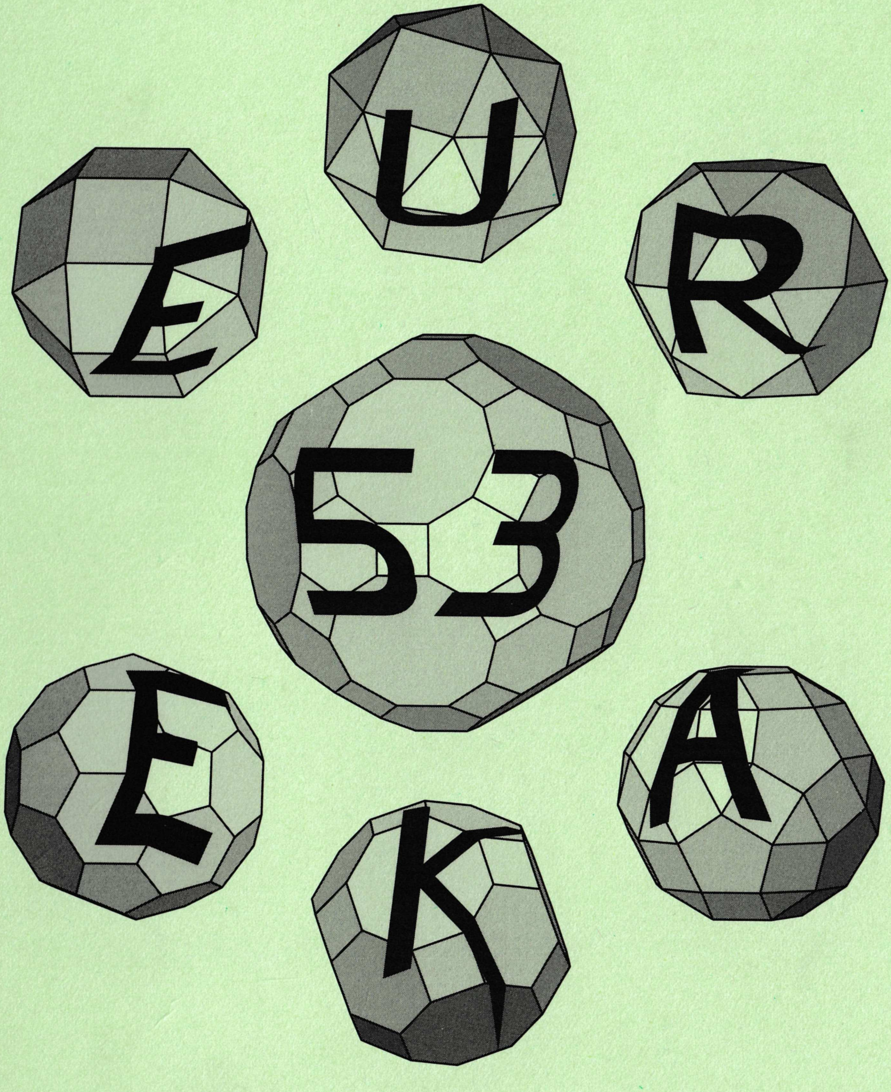
Eureka Digital Archive

archim.org.uk/eureka

Eureka Editor
archim-eureka@srcf.net
The Archimedeans
Centre for Mathematical Sciences
Wilberforce Road
Cambridge CB3 0WA
United Kingdom

Published by The Archimedeans, the mathematics student society of the University of Cambridge

Thanks to the Betty & Gordon Moore Library, Cambridge


EUREKA
53

Eureka

Eureka is the journal of the Archimedeans, which is the Cambridge University Mathematical Society, and a Junior Branch of the Mathematical Association. *Eureka* is published approximately annually, but since it, like the Society, is run entirely by student volunteers, it is impossible to guarantee precise publication dates. The Society also publishes QARCH, a problems journal.

Subscriptions

A subscription account may be opened by sending at least ten pounds to the Business Manager at the address below; issues will be sent as they are published and you will be informed when your account is running low. It is also possible to subscribe to *Eureka* by standing order. Back numbers of some of the more recent issues are available, and photocopies of any issue may be made available by special arrangement. TEX source for issues since *Eureka* **49** may also be available, subject to copyright; contact the Business Manager for details.

Electronic mail

The Society can receive electronic mail via JANET and the Internet, and can read most electronic media. Our electronic mail addresses are given below. There may be long delays in dealing with messages of any kind in the University vacations (especially June to October), and you are asked not to send large amounts of text electronically without prior warning.

Addresses

Correspondence, clearly marked "Editor" or "Business Manager", should be sent to:

Eureka, The Arts School, Bene't Street, Cambridge CB2 3PY, England

Our electronic mail address is:

`archim@phoenix.cambridge.ac.uk`

It should be noted that this address will cease to be valid at some point in late 1995 when Phoenix is decommissioned; our address from then will probably be:

`archim@hermes.cambridge.ac.uk`

but this is yet to be confirmed.

EUREKA

Editor: Colin Bell — Number 53, February 1994

Assistant Editor: Mark Walters — Business Manager: Peter Mennie

TABLE OF CONTENTS

ISSN 0071-2248

Editorial

As has become traditional in recent years in the editorial, and as I am something of a conformist, I shall start with the reasons why this particular issue has come out at the time it has, and not some other time. However (to preserve some sense of mystery) I shall announce that this *Eureka* is both earlier *and* later than could have been expected. As with *Eureka* **52**, the original intention was to produce an issue by the end of the Michaelmas term; this however was not to be, as by the beginning of December we had received insufficient material. Yet within the next two weeks enough material had come in, that not only could we produce the *Eureka* you see before you, but that some material had to be edited or left out altogether. Nevertheless, this issue follows less than a year behind its predecessor: the previous Editor made a valiant attempt to sabotage this by submitting an article full of diagrams even harder to typeset than those in his volume, but redeemed himself by not only submitting it with plenty of time to spare, but also putting considerable work in on its layout; for this I am considerably grateful.

It has been a pleasure to receive articles from both within and outside the University on a great variety of subjects; if nothing else, they have educated the Editor on areas of mathematics with which he had at best a passing acquaintance. The majority of the contributors are also students (or at least recent ex-students), which is a Good Thing: *Eureka* should be their voice. However, a brief perusal of the college loyalties of the contributors reveals that the vast majority of this issue was written by members of one particular college—this in the Editor's view does not bode well for the Archimedeans or the mathematical community in Cambridge at large. Trinity has been fairly dominant in the academic side of mathematics for some considerable time now (I shall not bore you with a list of famous names here), but (if the affiliations of past Editors of this journal and members of the Committee are anything to go by) this has not extended to more recreational mathematics in general or the Archimedeans in particular before. Yet this is now the eleventh successive year in which half or more of the committee have been from that same college, and in the Editor's five years in the Society it is fair to say this has been reflected in the active membership as a whole; the former can be put in perspective by noting that this state of affairs occurred only once in the thirty or so years up until 1970, when Girtonians (surprisingly enough) were in the majority for two years. (It is amusing to note that the first committee meeting in this period was held at Girton. None of the non-Girtonians turned up and the experiment has yet to be repeated.)

It should not be thought that I blame the authorities at Trinity for the current state of affairs; far from it: it is hardly surprising that they should wish to have as many good mathematicians in the college as possible. And there is no doubt that many are attracted by the prestige, both of the college's history, and also of having one of the most eminent mathematicians alive as Master.

So what is the problem? The Archimedeans does not obviously seem to have suffered from this bias: talks are still arranged, the subgroups and College Societies continue to function, and *Eureka* is still published (most years at least). Indeed the activities that the Society engages in are more numerous and more varied than those in other universities. Warwick, for example, a university of considerable size and standing in British mathematics, has had no mathematics society at all for the last few years (although it has now been relaunched and appears to be enjoying some success—we wish it well).

The Archimedeans has as one of its prime objects "to promote enjoyment and understanding of mathematics among students of all disciplines", and it is my belief that as a society we are doing the right sort of things to achieve this. But are we actually getting through to the membership? Why is such a small proportion of it active? Part of the reason may be the general decline in interest in many University societies in recent years; the theories for this are numerous and varied and I shall not dwell on them here. But those who think there is more to mathematics than the Tripos, and have come to Cambridge for an education rather than a degree, should still be able to and can find their spiritual home in the Archimedeans. That few have come is a concern. What is stopping them? The claim that no-one else is interested in mathematics is preposterous and can be discounted. The charge of an image problem is closer to home: a common perception seems to be that the Archimedeans is only for those who got a top-half First last year. This should not be, and indeed is not, the case: the Archimedeans never has been and never should be solely for the elite, but this must be made clear to the student population at large. If you are interested and/or inspired by what is contained herein, then you will find more of the same in the Archimedeans. If you have never been to a talk or other event, then why on earth not? If you regard yourself as too stupid or not knowledgeable enough, then this objection can easily be refuted: although you may well meet some of the top undergraduate mathematicians there are many of more average abilities and in any case they are, after all, normal human beings. And, furthermore, the more mathematics you experience, the better you'll get; everyone was young and ignorant once. If you are asking yourself, "Am I good enough for the Archimedeans?" then this is the wrong question: you should be asking "Is the Archimedeans good enough for me?" and the answer at least from this quarter must be yes.

