangle ABC$ and $\triangle ACD$, we find

$$4S = 2ab\sin B + 2cd\sin D.$$

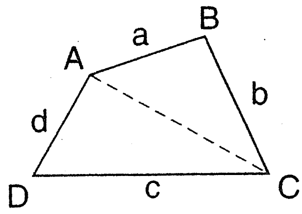


Figure 1

We may also apply the cosine rule to determine the length of the diagonal $\overline{AC}$ and so reach

$$a^2 + b^2 - 2ab\cos B = c^2 + d^2 - 2cd\cos D.$$

By rearranging this latter equation, squaring both equations and then adding them we reach, after a little work,

$$16S^2 + (a^2 + b^2 - c^2 - d^2)^2 = 4a^2b^2 + 4c^2d^2 - 8abcd\cos(B + D).$$

This last equation may in its turn be rewritten. Again the details will be omitted, but if we write

$$s = \frac{1}{2}(a + b + c + d),$$

we obtain

$$S^2 = (s-a)(s-b)(s-c)(s-d) - abcd\cos^2(\frac{B+D}{2}).$$

It is clear that the final term of this expression (following the minus sign) is always non-negative. So S^2, and hence S, will be maximised if this final term is zero. This will be achieved if $B + D = 180°$. This last equation is known to apply exactly if and only if the quadrilateral is *cyclic*, that is to say, can be inscribed in a circle (a result known to Euclid).

This proves the theorem in the case $n = 4$.[11]

Appendix B: Fasbender's deduction of the general case

Fasbender argued from Figure 2. He supposed that the polygon $ABCDEF\ldots$, whose sides $AB, BC, CD, DE, EF, \ldots$ are supposed known, was so set up as to maximise its area, but that it could not be inscribed in a circle. He next supposed that all the lower half of the

[11]This proof comes from a once very influential textbook: C V Durell and A Robson's *Advanced Trigonometry* (1930), pp. 24-27, where more detail is given. Although they give some historical details, it is not quite clear how old this particular proof is. It is, however, simpler and more elementary than that provided by Fasbender.

figure, $DEF \ldots A$ was held rigid, but that B and C were allowed to move as indicated. But if A, B, C, D did not lie on a circle, then the area of $ABCD$ could be increased by adjusting the positions of B and C so that they did.

This, however, would increase the total area of the entire polygon (as $DEF \ldots A$ was held rigid), and we have already assumed it to be maximal. Thus $ABCD$ must already lie on a circle. This circle is completely specified by the positions of B, C and D.

Next consider the quadrilateral $BCDE$. The same argument applies. These four vertices must lie on a circle, and since the positions of B, C and D suffice to determine this circle, it must be *the same circle* as we found before.

Continuing in this way, we find that *all the vertices* must lie on this same circle, and so the proof is complete.

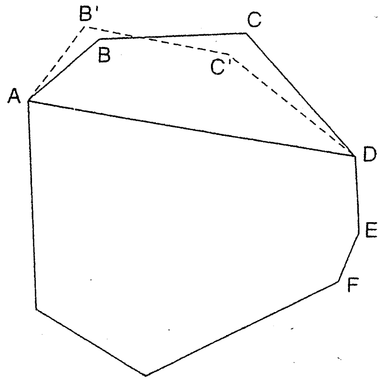


Figure 2

Appendix C: Niven's proof of the general case

Consider a polygon inscribed in a circle as shown in Figure 3a.[12] Each side of the polygon comes equipped with a corresponding arc and these arcs together make up a circle of radius r. Their total length is thus fixed at $2\pi r$.

Now suppose that we cut up Figure 3a, but keeping each side still attached to its arc. (The shape made up by the side and the arc is a *segment* of the original circle. We keep these segments intact.) Now rearrange the pieces, e.g. as shown in Figure 3b, in which the bounding curve is no longer circular. The length of the perimeter is quite unaltered, as is the total area of the segments.

However, the total enclosed area has been reduced (because of the classical isoperimetric property). Thus the cut and re-paste must have reduced the area of the enclosed polygon.

[12]The case illustrated is $n = 4$, but the argument is perfectly general.

It follows that the original area was the maximum that could be achieved.

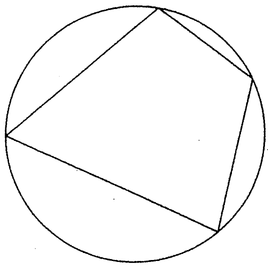


Figure 3a

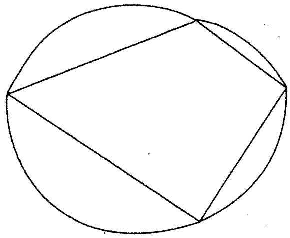


Figure 3b

Appendix D: Proof that the vertices may be made to lie on a circle

All the technical detail in appendices A, B, C above assumes that it is actually possible so to arrange the vertices of the polygon that they lie on a circle. This is in fact the case, but the matter requires proof. It's far from intuitively obvious; so let us now proceed to this proof.

If we take n given lengths, these may be connected into a polygon in exactly the case that the longest of them is shorter than the total length of all the others.[13] So, *find* the longest side of the polygon. (If two or more are equally long, choose any one of these.) Now imagine the polygon disconnected at one vertex and stretched out to form a straight line with this longest side at the extreme right-hand end.

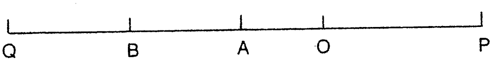


Figure 4

Call the right-hand end-point P and the other end of this longest side O. To the left of O are strung out the other vertices: $A, B, \ldots, Q$ (say). See Figure 4. On the point O place a very large circle such that the line formed from the polygon is tangent to the circle, at the point O. Having done this, push the side OP upwards so that P lies on the circle. Do the same with the points on the left of O, so that all of these also lie on the circle. The result is Figure 5.

Also shown on Figure 5 is a point P', which is the point as far up the circle on the left of O as P is on the right.

Now let the radius of the circle gradually decrease. The points P and P' will slowly rise, and if the process of shrinkage goes on long enough, they will meet at a point vertically above O. While this is going on, the point Q will also slowly rise up on the left-hand side. Initially Q lies between P' and P. As the shrinkage goes on, we could have one of the following scenarios.

[13]This is obvious once you think about it. For an interesting extension of this question, see *Function*, *Vol* 5, *Part* 3, p. 8.

1. The point P' meets Q at some stage of the process;
2. The point Q stays ahead of P', continues over the top of the circle and as it comes down meets P on its way up;
3. The points P, P' and Q all come together at the top.

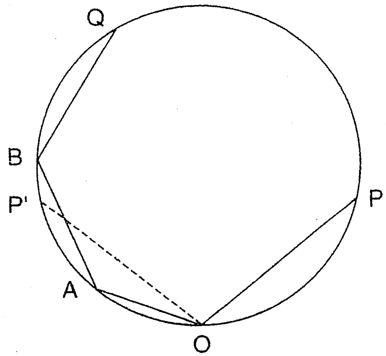


Figure 5

These are the only possibilities and each corresponds to a situation in which the various vertices are so arranged as to lie on the circumference of a circle. In case 1, the circle is shrunk until P' and Q coincide; at that stage the edge $\overline{OP'}$ should replace the edge $\overline{OP}$ to obtain the polygon.

Note also that in case 1, the centre of the circle lies outside the polygon; in case 2 it is inside, and in case 3 it just *happens* to lie at the mid-point of the longest side. These are the only possibilities.[14]

Problems

See if you can prove the technical results used by Umpfenbach, and evidently once widely known. They refer to properties of the quadrilateral and will be stated in the notation of Figure 1 and Appendix A. The first is related to a formula for area in the case of triangles (and itself closely related to the sine rule); the second is the appropriate generalisation of the cosine rule.

$$2S = bc\sin C + cd\sin D - bd\sin(C+D)$$

$$a^2 = b^2 + c^2 + d^2 - 2bc\cos C - 2cd\cos D + 2bd\cos(C+D).$$

* * * * *

[14]This proof is Melzak's, as rediscovered by Sah. McNab's earlier discussion failed to include the point P' and thus overlooked case 1.