If this issue of *Eureka* has whetted the appetite of just one person to think about mathematics in general and to become involved with what is, after all, the society for *all* Cambridge students interested in mathematics then I shall be able to consider it a success. But enough of the Editor's introspection. *Eureka* is a celebration of all that is best in the Archimedeans (or at least that is the idea), and I shall keep you from it no longer. On with the mathematics!

Colin Bell

Acknowledgements

Despite what it might seem, *Eureka* is not a one- or even two-man show: there are many behind the scenes without whom it would not be possible, and we have been blessed with extensive enthusiastic and capable help. In particular, we are indebted to our predecessor, Michael Greene, who apart from the work on his own article gave a large amount of help, advice, encouragement, and (perhaps above all) cups of tea. Further thanks go to the Business Manager, Peter Mennie, the Subscriptions Manager, Peter Benie (no relation), and the various people involved with distribution and the more mundane jobs of getting *Eureka* printed and sent out. We were also aided and abetted by the army of proof-readers who have ensured that the number of mistakes has been kept within reasonable bounds; any remaining ones are of course ours. Colin Yakeley and Mark Owen are worthy of mention for suggesting alternative titles to two of the articles. And last, but by no means least, we must thank all those who have contributed to this *Eureka*—without you it would be very much shorter and considerably less interesting.

Colin Bell and Mark Walters
February 1994

The Archimedeans

Patricia Smart (Chronicler 1992–1993)

Well, another year has passed in the Archimedeans' calendar ...

The year began after the AGM in March, but not very much happened in the life of the Archimedeans between then and the May Week events. However, before then we did have the Lecturer of the Year competition, which was one of the first events organised by the new committee—this was won by Prof. Frank Kelly. The Archimedeans had a variety of social events during the latter part of the Easter Term. Once again, we had a very enjoyable Garden Party in Clare Memorial Court in fine weather (for many people this was the highlight of the term). The Barbershop Subgroup was in full voice and provided entertainment alongside the refreshments which included the traditional strawberries and cream. The Punt Trip to Grantchester was also a pleasant occasion. Some members of the Archimedeans took up the challenge from the Invariants (the Oxford Mathematical Society) to compete in a croquet match, and lost, possibly due to their lack of experience at playing on flat croquet lawns.

After the summer vacation, we got into the swing of the routine of speaker meetings, College Society meetings, business meetings, and other activities. First of all, though, we had to join in the crush at Kelsey Kerridge for the Societies' Fair where we had a stall. This was followed later in the same week by our squash. Our membership drive was very successful, both at the beginning of the Michaelmas Term and during the course of the year.

We held seven main speaker meetings during the Michaelmas and Lent Terms. The number of people attending these meetings was good: indeed the Cockcroft Lecture Theatre was filled to its maximum capacity (without breaking fire regulations) for Prof. Stephen Hawking's talk on *The Seeds of the Universe*. Prof. D. Singmaster of London, Prof. G. Joseph of Manchester, Mr N. Lord of London, Dr T. Hawkes of Warwick, Prof. M. MacCullum of London, and Dr T. Gardiner of Birmingham also gave interesting talks this year, between them covering a wide spectrum of mathematical areas. Prof. Singmaster produced, after his talk, a suitcase of interesting and unusual puzzles, and many members stayed for a 'hands-on' session.

The College Mathematical Societies continued to have meetings, giving the opportunity for many mathematically filled hours. Various subgroups have been in operation this year, some lasting the whole year (notably the Puzzles and Games Ring and the Music Appreciation Subgroup), others creeping quietly into non-existence. The Barbershop Subgroup met during the Easter Term, and the Mathematical Models Subgroup met a few times during the year.

Other activities included the annual Puzzle Hunt in the Michaelmas Term and the Problems Drive in the Lent Term, both of which proved to be enjoyable events.

Financially, we are in a much better position than we were a few years ago. The bookshop has continued to prosper. *Eureka* **51** was published early on in the year.

In March 1993, the Committee passed the baton to its successor. In keeping with tradition, I shall end this report by saying that the Archimedeans had a successful year. We hope that next year will be even more successful.

Train Sets

Adam Chalcraft and Michael Greene

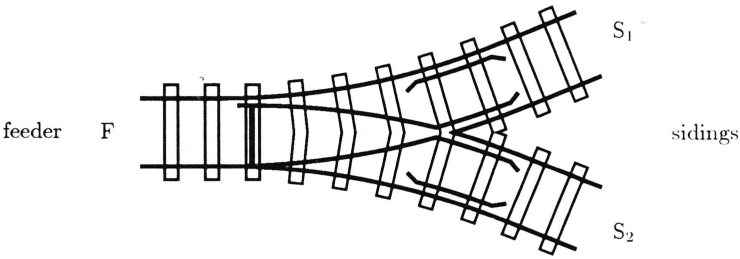
Suppose we have a train set—large stocks of straight and curved track, bridge-building materials, different kinds of sets of points, and a single engine, say clockwork for the sake of nostalgia. What can we do?

This depends, of course, on what kinds of point we have (for simplicity, a *set of points* or *pair of points* will be called a *point*). We use the following general notation:

Track will be drawn as a single line: 

Bridges will be drawn simply as crossovers: 

Lazy points



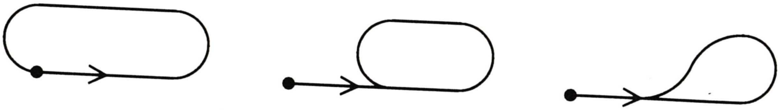
Lazy points will be taken to behave in the way you might expect from studying the physical piece of track above. With the point in the state shown, a train entering from F will leave by S_1, and vice versa, and a train from S_2 will change the point to its other state (with the central A-shaped piece up rather than down) and leave by the feeder. If the state has changed, the behaviour is exactly the same with S_1 and S_2 swapped.

When it is important which siding is *live*, i.e., which siding a train entering the feeder will leave by, a dot will be added to that siding; the other siding will be called *dead*. The lazy point above will therefore be represented by one of these:

Suppose we have a finite *layout* (a connected arrangement of *sections* of track between feeders and sidings, with no ends, or 'buffers') with lazy points. Define the *state* of the system to be the setting of each of the lazy points, the section of track which the train T is on, and the direction in which T is travelling; then there are only finitely many states. We allow the train to run until the states cycle, and then remove any unused pieces of track from the layout and replace partially-used points by pieces of track. We would like to classify all possibilities for the *stabilised* layout which remains.

Follow the train from the middle of some section of track.

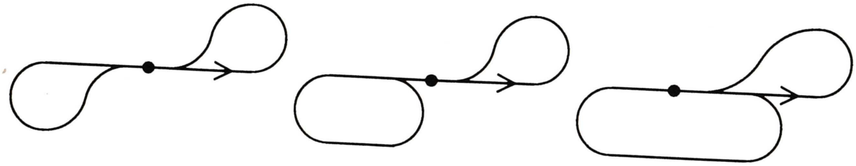
The layout is finite. Consider the first time T reaches track it has already traversed—then we have one of the following:



(a) (b) (c)

(Note that, although T may have crossed bridges on its route so far, these are the only essentially different connections.) Case (a) is a simple loop and is certainly a solution. In (b), T will continue around the loop and so had not settled down to a cycle of states; this case therefore does not occur.

In the last case, we need to follow the train round further. Again there are three essentially different cases:



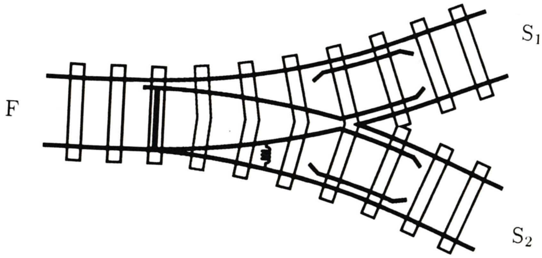
(d) (e) (f)

Both (e) and (f) fall into a smaller cycle and simplify to case (a). (d), however, is a solution, and is traversed in an unexpectedly complex way (and its behaviour shows that these are all the cases).

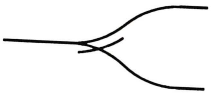
This classifies all finite stabilised lazy point layouts. □

We now introduce another natural kind of point.