COMPUTERS AND COMPUTING

Finding x–intercepts with Mathematics Software

Cristina Varsavsky

Computers are radically changing the way we "do mathematics". Nowadays we have access to sophisticated software capable of performing long calculations or displaying complicated graphs at the touch of a button. Of particular importance are the so-called *computer algebra systems* which manipulate not only numbers but also symbols.

In a computer algebra system, irrational numbers such as $\sqrt{2}$ or π can be manipulated as symbols, that is, they are not converted into their approximate decimal forms. When a calculator is used to find $\sqrt{72}$, it gives the approximate answer 8.4852814; a computer algebra system returns the exact simplified expression, $6\sqrt{2}$. Because they can handle symbols, these systems can also solve equations, calculate integrals, find derivatives and simplify mathematical expressions; they can actually do most of secondary and tertiary level mathematics, and more[1].

You might have heard of or used one such system; among the most widely available are DERIVE, MAPLE, MATHEMATICA, and THEORIST. There are also hand-held calculators which can manipulate symbols.

A typical problem a high school student will try to solve with a computer is to find the x–intercepts of a polynomial function $p(x)$, that is the real solutions to the equation $p(x) = 0$. This is done through a command usually called **solve**. By using this command students do not have trouble finding solutions to a quadratic, cubic, or quartic equation, although some equations may require a computer to be working for a long while, returning a long string of surds and quotients.

When it comes to a quintic (i.e. a fifth degree polynomial), students usually become puzzled with the behaviour of the computer program as it echoes back the same equation. Why is the program refusing to produce solutions? Well, this is because it does not have an in-built formula to solve such polynomial equations. Furthermore, it will never have one. The Norwegian mathematician Niels Abel proved in 1826 that polynomial equations of degree greater than 4 cannot be solved, in general, using the

[1]See also related articles in *Function, Vol* 15 *Part* 5, and *Function, Vol* 16 *Part* 1.

four operations and the extraction of roots. In other words, there will never be a "quintic formula", as there is for the quadratic, cubic and quartic.

If a polynomial equation cannot be solved exactly, it can always be solved approximately. Computer algebra systems usually do so using numerical methods for successive approximations (you must have learnt at school some of these methods, most probably the Newton-Raphson method). To apply a numerical technique a first approximation is needed, usually an upper and a lower bound for the true solution. If two or more solutions lie in the starting interval, such a numerical method will produce at most one of them and miss the rest. Consequently we need to start the numerical method with an *isolating interval*, an interval which contains only one solution.

So, unless the polynomial equation of degree higher than 4 is of a particular kind, only approximate real solutions can be found. In this case there are two steps involved in the search for real solutions: firstly, finding an isolating interval for each of them; secondly, obtaining the solution within each interval to the desired accuracy.

In simpler computer algebra systems like DERIVE, the isolation has to be done by the user. A graphical approach usually helps to visualise the isolating intervals. This isolating task must be performed with care. In some cases the scale used to produce the graph may lead to a false conclusion; also, some solutions may lie outside the screen. We always need to ask "Are there any other solutions?" and have a solid argument for the answer. Once the isolating intervals are found, the approximation has to be performed by first switching to *approximate mode* and then using repeatedly the **solve** command to find each solution.

In more sophisticated systems like MAPLE, we only need to use the command for approximate solutions[2] to obtain all $x-$intercepts; these systems have an in-built isolating routine. Although the details of the isolating algorithms are rather complex, they are based on a very simple technique of sign variations of a polynomial sequence within the interval under consideration. One such sequence is the *Fourier sequence* of the successive derivatives of a polynomial. An important result says that the number of sign variations lost from the left end to the right end of an interval must either be the number of solutions in it or exceed that number by a multiple of two.

[2]In MAPLE this command is **fsolve**.

We illustrate this with an example. Take the polynomial

$$p(x) = 36x^3 + 27x^2 - 4x - 3,$$

with the corresponding derivatives

$$p'(x) = 108x^2 + 54x - 4, \quad p''(x) = 216x + 54, \quad \text{and} \quad p'''(x) = 216.$$

To find out the number of solutions in the interval $(-1, 1)$, we evaluate the Fourier sequence at the two end values. For $x = -1$ we have

$$\{p(-1), p'(-1), p''(-1), p'''(-1)\} = \{-8, 50, -162, 216\}$$

Within this sequence, the sign changes 3 times. Now for $x = 1$, we have

$$\{p(1), p'(1), p''(1), p'''(1)\} = \{56, 158, 270, 216\},$$

with no sign variations. Then the number of sign variations lost from -1 to 1 is 3, meaning that there are one or three solutions in the interval $(-1, 1)$. Since this information is not very helpful we split the interval in half, and obtain the Fourier sequence for $x = 0$:

$$\{p(0), p'(0), p''(0), p'''(0)\} = \{-3, -4, 54, 216\}$$

It has only one sign variation. Then the number of sign variations lost from -1 to 0 is 2, and from 0 to 1 is 1. Therefore there is exactly one solution to the equation $p(x) = 0$ in the interval $(0, 1)$ and two or none in the interval $(-1, 0)$. We investigate further and split this latter interval in half. We have

$$\{p(-0.5), p'(-0.5), p''(-0.5), p'''(-0.5)\} = \{1.25, -4, -54, 216\}$$

with two sign variations. This fully isolates the three solutions in the intervals $(-1, -0.5)$, $(-0.5, 0)$, and $(0, 1)$.

The Fourier sequence isolation technique is based on the following basic result:

> If α is a simple real solution to $p(x) = 0$, then it is possible to find a (small) interval around α, $(\alpha - \epsilon, \alpha + \epsilon)$, so that $p(x)$ and $p'(x)$ have opposite signs in the interval $(\alpha - \epsilon, \alpha)$ and equal signs in the interval $(\alpha, \alpha + \epsilon)$.

That is, the sequence $\{p, p'\}$ loses one sign variation at a simple solution. Can you see why this is true?

The higher derivatives in the Fourier sequence are there for detecting solutions of higher multiplicity.

* * * * *

PROBLEM CORNER

SOLUTIONS

PROBLEM 20.1.1

If a hen and a half lay an egg and a half in a day and a half, how many hens will lay two eggs in three days?

SOLUTION

A hen and a half lay an egg and a half in the same time as one hen lays one egg. One hen therefore lays one egg in a day and a half and hence two eggs in three days.

PROBLEM 20.1.2 (Claudio Arconcher, São Paulo, Brazil)

Find solutions in positive integers for

(a) $x^x + y^y = 2xy$ (b) $x^x + y^y + z^z = xy + xz + yz$.

SOLUTION

Suppose firstly that $x \geq y$ in equation (a). If $x \geq 3$ then $x^x + y^y > x^x \geq x^3 > 2xy$, so there are no solutions with $x \geq 3$. Checking the potential solutions (1,1), (2,2) and (2,1), we discover that only the first two are solutions. Because of the symmetry between x and y in the equation, the case $x \leq y$ yields no further solutions. The solutions are $(x, y) = (1, 1)$ and $(x, y) = (2, 2)$.

Equation (b) can be solved in a similar way. Begin by assuming that $x \geq y \geq z$. If $x \geq 3$, then $x^x + y^y + z^z > x^x \geq x^3 \geq 3x^2 \geq xy + xz + yz$, giving no solutions. Checking all potential solutions with x equal to 1 or 2 gives just the two solutions (1,1,1) and (2,2,2). Dropping the assumption $x \geq y \geq z$ gives no further solutions, because of the symmetry between the three variables in the equation.

Also solved by Peter Bullock (Norwood Secondary College).

PROBLEM 20.1.3

Prove that for any positive integer n, $\lfloor n + \sqrt{n} + \frac{1}{2} \rfloor$ is not a perfect square. (The notation $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x.)