Sprung points



Suppose we take a lazy point and attach a spring under tension as shown. Here, although a train from S_2 will push the central piece up as it runs over the point, the piece will not stay there, and any subsequent train from F will leave by S_1. This means that the live siding never changes. We call this a *sprung* point and use the notation



to represent it; the extra arc should be thought of as guiding a train from the feeder along the correct siding.

Suppose we have a finite layout with lazy and sprung points. As before, allow the train to run until the layout stabilises; then every point which was only going to change state finitely often will not change again.

We simplify the layout in three stages, without altering the route of the train.

Stage 1 We first simplify the network by removing any point and section of track which is unused, and replacing a point by track if one of its sidings is unused.

Now every section of track is either used infinitely often in both directions, or infinitely often in one direction and not at all in the other. The former sort of track is called *two-way*, and the latter *one-way*.

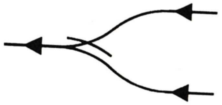
Stage 2 For the next stage, we regard sprung points as simpler than lazy points, and simplify the network by replacing every lazy point with a sprung point if this can be done without altering the route of the train.

If a lazy point has a siding through which the train never leaves, then the point may be replaced by the sprung point which does not allow the train to leave through that siding.

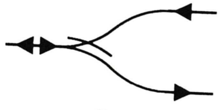
If a lazy point has a siding through which the train never enters, then the point will only be able to change state once. It will therefore not change state at all, and may as well be a sprung point.

Having done this, both sidings on every lazy point are two-way, and so therefore is the feeder.

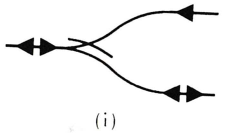
Stage 3 Now consider the sprung points. The dead siding must be one-way. There are three remaining possibilities, shown in cases (g)–(i).



(g)



(h)



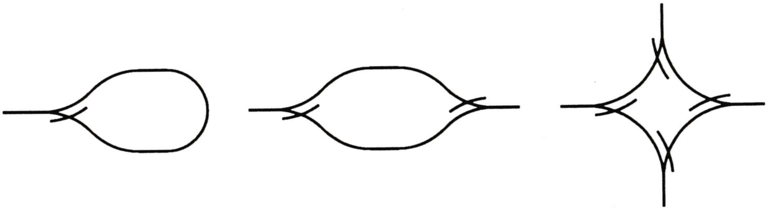
(i)

If we count the number of inputs and outputs in each of these cases, we find that (g) and (i) have one more input than output, whereas (h) has equal numbers of each. By Stage 2, all the lazy points have equal numbers of inputs and outputs. Across the entire layout, the number of inputs must equal the number of outputs, and so cases (g) and (i) cannot occur.

The sprung points must therefore all be as in case (h), and be joined to each other in circles, live siding joined to dead siding. These circles are called *roundabouts*, and the following figure (overleaf) shows roundabouts with 1, 2, and 4 sprung points. A train entering a roundabout must turn left, and then leave by the next exit.

The layout now consists of roundabouts and lazy points. Every section of track is two-way except the track inside a roundabout. This completes the classification of stabilised layouts made from lazy and sprung points. □

There is one more observation to be made, however. Suppose Stage 1 of the simplification has just been done, and we are about to embark on Stage 2. The only thing that



Stage 2 does is to turn some of the lazy points into sprung points. When this is over, all the sprung points look like case (h). Therefore any lazy point which was turned into a sprung point in Stage 2 had a one-way and two-way diagram as in case (h), but this is impossible for a lazy point.

The conclusion is that Stage 2 did not do anything. After Stage 1 was complete, all the lazy points were already two-way on both sidings and the sprung points were all arranged into roundabouts.

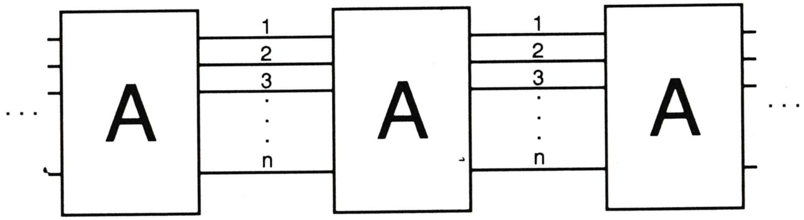
Can we build a computer?

We would like to know how complicated the route of the train can get. At one extreme, we might be able to simulate a general computer. The usual formulation of this is a *Turing machine*, which works as follows. Imagine a box of hardware which is able to move along an infinite 1-dimensional tape on which is written an infinite discrete string of 0s and 1s. The box can be in any of a finite number of *states* at any point on the track. It has a list of instructions as to what to do for any state looking at either digit. Each of these instructions is of the form

> leave unaltered or change the current digit,
> move one place to the right or left along the tape, and
> enter a particular state before performing the next instruction.

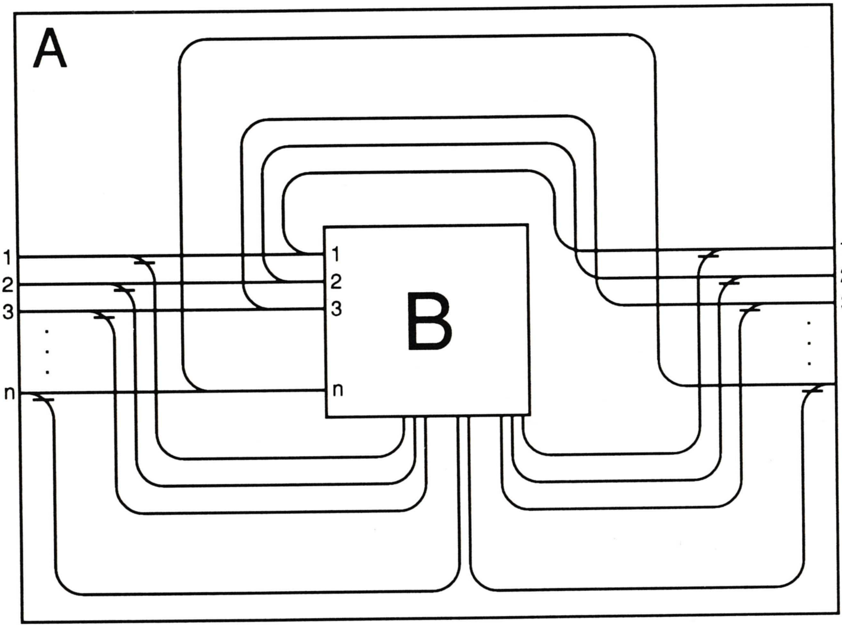
Now the process repeats. It is 'well-known' that any algorithm which can be run on an ordinary computer can be encoded to run on such a device.

In the search for such a model, we shall try to find a cell A which simulates the behaviour of the box at one point on the tape, and then plug these together to represent the whole tape:



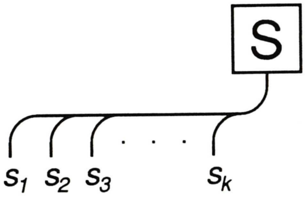
If T is in a particular copy of A, this will correspond to the box being at the equivalent position on the tape. If T travels from one copy of A to an adjacent copy along a piece of track labelled k, this will correspond to the box moving one step in the same direction along the tape, about to start the next action in state k. Note that this means there are two fundamentally different kinds of states which A can be in, representing either a 0 or a 1 at the corresponding point on the tape.

Now examine A. Around the outside, the tracks are used as both input and output, and the Turing machine does not remember which direction it came from to execute its next action. This track can therefore be used for the outer shell:



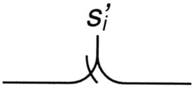
The B shown has n inputs on the left side and a selection of outputs coming out of the bottom. The outer shell sends the train out left or right on the appropriate line, as required. The sprung points ensure that the uses of the lines for output and input are split cleanly apart.

It will be useful to introduce the concept of a *subroutine*. This is a section of track with one input, which is also used as the output. When a train is sent in, it wanders around inside and eventually comes out again on the same section. We can give a subroutine more than one input:



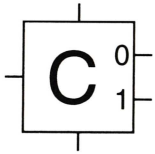
(*)

A train entering this diagram at s_i will set the lazy points to the route it took, enter and leave S, and then return along the same route—in particular it will return to s_i. We can *call* S by connecting s'_i (overleaf) to s_i:



A train travelling from left to right in this diagram will call S and continue.

Suppose we have built a subroutine S, with entry points s_1, s_2, …, s_k as above, which changes the digit represented by the current copy of A from 0 to 1 or vice versa. Suppose further that we have a piece of layout which looks like



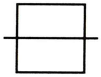
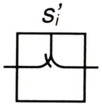
and behaves as follows:

i) a train from the left comes out on either 0 or 1 depending on the digit at the current point on the tape, and

ii) a train from the top changes the state recorded and comes out the bottom.

We now connect these up (see opposite).