SOLUTION

Since n is an integer, $\lfloor n + \sqrt{n} + \frac{1}{2} \rfloor = n + \lfloor \sqrt{n} + \frac{1}{2} \rfloor = n + m$, where $m = \lfloor \sqrt{n} + \frac{1}{2} \rfloor$. In order to show that $\lfloor n + \sqrt{n} + \frac{1}{2} \rfloor$ is not a perfect square, it is sufficient to show that $n + m$ lies strictly between m^2 and $(m+1)^2$. We can do this in the following way. From the definition of m, we deduce $m \leq \sqrt{n} + \frac{1}{2} < m + 1$. Subtracting $\frac{1}{2}$ from each expression and then squaring, we obtain

$$(m - \frac{1}{2})^2 \leq n < (m + \frac{1}{2})^2.$$

Therefore

$$(m - \frac{1}{2})^2 + m \leq n + m < (m + \frac{1}{2})^2 + m.$$

Expanding the brackets and simplifying, we obtain

$$m^2 + \frac{1}{4} \leq n + m < m^2 + 2m + \frac{1}{4}.$$

Rewriting this as $m^2 + \frac{1}{4} \leq n + m < (m+1)^2 - \frac{3}{4}$, we conclude that $n + m$ is between m^2 and $(m+1)^2$, as required.

PROBLEM 20.1.4 (1994 Old Mutual Mathematics Olympiad, South Africa; reprinted from *Mathematical Digest*, January 1995, University of Cape Town)

A, B, C, D and E are distinct points in three-dimensional space, such that A, B and C lie on the surface of a sphere. Prove that at most one of the four-sided figures $XYDE$, where X and Y are two of the points A, B and C, can be a parallelogram.

SOLUTION from *Mathematical Digest*

Since any three non-collinear points lie on a circle (and hence on the surface of a sphere), the condition on A, B and C means simply that they are non-collinear.

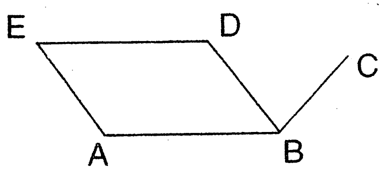


Figure 1

Suppose $ABDE$ is a parallelogram (see Figure 1). Then $AB \| ED$ and hence neither BC nor AC is parallel to ED. So neither $BCDE$ nor $ACDE$ is a parallelogram. So at most one of the quadrilaterals is a parallelogram. In fact, at most one is a trapezium.

It is, however, possible that $ABDE$ and $ADCE$ are parallelograms (see Figure 2) but the problem specifies that DE must be a side, not a diagonal, of any parallelogram.

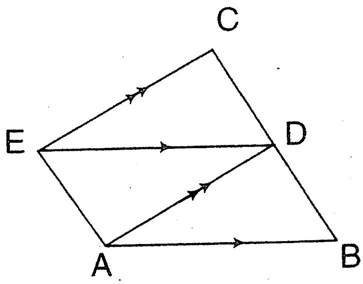


Figure 2

PROBLEM 20.1.5 (from *Mathematical Mayhem*, Vol 7, Issue 5, University of Toronto)

Let $G(n)$ be the number of strictly increasing or decreasing sequences formed using the values $1, 2, \ldots, n$, e.g. for $n = 2$ there are 4 sequences (1), (2), (1,2), (2,1). Find an explicit formula for $G(n)$ in terms of n.

SOLUTION

There are 2^n different ways of selecting the values in the sequence, since each of the n values $1, 2, \ldots, n$ is either included in a sequence or not. One of these selections corresponds to choosing none of the numbers, there are n selections in which one number is chosen (each of which yields just one sequence), and the remaining $2^n - n - 1$ selections each yield two sequences, one increasing and one decreasing. The total number of sequences is therefore $n + 2(2^n - n - 1)$, or $2^{n+1} - n - 2$.

PROBLEM 20.1.6 (from Swedish Mathematical Olympiad, 1979 Qualifying Round; reprinted from *Mathematical Mayhem*, Vol 8, Issue 1, University of Toronto)

For which real values of $a, a \geq 1$, is $\sqrt{a+2\sqrt{a-1}}+\sqrt{a-2\sqrt{a-1}}=2$?

SOLUTION

Let $b=\sqrt{a-1}$. Then $a=b^2+1$. Substituting into the given equation, we obtain

$$\sqrt{b^2+1+2b}+\sqrt{b^2+1-2b}=2.$$

Therefore

$$\sqrt{(b+1)^2}+\sqrt{(b-1)^2}=2.$$

We can write this as $|b+1|+|b-1|=2$. Since $b \geq 0$, we can remove the first pair of absolute value signs and write $b+1+|b-1|=2$, which simplifies to $b+|b-1|=1$. The case $b \geq 1$ yields $b+b-1=1$, so $b=1$ and hence $a=2$. The case $b<1$ yields $b+1-b=1$, which imposes no further restriction on b; in this case, the restriction $b<1$ implies $a<2$. The possible values of a are therefore $1 \leq a \leq 2$.

PROBLEM 20.1.7 (based on a problem seen on the Internet)

1. Let two circles C_1 and C_2 be given, with C_1 inside C_2. A third circle, C_3, moves around the region between C_1 and C_2, in such a way that it is always tangent to both circles. Prove that the locus of the centre of C_3 is an ellipse.

2. Now suppose the problem in (1) is modified so that C_1 is outside rather than inside C_2. What type of figure is described by the locus of the centre of C_3 in this situation?

3. What happens if C_1 and C_2 overlap?

SOLUTION

With C_1 inside C_2, let the centres of C_1, C_2 and C_3 be A, B and C respectively, and let their radii be r_1, r_2 and r_3 respectively; note that r_1 and r_2 are constants and r_3 is a variable. Then $AC=r_1+r_3$ and $CB=r_2-r_3$, so $AC+CB=r_1+r_2$, which is a constant, independent of the location of the circle C_3. By one of the defining properties of an ellipse, the locus of C is an ellipse with foci at A and B.

With C_1 outside C_2 and using the same notation as before, $AC = r_1 + r_3$ and $CB = r_2 + r_3$, so $AC - CB = r_1 - r_2$, which is a constant. This is a defining property of one branch of a hyperbola, so the locus of C is a branch of a hyperbola with foci at A and B. (The degenerate case $r_1 = r_2$ is an exception; in this case, the locus is a line.)

Now let C_1 and C_2 overlap. If C_3 is in the region of overlap, we get part of one branch of a hyperbola. The rest of the branch is obtained when C_3 is outside both C_1 and C_2. If C_3 is in one of the other two regions (inside one circle and outside the other), we get part of an ellipse, the two parts together forming the complete figure.

Also solved by Claudio Arconcher (São Paulo, Brazil).

PROBLEM 20.1.8 (from the Internet, author unknown)

From each corner of a unit square draw a quarter of an inscribed unit circle. Find the area of the central diamond shape where the four quarter circles overlap.

SOLUTION

This problem was poorly worded; it should have said "an inscribed quarter of a unit circle" rather than "a quarter of an inscribed unit circle", and the central region is where the four quadrants (rather than quarter circles) overlap.

Let the areas of the three differently shaped regions arising in the problem be x, y and z as shown in Figure 3.

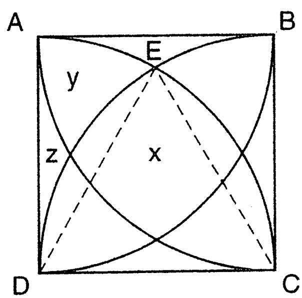


Figure 3

From the area of the square and the area of a quadrant respectively, we obtain the following two equations connecting x, y and z:

$$x + 4y + 4z = 1 \tag{1}$$

$$x + 3y + 2z = \frac{\pi}{4}. \tag{2}$$

A third equation is obtained by dividing the square into four regions: the equilateral triangle with vertices at C, D and E (indicated by dotted lines in the figure), the two sectors DAE and CBE, and the region bounded by the side AB and the arcs AE and BE:

$$\frac{\sqrt{3}}{4} + 2 \times \frac{1}{2} \times \frac{\pi}{6} + z = 1. \tag{3}$$