Each 

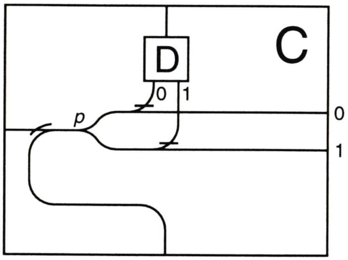
is either a  or a 

.

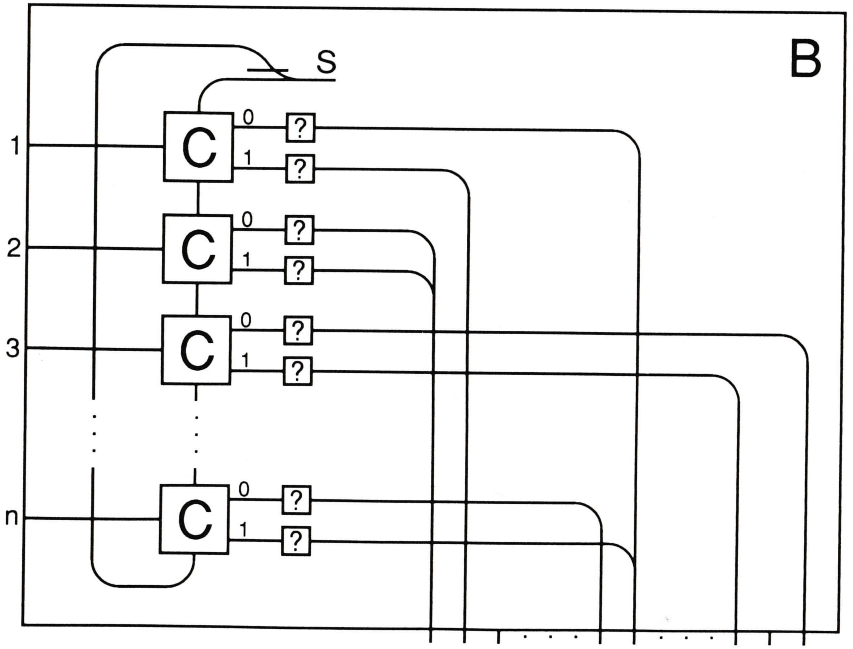
If T enters B on track k and passes through one of the Cs, we know exactly what to do to simulate the original Turing machine: we know whether to change the digit recorded in this position on the tape (that is, whether to call S), and we then know which way to leave this cell and on which track. Thus this diagram encodes the program. All we need to do is give S multiple entry points (as in (*)) and connect s_i to s_i' for each i, and the track will behave as wanted. (For neatness, we also should remove the unused track in A.)

Now to construct C. Let us suppose the existence of a *distributor* D, which has two outputs, 0 and 1, and one input. Trains sent in the input come out of 0 and 1 alternately. The reader is strongly advised to try constructing one of these—it makes an infuriating problem.

Suppose for the moment that such a piece of track exists. Take C to be this:



When we store our program's input data on the tape, we must set the (lazy) point p to



direct trains from the left out of the appropriate exit on the right, and make sure that the first train to enter by the top will come out of D on the correct side to flip p. From then on, C will behave as we want it to, and we nearly have a Turing machine—all we need is a distributor.

Distributors are impossible

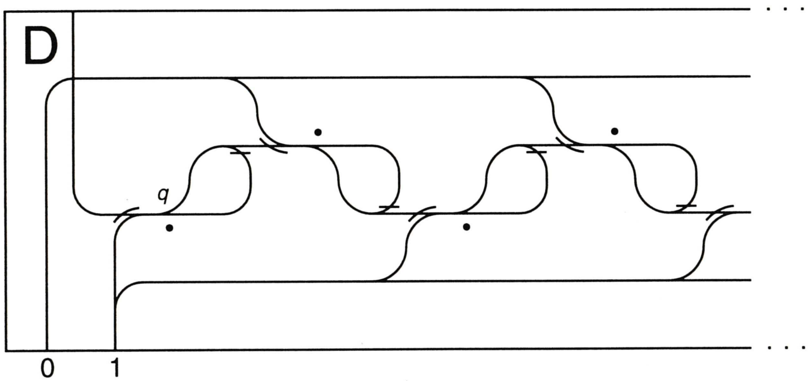
Suppose D is a distributor made from a finite number of lazy points and sprung points, with one input and two outputs as shown in the first figure below. We can add one extra sprung point and some extra track to get the network shown in the second figure.

Now run this network until it stabilises. Since D will work as a distributor for ever, the stable form of D is still a distributor, so the network stabilises to a form in which the extra sprung point is being used like case (g) earlier.

Unfortunately, we showed there that case (g) could not happen, so distributors are impossible. □

A distributor

Now we shall exhibit a distributor. The construction is rather naïve:

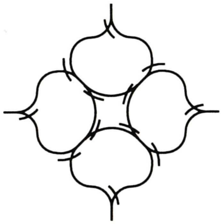


All lazy points are initially set to send trains across horizontally, except perhaps the one labelled q, which should be set the other way exactly when D is in a copy of A representing a 1 on the tape. Of course, D can be bent round to point upwards to avoid too many bridges or arbitrarily short sections of track.

This completes the construction of a Turing machine simulator. □

One further result

The reader may object to the impurity of using bridges. If we have sprung points, it is easy to simulate a bridge in the plane:



Conjectures

1. The Turing machine given here suffers from complexity problems: an algorithm which uses bits of the tape over and over again will get slower as more distant parts of the distributors are used (judging time by counting the number of points T crosses). We conjecture that no Turing-powerful layout has the same complexity for every algorithm as a conventional Turing machine.
2. Suppose we have a *random* point, which sends trains in a siding out of the feeder and trains in the feeder out of one of the sidings with equal probability, independent of previous behaviour. It seems likely that even with lazy, sprung, and random points we cannot build a finite distributor. (We now only require a box which always sends the train out of the correct exit if it sends it out at all, and does so in finite time with probability 1.)
3. A weaker version of 2: given lazy, sprung, and random points, we have been unable to construct a *partial* distributor, which never sends T out of the wrong exit, has for any n a positive probability of working at least n times in a row, but may have a chance of trapping the train inside after some time.

Representations as Sums of Squares

Mark Walters

This is an extension of a problem which appeared in the 1992 International Mathematical Olympiad. The problem is:

Given a positive integer n, find $S(n)$ which is defined to be the greatest integer such that for all $k \leqslant S(n)$, n^2 can be written as the sum of k positive square numbers.

We start with a definition: a *deficient square* is a number which is one less than a positive square. The following result is immediate:

LEMMA 1. *n^2 is expressible as the sum of k positive squares if and only if $n^2 - k$ is expressible as the sum of at most k deficient squares.*

PROOF. The proof is obvious:

$$n^2 = \sum_{i=1}^{k} a_i^2 \quad \Leftrightarrow \quad n^2 - k = \sum_{i=1}^{k}(a_i^2 - 1).$$

If we start on the right hand side with fewer than k deficient squares, we can add in the required number of zeros—0 is a deficient square. □

From this we can immediately find a global upper bound: since 13 is not expressible as the sum of deficient squares, the only relevant ones being 3 and 8, we can deduce that n^2 is not representable as the sum of $n^2 - 13$ positive squares, and hence $S(n) \leqslant n^2 - 14$ for all $n \geqslant 4$ (clearly for $n < 4$, $S(n) = 1$). The next lemma tells us that this is the best possible bound of this form:

LEMMA 2. *Any integer $n > 13$ except 28 can be expressed as the sum of five deficient squares. Moreover 28 is expressible as the sum of six deficient squares.*

PROOF. If $n \geqslant 165$ then $n^2 + 5 - 169 > 0$ and so is expressible as the sum of at most four squares (Lagrange's Theorem). Therefore, since

$$169 = 13^2 = 12^2 + 5^2 = 12^2 + 4^2 + 3^2 = 10^2 + 8^2 + 2^2 + 1^2,$$

$n+5$ is representable as the sum of exactly five squares, by combining the representation with the appropriate decomposition of 169. Therefore n is expressible as the sum of five deficient squares.

If $n \leqslant 164$ then the result is also true. I have not found an elegant proof of this but it is easy to verify by a direct search.