Solving equation (3) for z and substituting into equations (1) and (2), we obtain

$$x + 4y = -3 + \sqrt{3} + \frac{2\pi}{3} \tag{4}$$

$$x + 3y = -2 + \frac{\sqrt{3}}{2} + \frac{7\pi}{12}. \tag{5}$$

Eliminating y from equations (4) and (5) yields the solution for x, which is the required area:

$$x = 1 - \sqrt{3} + \frac{\pi}{3}.$$

K R S Sastry (Dodballapur, India) obtained a solution for the more general problem in which the square is replaced by a rhombus with angle $BAD = \theta$, where $60° < \theta \leq 90°$. (The restriction is necessary in order to ensure that the region whose area is to be found falls within the rhombus.) He showed that the area in this case is $\sin\theta - \sqrt{3} + \frac{\pi}{3}$. Of course, the solution to the original problem can then be obtained by putting $\theta = 90°$.

This problem was also solved by Peter Bullock (Norwood Secondary College).

PROBLEMS

Readers are invited to send in solutions (complete or partial) to any or all of these problems. All solutions received in sufficient time will be acknowledged in the next issue but one, and the best solutions will be published.

PROBLEM 20.3.1 (modified from a problem in *Alpha*, August 1995)

Find a straightedge and compasses construction of a triangle ABC, given three points X, Y, Z where the circumscribed circle intersects respectively the extension of the median through C, the extension of the altitude through C, and the extension of the angle bisector at C.

PROBLEM 20.3.2 (based on a problem on the Internet)

Ten girls and ten boys are at a party. All the girls prefer cakes, and all the boys prefer ice creams. The children sit around a round table in no particular order, and each of them is served either a cake or an ice cream. Show that it is possible to rotate the table in such a way that at least ten children get what they prefer.

PROBLEM 20.3.3 (1995 Old Mutual Mathematics Olympiad, South Africa; reprinted from *Mathematical Digest*, October 1995, University of Cape Town)

Suppose that $a_1, a_2, a_3, \ldots, a_n$ are the numbers $1, 2, 3, \ldots, n$ but written in any order. Prove that

$$(a_1 - 1)^2 + (a_2 - 2)^2 + (a_3 - 3)^2 + \ldots + (a_n - n)^2$$

is always even.

PROBLEM 20.3.4 (from *Alpha*, July 1995)

Prove that for every natural number n such that $n \geq 2$ and every natural number k, the number $(1 + k + k^2 + \ldots + k^n)^2 - k^n$ is not prime.

PROBLEM 20.3.5 (posted on the Internet by Bill Taylor, University of Canterbury, New Zealand)

Given any four points on the circumference of a circle, mark the four midpoints of the arcs between adjacent pairs of points. Form two chords by joining the opposite pairs of midpoints. Prove that these two chords cross at right angles.

(Bill Taylor has pointed out that the assertion in this problem appeared originally as a statement of fact, without any proof, as part of the proof of another result in a published paper. Presumably the author of the paper thought it was obvious!)

* * * * *

OLYMPIAD NEWS

Hans Lausch

1. The Eighth Asian Pacific Mathematics Olympiad

The Asian Pacific Mathematics Olympiad (APMO), an annual competition, was started in 1989 by Australia, Canada, Hong Kong and Singapore. Since then the number of participating Pacific Rim countries has grown to nearly twenty. Moreover, Argentina, South Africa, and Trinidad and Tobago were using the contest questions for their national competitions. In Australia, 26 students sat this four-hour examination on 12 March.

Time allowed: 4 hours. No calculators to be used. Each question is worth 7 points.

1. Let $ABCD$ be a quadrilateral with $AB = BC = CD = DA$. Let MN and PQ be two segments perpendicular to the diagonal BD and such that the distance between them is $d > BD/2$, with $M \in AD$, $N \in DC$, $P \in AB$, and $Q \in BC$. Show that the perimeter of the hexagon $AMNCQP$ does not depend on the position of MN and PQ so long as the distance between them remains constant.