Finally note that $28 = 8 + 8 + 3 + 3 + 3 + 3$. □

LEMMA 3. *For any integer n and for all k such that $5 \leqslant k \leqslant n^2 - 14$, n^2 can be written as the sum of exactly k positive squares.*

PROOF. For $5 \leqslant k \leqslant n^2 - 14$, we have $n^2 - k > 13$, and so we can apply the above lemma: $n^2 - k$ is always expressible as the sum of k deficient squares by adding zeros. (The case $n^2 - 5 = 28$ obviously can't occur.) Now using Lemma 1, n^2 is expressible as the sum of k squares. □

Hence the problem reduces to that of whether n^2 is the sum of k positive squares for $2 \leqslant k \leqslant 4$, since it is the sum of k positive squares for $k = 1$ and for $5 \leqslant k \leqslant n^2 - 14$, and not for $k = n^2 - 13$. We consider $k = 4$ first:

LEMMA 4. *If a square n^2 can be expressed as the sum of two positive squares then it can be expressed as the sum of four positive squares.*

PROOF. Clearly if $n^2 = a^2 + b^2$ then n, a, b form a Pythagorean triple. Therefore they are of the form

$$n = m(q^2 + r^2), \qquad a = m(q^2 - r^2), \qquad b = 2mqr.$$

In particular,

$$n^2 = m^2(q^4 + 2q^2r^2 + r^4) = (mq^2)^2 + (mqr)^2 + (mqr)^2 + (mr^2)^2. \qquad \square$$

So the only remaining cases to consider are $k = 2$ and $k = 3$. The case $k = 2$ is relatively well-known (see any standard number theory textbook, for example [**1**], for this and the other results quoted): an odd prime number is the sum of two squares if and only if it is of the form $4t + 1$. So if n has a factor p of this form with $p = q^2 + r^2$ we can write $n^2 = (2qrn/p)^2 + ((q^2 - r^2)n/p)^2$ using the Pythagorean triple formulae above. The converse is a little harder. Another relatively well-known result is that the number of ways of writing $m = x^2 + y^2$ is $4(d_1 - d_3)$ where d_q is the number of divisors of m congruent to q (mod 4). A small amount of thought gives us that n^2 is the sum of two positive squares if and only if n has a prime factor of the form $4t + 1$. This leaves us the case $k = 3$.

LEMMA 5. *If n^2 is expressible as a sum of two positive squares and $5 \nmid n$ then n^2 is expressible as the sum of three positive squares.*

PROOF. Consider the equation $n^2 \equiv a^2 + b^2 \pmod 5$. Since a square is congruent to 0 or ± 1 modulo 5 and $n \not\equiv 0$ modulo 5, either a or b must be congruent to 0 mod 5. Without loss of generality a is, whence $a = 5r$ and

$$n^2 = (4r)^2 + (3r)^2 + b^2. \qquad \square$$

LEMMA 6. *If $n = 5.2^m$ then n^2 is not expressible as the sum of three positive squares.*

PROOF. Suppose

$$(5.2^m)^2 = a^2 + b^2 + c^2.$$

Then (working modulo 4) a, b, and c must all be even: say $a = 2a'$ etc. We then have

$$(5.2^{m-1})^2 = a'^2 + b'^2 + c'^2.$$

By induction,

$$5^2 = A^2 + B^2 + C^2, \qquad \text{for some } A, B, C$$

and this is clearly impossible. $\square$

The only remaining case is $n = 5r$ and $r \neq 2^m$. For this result I am indebted to Gareth McCaughan, who spotted it as an exercise in [**1**] (exercise 5, chapter 9) and provided the following solution. Lemma 7 is of a more technical nature and the proof can be safely skipped if desired.

LEMMA 7. *If $m \equiv 1 \pmod 4$ then there are x, y, z such that $x^2 + y^2 + z^2 = m$ and $(x, y, z) = 1$.*

PROOF. By Dirichlet's Theorem, there are infinitely many primes in the arithmetic progression with first term $(3m-1)/2$ and common difference $4m$. So, let p be a prime of the form $4mu+(3m-1)/2$. Write $c = 8u+3$, and notice that $2p = cm-1$. It is now easy to show (using quadratic reciprocity and related results) that $-c$ is a quadratic residue modulo p; since everything is a quadratic residue modulo 2 the Chinese Remainder Theorem implies that c is a quadratic residue modulo $2p$. Let a be a witness to this: that is, suppose $2p \mid a^2+c$. This makes $(a^2+c)/(cm-1)$ an integer.

Now consider the ternary quadratic form

$$F(X,Y,Z) = \frac{a^2+c}{cm-1}X^2 + 2aXY + (cm-1)Y^2 + 2XZ + mZ^2.$$

It's easy to check that this is positive definite and has determinant 1, and all its coefficients are integers. There is a well-known theorem to the effect that a form with these properties and degree at most 7 must be equivalent to $X^2+Y^2+Z^2$, where "equivalent" means that there is an integer matrix A of determinant 1 such that (identifying F with the corresponding 3×3 matrix) $AA^T = F$. This implies that $a_{31}^2 + a_{32}^2 + a_{33}^2 = m$, so we shall be done if $(a_{31}, a_{32}, a_{33}) = 1$. Let B be the inverse of A (which of course also has integral entries), so that $BFB^T = 1$, implying that $F(b_{11}, b_{12}, b_{13}) = 1$; but any common factor of a_{31}, a_{32}, a_{33} must also divide b_{11}, b_{12}, b_{13}. Hence indeed $(a_{31}, a_{32}, a_{33}) = 1$. □

With that result, we can finish the problem off relatively easily.

LEMMA 8. *If $n = 5r$ where r is not a power of 2, then n is the sum of three positive squares.*

PROOF. r must have an odd prime factor. If we can express $(5p)^2$ as the sum of three positive squares then multiplying all of them by $(n/5p)^2$ will give us a solution. There are three cases to consider:
If $p = 5$ then we have $625 = 20^2 + 12^2 + 9^2$.
If $p \equiv 1 \pmod 4$ but $p \neq 5$ then p is the sum of *two* non-zero squares and so p^2 is: the result follows from Lemma 5.
If $p \equiv 3 \pmod 4$ then p^2 is the sum of three squares with no common factor (Lemma 7); so at most one of them can be 0. If one of them *is* 0 then p^2 is the sum of two positive squares, which is a contradiction. □

We can summarise the results as follows:

THEOREM *$S(n)$ takes only the values 1, 2, and $n^2 - 14$*
(i) *If n does not have a prime factor of the form $4k+1$ then $S(n) = 1$.*
(ii) *If n is of the form 5.2^m then $S(n) = 2$.*
(iii) *If n does have a prime factor of the form $4k+1$ but is not of the form 5.2^m then $S(n) = n^2 - 14$.* □

Reference

[1] H. E. Rose. A Course in Number Theory (Exercise 5, chapter 9).

Fractional Iterations of Functions

Robin Michaels

The idea of iterating a function is commonly used in mathematics. Many techniques of numerical analysis involve iterating a carefully chosen function. However, the structure of the iteration process itself is of some interest. t_f

Under certain conditions it is clearly possible to iterate a function any positive integral number of times. If f is a function such that

$$\text{codomain}(f) \subseteq \text{domain}(f) \tag{1}$$

then as usual we shall write

$$f^n(x) = \underbrace{f(f \ldots f(x) \ldots)}_{n \text{ times}}$$

(unless otherwise stated, we shall always assume that $n \in \mathbb{N}$).

The restriction on the domain and codomain of f is important, since otherwise it will be impossible to evaluate $f^n(x)$ if $n \geqslant 2$ and $x \in \text{domain}(f)$ but $f(x) \notin \text{domain}(f)$. In other words, the restriction is necessary to ensure that $\text{domain}(f) = \text{domain}(f^n)$.

If an inverse function of f exists, we will write it as f^{-1}. It would clearly be convenient if it were possible to define $f^{-n} \equiv (f^{-1})^n$, but this is only possible if

$$\text{codomain}(f^{-1}) \subseteq \text{domain}(f^{-1}),$$

or equivalently

$$\text{domain}(f) \subseteq \text{codomain}(f). \tag{2}$$

We further define $f^0(x) = x$ (in other words $f^0 = I$, the identity function).

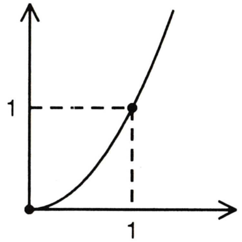
Looking at (1) and (2), we can see that, in order for f^m to be well defined for all $m \in \mathbb{Z}$, we need to have

$$\text{domain}(f) = \text{codomain}(f),$$

and since both f and f^{-1} exist over $\text{domain}(f)$, f must be a bijection.

So far only general abstract properties of the function have been mentioned, but, from now on, only continuous functions on some interval of $\mathbb{R}$ will be considered. In fact, even more restrictive conditions will be placed on these functions.

We call a function f *minimal* if, for every proper non-empty $S \subset \text{domain}(f)$, there exists an $x \in S$ with $f(x) \notin S$ or $f^{-1}(x) \notin S$. Equivalently, f is minimal if there is no proper non-empty $S \subset \text{domain}(f)$ with $S = \text{codomain}(f|_S)$. For example, consider the function $f(x) = x^2$, defined on $[0, \infty)$:



This is not minimal, since, for example, the image of $(0, 1)$ under f is precisely $(0, 1)$. However, it can be seen that if f is restricted to $(0, 1)$ then it becomes a minimal function.