2. Let m and n be positive integers such that $n \leq m$. Prove that

$$2^n n! \leq \frac{(m+n)!}{(m-n)!} \leq (m^2+m)^n.$$

3. Let P_1, P_2, P_3, P_4 be four points on a circle, and let I_1 be the incentre of the triangle $P_2P_3P_4$, I_2 be the incentre of the triangle $P_1P_3P_4$, I_3 be the incentre of the triangle $P_1P_2P_4$, I_4 be the incentre of the triangle $P_2P_3P_1$. Prove that I_1, I_2, I_3, I_4 are the vertices of a rectangle.

4. The National Marriage Council wishes to invite n couples to form 17 discussion groups under the following conditions:

   (1) All members of a group must be the same sex, i.e. they are either all male or all female.

   (2) The difference in the size of any two groups is either 0 or 1.

   (3) All groups have at least one member.

   (4) Each person must belong to one and only one group.

Find all values of n, $n \leq 1996$, for which this is possible. Justify your answer.

5. Let a, b, c be the lengths of the sides of a triangle. Prove that

$$\sqrt{a+b-c} + \sqrt{b+c-a} + \sqrt{c+a-b} \leq \sqrt{a} + \sqrt{b} + \sqrt{c}$$

and determine when equality occurs.

2. Australians at the XXXVII International Mathematical Olympiad

The performance of students at the APMO was used in selecting ten candidates for the team which is to represent Australia at this year's International Mathematical Olympiad (IMO). Also, fifteen highly gifted students, with at least one more year of secondary education ahead of them, were singled out for further training.

These 25 students participated in the ten-day Team Selection School of the Australian Mathematical Olympiad Committee. Following a tradition, the School was held in Sydney. Participants had to undergo a day and evening programme consisting of tests and examinations, problem sessions and lectures by mathematicians. Finally, the 1996 Australian IMO team was selected.

Bombay (Mumbai) is the venue of the XXXVII IMO scheduled for July. There the Australian team will have to contend with six problems during 9 hours spread equally over two days in succession. The Australian team is:

John Dethridge, Year 12, Melbourne Grammar School, Melbourne, Victoria; Daniel Ford, Year 12, James Sheahan High School, Orange, NSW; Jian He, Year 12, University High School, Melbourne, Victoria; Alexandre Mah, Year 12, North Sydney Boys' High School, Sydney, NSW; Daniel Mathews, Year 11, Scotch College, Melbourne, Victoria; Brett Parker, Year 12, Penleigh & Essendon Grammar School, Melbourne, Victoria.

Reserve: Brian Scerri, Year 12, Canberra Grammar School, Canberra, ACT.

Good luck to them all!

* * * * *

Function is a mathematics magazine produced by the Department of Mathematics at Monash University. The magazine was founded in 1977 by Prof G B Preston. *Function* is addressed principally to students in the upper years of secondary schools, and more generally to anyone who is interested in mathematics.

Function deals with mathematics in all its aspects: pure mathematics, statistics, mathematics in computing, applications of mathematics to the natural and social sciences, history of mathematics, mathematical games, careers in mathematics, and mathematics in society. The items that appear in each issue of *Function* include articles on a broad range of mathematical topics, news items on recent mathematical advances, book reviews, problems, letters, anecdotes and 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
900 Dandenong Rd
Caulfield East VIC 3145, Australia
Fax: +61 (03) 9903 2227
e-mail: function@maths.monash.edu.au

Function is published five times a year, appearing in February, April, June, August, and October. Price for five issues (including postage): \$17.00*; single issues \$4.00. Payments should be sent to: The Business Manager, *Function*, Mathematics Department, Monash University, Clayton VIC 3168; cheques and money orders should be made payable to Monash University. Enquiries about advertising should be directed to the Business Manager.

For more information about *Function* see the magazine home page at `http://www.maths.monash.edu.au/~cristina/function.html`.

*\$8.50 for *bona fide* secondary or tertiary students.