It is clear that a continuous bijective function from an interval to itself must be monotonic, and we shall restrict attention to strictly increasing functions. So henceforth all functions to be iterated will be minimal increasing functions from some interval of $\mathbf{R}$ to itself.

The problem

Given a suitable function f, we now know how to define f^n for n an integer. (Note that in this section, for the sake of simplicity, we shall assume that f^{-1} exists; the argument carries through with the obvious restrictions when it does not.) It seems a natural question to ask if we can talk about f^t when t is rational, or even irrational. To ascertain whether this is a meaningful concept, we first need to clarify what properties we require f^t to have. For integral n and m the following hold:

$$f^n f^m = f^{n+m} = f^m f^n \quad \text{and} \quad (f^n)^m = f^{nm}.$$

It seems natural to extend these properties to these fractional powers of f (using fractional to mean non-integral) so we require:

$$f^s f^t = f^{s+t} \quad \text{and} \quad (f^s)^n = f^{sn} \quad \forall s, t \in \mathbf{R}, n \in \mathbf{Z}.$$

We also need to make sure that this definition of f^n corresponds to the standard one when $n \in \mathbf{Z}$. Two extra conditions are needed to ensure that these fractional powers behave as we would expect them to intuitively. The first is one of continuity—we require $f^t(x)$ to be continuous with respect to t for fixed x. The second is that $f^t(x) \neq x$ unless $f = I$ or $t = 0$. The last two requirements prevent attempts to use some unpleasant definition of f^t. It is not certain that all these conditions are independent but that question is not relevant to the following discussion.

We can now look at an example of a function satisfying these conditions. Let

$$f(x) = x^2, \quad x \in [0, \infty).$$

We have already shown that f is not minimal. However, if one considers the graph of $f(x)$, it is not hard to see that the following restrictions of f are minimal:

$$\begin{aligned} f_1(x) &= x^2, & x &\in S_1 = \{0\}; \\ f_2(x) &= x^2, & x &\in S_2 = (0, 1); \\ f_3(x) &= x^2, & x &\in S_3 = \{1\}; \\ f_4(x) &= x^2, & x &\in S_4 = (1, \infty). \end{aligned}$$

f_1 and f_3 are not exceptionally interesting functions; we will consider f_4 more closely. It is not hard to prove by induction that

$$f_4^n(x) = x^{2^n} \quad \forall n \in \mathbf{Z},$$

so the natural way to define $f_4^t(x)$ is

$$f_4^t(x) = x^{2^t} \quad \forall t \in \mathbf{R}.$$

It is not hard to check that this satisfies all our requirements. It is tempting to assume that for this particular function we have solved the problem completely, but we have not yet shown whether this is the only possible way of defining fractional powers: this issue will be considered later. If we now consider another function, say

$$g(x) = x^2 + x, \quad x \in (0, \infty),$$

then it is not easy to find any closed form for g^n with $n \in \mathbb{Z}$, let alone generalising it to fractional powers of g. A similar problem occurs with most other suitable functions, for example

$$h(x) = e^x - 1, \quad x \in (0, \infty)$$
$$\text{or} \quad j(x) = x^x, \quad x \in (1, \infty).$$

When we look at these cases the problem seems to be that h^n and j^n grow too quickly, in some sense, to be captured by our usual notation. However, even if no explicit formula for arbitrary powers can be found, there is no reason not to try and discover if arbitrary powers of these functions do in fact exist.

The transfer function

To tackle this problem, it will be useful to introduce some new notation. First, we define the shift functions as

$$s_\alpha(x) = x + \alpha, \quad x \in \mathbb{R}.$$

We will often write s for s_1, if we can do so without ambiguity. It is clearly possible to form arbitary powers of s_α in the following way:

$$s_\alpha^t = s_{\alpha t}.$$

This satisfies the conditions imposed on fractional powers of functions earlier. The problem of finding fractional powers of suitable functions over suitable intervals will be reduced to that of finding such powers of s, and this problem will be returned to later.

Secondly, the transfer function t_f is defined as follows (assuming for the time being that it exists):

$$t_f\colon \text{domain}(f) \to \mathbb{R} \quad \text{such that} \quad f^{t_f(x)}(x_0) = x.$$

The transfer function can be thought of as a sort of logarithm of its parameter with respect to function and a base point, x_0. To see intuitively what this means, assume that you have a graph of the inverse of the transfer function of f. This can be used to evaluate $f(z)$ in the following way. Look across from z on the y-axis until you reach the graph. Then see what the associated x-coordinate is, add one to it, and then find the value of t^{-1} for this value. It is not hard to see why this is $f(z)$. As an example, we take

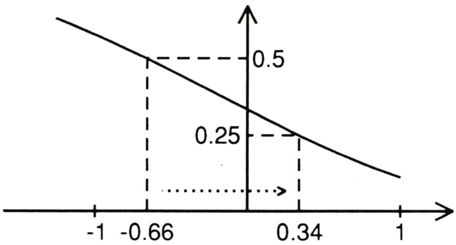
$$f(x) = x^2, \quad x \in (0, 1).$$

Let x_0 be $\frac{1}{3}$, so that

$$t_f(x) = \frac{\log\log(1/x) - \log\log 3}{\log 2},$$

and

$$t_f^{-1}(x) = (\tfrac{1}{3})^{2^x}.$$



Using the graph of t_f^{-1} above, it is possible to calculate, for example, $f(0{\cdot}5) = 0{\cdot}25$ by the above method.

Before we consider the relationship between transfer functions and fractional iterations, it is useful to prove some properties of minimal functions of the type referred to earlier:

PROPOSITION. For every minimal increasing function f from an interval of $\mathbf{R}$ to itself
(i) $D = \text{domain}(f)$ is either a point or an open interval;
(ii) in the latter case, $f(x) - x$ is of the same sign for all $x \in D$.

PROOF. It is clearly possible for D to contain a single element x with $f(x) = x$. If D is not a point or an open interval then it must be closed at at least one end. The case $D = (a, b]$ will be considered here—$D = [a, b)$ and $D = [a, b]$ can be dealt with similarly.

There must be some pair $x, y \in D$ such that $f(x) = b$ and $f(b) = y$. $x \neq b$ and $y \neq b$ since otherwise f is not minimal. Now consider $f(z)$ on the interval $[x, b]$. $f(x) = b > x$ and $f(b) = y < b$, so by the Mean Value Theorem, there is some c between x and b with $f(c) = c$ which contradicts the minimality of f.

Now if $D = (a, b)$ then either $f(x) < x$ for all $x \in (a, b)$ or $f(x) > x$ there, since otherwise, by the continuity of f, we can find some $y \in (a, b)$ with $f(y) = y$, and this contradicts the minimality of f. So $f(x) - x$ is of the same sign for all $x \in D$, as required. □

PROPOSITION. If $D = (a, b)$ then the sequences (a_n) and (a_{-n}), where $a_n = f^n(x_0)$, tend to a and b (not necessarily respectively) for all $x_0 \in D$.

PROOF. From the result proved above, we may assume that $f(x) > x$ for all $x \in (a, b)$. Then (a_n) forms an increasing sequence bounded above by b, so it must converge, to c say. Then $c \notin (a, b)$ since otherwise, by continuity of f, $f(c) = c$, which would mean f were not a minimal function. Thus since $c \leqslant b$, $A = b$.

Similarly (a_{-n}) is a decreasing sequence converging to a. (The proof can be extended to work for infinite open intervals: in this case the appropriate sequence diverges to infinity.) If we have $f(x) < x$ then $a_n \to a$ and $b_n \to b$. □

We are now ready to define the transfer function, and the associated fractional iterates. Set $t_f(x_0) = 0$ for some $x_0 \in D$; then we also need (rearranging the defining property of transfer functions) $t_f(f^n(x_0)) = n$ for all $n \in \mathbf{Z}$. We shall fill in the gap from $a_0 (= x_0)$ to a_1 by setting t_f to be any continuous monotonically increasing bijection from $[a_0, a_1]$ to $[0, 1]$. We also need $t_f(f^r(x_0)) = r$ for all $r \in \mathbf{R}$, which gives us $f^r(x_0) = t_f^{-1}(r)$ for $r \in [0, 1]$. More generally, we have

$$t_f(f(x)) = t_f(x) + 1 \quad \text{and} \quad t_f(f^{-1}(x)) = t_f(x) - 1.$$

and these provide a unique extension of t_f^{-1} to a function from $\mathbb{R}$ to D (by induction on the intervals $[a_n, a_{n+1}]$ which we know are disjoint). This function is clearly continuous everywhere, and has an inverse, for if not then $t_f(a) = t_f(b)$ with $a \neq b$. Since t_f^{-1} is continuous, we must have a local extremum between a and b. But if such an extremum exists then it must be at an image of x_0, in other words a_n for some n, since our definition of t_f precludes it occurring elsewhere. However, we know that $t_f([a_{n-1}, a_n]) = [n-1, n]$ and $t_f([a_n, a_{n+1}]) = [n, n+1]$, and since both pairs of intervals are disjoint, this case cannot happen either. So we have defined a function t_f, and from it we can derive fractional powers: we already know that $f^r(x_0) = t_f^{-1}(r)$ and we clearly want $f^r(x) = t_f^{-1}(r + t_f(x))$. We merely need to check the required properties: if we note that $f^r = t_f^{-1} s t_f$ then we get $f^s f^t = f^{s+t}$ and $(f^s)^n = f^{sn}$ easily; continuity of $f^t(x)$ with respect to t follows from the continuity at x_0; $f^t(x) \neq x$ is similarly simple. The final properties we need to check are domain(t_f) = domain(f) and codomain$(t_f) = \mathbb{R}$; these are also obvious from the derivation.

Note that if we already have fractional powers of f then a transfer function can be generated which is consistent with them: just let $t_f^{-1}(r) = f^r(x_0)$ for $r \in [0,1]$ and everything works.

Examples

(i)

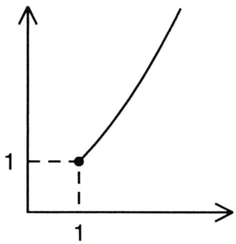
$$f : x \to x^2, \quad x \in (1, \infty).$$

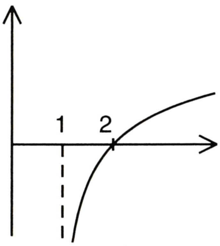
$$t_f(x) = \frac{\log\log x - \log\log x_0}{\log 2}$$

is acceptable and after a little manipulation

$$f^r(x) = x^{2^r}, \quad r \in \mathbb{R}.$$

$$f(x) = x^2 \qquad\qquad t_f(x) = \frac{\log\log x - \log\log 2}{\log 2}$$





(ii)

$$f : x \to \frac{2x}{1+x}, \quad x \in (0,1).$$

$$t_f(x) = \frac{\log(x/(1-x)) - \log(x_0/(1-x_0)}{\log 2},$$

so that

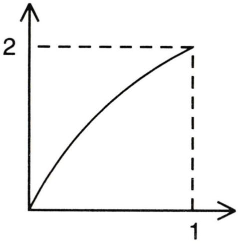
$$f^r(x) = \frac{1}{1 + ((1-x)/x)2^{-r}}.$$

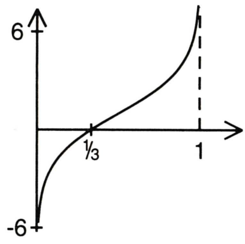
NOTE. In this case, f is a Möbius transformation, and in the transfer function the matrix diagonalisation

$$\begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$$

has appeared, albeit in a rather disguised form.

$$f(x) = \frac{2x}{1+x} \qquad\qquad t_f(x) = \frac{\log(\frac{x}{1-x}) - \log(\frac{1/3}{2/3})}{\log 2}$$





(iii) It is possible to start with the transfer function:

$$t_f(x) = x^3, \quad x \in (-\infty, \infty),$$

and then

$$f = t_f^{-1} s t_f,$$

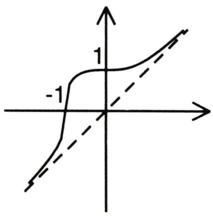
so

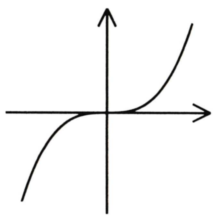
$$f(x) = \sqrt[3]{x^3 + 1},$$

giving

$$f^r(x) = \sqrt[3]{x^3 + r}.$$

$$f(x) = \sqrt[3]{x^3 + 1} \qquad\qquad t_f(x) = x^3$$





So, it seems that a way of producing fractional powers of all suitable functions has been found—but all is not so simple.

The Serpent in the Garden

Unfortunately, although it has been proved that fractional powers of suitable functions exist, it has not been proved that they are unique. In fact they are very non-unique.

If we can find a non-trivial transfer function for s, then any function can have several transfer functions: if

$$f = t_f^{-1} s t_f \quad \text{and} \quad s = t_s^{-1} s t_s,$$

then

$$f = t_f^{-1} t_s^{-1} s t_s t_f = (t_s t_f)^{-1} s (t_s t_f),$$

so $t_f{}' = t_s t_f$ is also a transfer function for f.

It has already been proved that a transfer function leads to a definition of fractional powers of a function which is unique up to the value of x_0, so different transfer functions will lead to different definitions of fractional powers.

$$t_s(x) = x + \tfrac{1}{7}\sin 2\pi x, \quad x \in \mathbf{R},$$

is, it is easy to check, a suitable non-trivial transfer function for s. However, there is no simple way to write t_s^{-1} in an explicit form, so this phenomenon is easily overlooked. In fact if p is a continuous function with $|2\pi p(x) + p'(x)| < 1$ for all x, and p is periodic with period 1, then

$$t_s(x) = x + p(x)\sin 2\pi x, \quad x \in \mathbf{R},$$

is a suitable transfer function for s. Thus fractional powers of any suitable function are not uniquely defined. The set of possible transfer functions for a given function can be thought of as an equivalence class, and the set of transfer functions for s form a group, capturing the structure of the equivalence classes.

Other uses for transfer functions

Unexpectedly, transfer functions have proved useful in solving a certain class of functional equation. If we have

$$\sum_{n=1}^{N} a_n f^n(x) = 0,$$

a functional equation in f, then, if we make the unwarranted assumption that f has a transfer function, this leads to:

$$\sum_{n=1}^{N} a_n t^{-1} s_n t = 0,$$

or

$$\sum_{n=1}^{N} a_n t^{-1}(x+n) = 0.$$

This is a difference equation for t^{-1} and may well be soluble by standard methods. This approach is not certain to work, but may lead to unexpected results. For example:

$$2f^3(x) - 7f^2(x) + 7f(x) - 2x = 0$$

transforms to

$$2t^{-1}(x+3) - 7t^{-1}(x+2) + 7t^{-1}(x+1) - 2t^{-1}(x) = 0$$

and since this factorises as

$$2y^3 - 7y^2 + 7y - 2 = 2(y - \tfrac{1}{2})(y-1)(y-2)$$

it has solutions of the form

$$t^{-1}(x) = a2^x + b2^{-x} + c, \quad a, b, c \in \mathbf{R}$$

and so

$$t(y) = \log\left[\frac{(y-c) \pm \sqrt{(y-c)^2 - 4ab}}{2a}\right] / \log 2.$$

If this is defined, then we can find a solution for f. Putting $a = b = 1$ and $c = 0$ we get

$$t^{-1}(x) = 2^x + 2^{-x},$$

$$t(y) = \log\left(\frac{x + \sqrt{y^2-4}}{2}\right) / \log 2.$$

Now

$$f = t^{-1}st,$$

so

$$\begin{aligned} f(x) &= 2\frac{x + \sqrt{x^2-4}}{2} + \tfrac{1}{2}\frac{2}{x + \sqrt{x^2-4}} \\ &= x + \sqrt{x^2-4} + \frac{1}{x + \sqrt{x^2-4}}, \end{aligned}$$

and this does indeed turn out to be a solution of the functional equation.

Another use of the transfer function—although in this case it is more of a transfer operator—occurs when attempting to define the fractional derivative of a function. The operator D, which differentiates a suitable function, can be written as $D = TST^{-1}$ where

$$\begin{aligned} S &: f(x) \mapsto f(x) + 1, \\ T(g) &= F^{-1}e^{\log(i\omega)g}, \end{aligned}$$

so that

$$T^{-1}(g) = \frac{\log Fg}{\log i\omega},$$

where F is the Fourier transform operator. This is just another way of noting that

$$F(Dg) = i\omega Fg,$$

a standard property of Fourier transforms. It would seem natural to define the rth derivative of g for $r \in \mathbf{R}$ as

$$F(D^r g) = (i\omega)^r Fg,$$

but it was shown earlier that in such a case the transfer function, or in this case operator, is not unique.

The situation has not been rigorously investigated but it seems likely that the same ambiguity that occurred with fractional powers of functions will occur with operators.

The transfer function can be seen as a way of conjugating a function to the shift function, and, looked at from this point of view, it is not so unexpected that it appears to be useful in different problems involving iteration theory.

Gess the Game

Paul Bolchover

Gess (pronounced "guess") is a game for two players that was originally conceived by the Puzzles and Games Ring of the Archimedeans as a generalised form of chess.

The rules

Gess is played on a grid of 18 squares by 18 squares. Each player starts with 43 identical stones, one player being *Black*, and the other being *White*. (When playing gess, the most practical arrangement is to use the squares of a go board, each player using 43 go stones.) Play proceeds alternately, with Black starting.

Each 3×3 square which is entirely or partially on the board, and which contains at least one of a player's stones and none of his opponent's is known as a *piece*, and is referred to by the location of its central square. (If the centre is off the board, it is referred to in the obvious way, and this accounts for the slightly odd manner in which the board is labelled.) Assuming there are no obstructing stones (see below), pieces move as follows.

The non-central squares in the piece determine the directions in which the piece is able to move; for example, if a piece has a stone in its central forward square, then it is able to move straight forwards; if it has a stone in its rear left square, then it is able to move diagonally backwards and left.

If the piece has a stone in the central square, then it may move an unlimited number of squares in any of the permitted directions, otherwise it may only move up to 3 squares in the permitted directions.

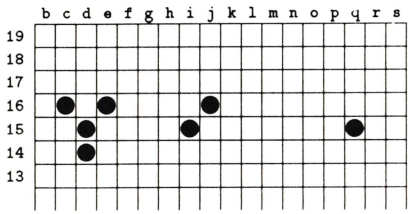


Figure 1

For example, in Figure 1, the piece at d15 may move an unlimited number of squares diagonally forwards and left, forwards and right, or straight backwards. The piece at j15 may move up to 3 squares, either forwards or to the left. The piece at q16 may only move backwards, up to 3 squares, whereas the piece at q15 may not move at all, since it does not have a valid direction in which to move. (It can move an unlimited number of squares but in no direction. It should also be noted that a 'move' that does not change the appearance of the board is not allowed.)

Note that you can have pieces which are partly off the board (for example, if you have a single stone in a corner, you may move it diagonally inwards for up to 3 squares), and pieces may also be moved partially, but not entirely, off the board, and any stones in the piece which end up off the board are removed.

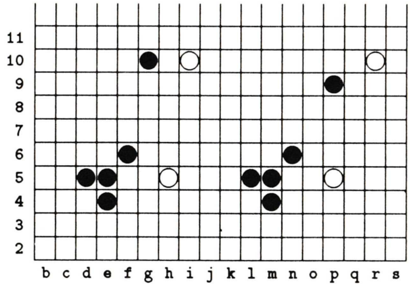


Figure 2

A piece may only continue moving in a given direction if at each stage its *footprint* (in other words the 3×3 square) doesn't cover any other stones of either colour. If stones are covered, then they are taken (removed from the board) and the move ends. To give an example, the piece at e5 in Figure 2 may move to i9 and take the white stone at i10, but in the similar position to the right, the piece at m5 may only move as far as p8, and if it does it takes the black stone at p9.

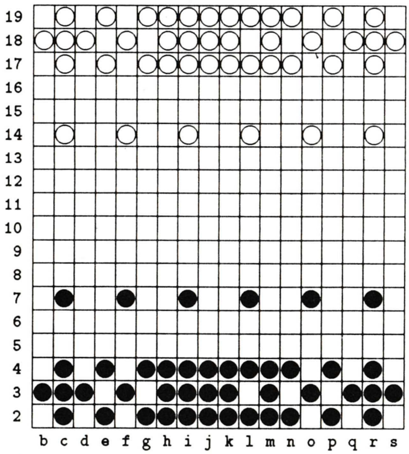


Figure 3: initial position

The initial position is shown in Figure 3. You should recognise it as having pieces analogous to rook, bishop, queen, king, bishop, and rook in that order, with six pawns in front. (There is obviously no possibility of producing a knight.) Note however that there are distinct differences; for example, the 'bishops' cannot move out until a space on one side or the other is cleared. A piece that moves like a king—a 3×3 square with all of its outer squares containing stones, but with no central stone—is known as a *ring*. The object of the game is to capture (or disable) your opponent's ring or rings; if at the end of a move either player has no ring then he loses: the player who has just moved being considered first, so you cannot use part of your ring to take your opponent's ring or rings. It is possible to have more than one ring at a time—indeed this may be considered desirable—and you may destroy one or more of your own rings provided that you still have at least one at the end of your move.

A sample game

Black: P. Bolchover
White: R. Michaels

1 f6–f7 — Preparing to form a powerful diagonal piece later after e3–e6.

... p15–m12 — Controls the centre, and blocks a long-range attack upon White's ring. It also builds up an attack on the central block. Both the last moves announce the players' intention to form a second ring.

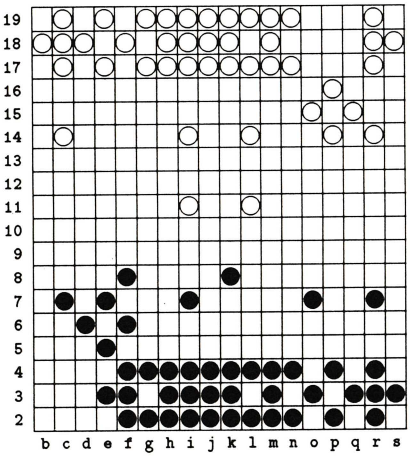
2 e3–e6 — Forms a diagonal piece attacking White's ring, and prepares to form a second ring. However, it doesn't do much defensively.

... p18–p15 — Defends the centre.

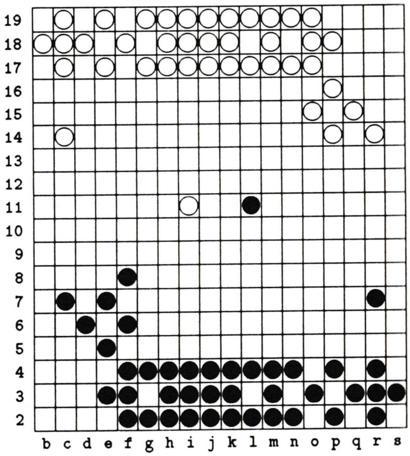
3 b3–e3 — Forms a ring. Note the many strong attacks forwards, but each of these attacks will disable one of Black's rings.

... e15–h12 — Opens a line against Black's second ring and reinforces the centre. b13–d15 might be a good move in the future.

4 m6–l7 — Opens a line for the forwards piece centred on m3.



After Black 4



After Black 8

... m12–j9 — White goes for the exchange but pins l14 to the White ring. It does, however, open a diagonal line for White, while blocking one for Black.

5 i6–i7 — h6–i7 or j6–k7 might be better.

... h15–k12 — Reinforces White's ring and opens up his queen.

6 i7–i10 — Opens up the centre for both players. It is unclear who has the advantage in this position. Overall, the move is probably bad, as it exposes Black's second ring.

... m15–j12 — The White pawns are starting to look slightly too far advanced.

7 o6–o7 — Attacking the pinned White pawn.

... s18–p18 — Forming a double ring.

8 p7–m10 Opening up an attack on the weak l11 square.

... k18–k12 Probably the best way of defending.

9 j3–j10

... k13–j12

10 j9–j10 j9–k10 would leave the Black ring under attack.

... q15–r14 Initiating a diagonal attack on the Black ring.

11 e6–h6 Blocking the attack.

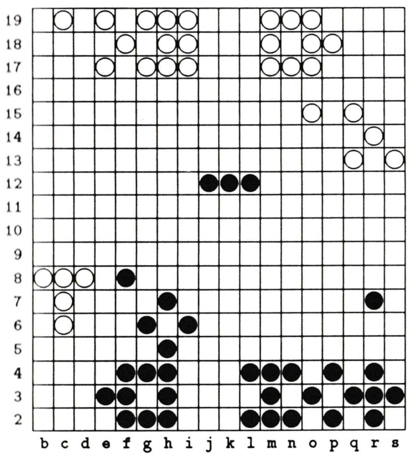
... j13–j12

12 j9–j10

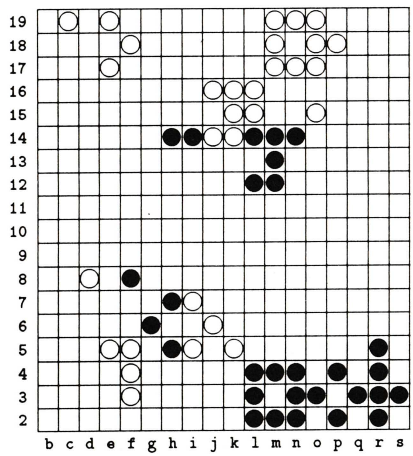
... c13–c15 Preparing for a flank attack and increasing the pressure on Black, who looks very weak defensively, but has a strong attack against White's ring.

13 j10–k11 Removing the obstructing stone.

... c17–c7 Note that this move is illegal, since the piece can only move as far as c8, but as neither player noticed it at the time the move stands.



After White 13



Final position

14 g3–j3 Forming a double ring for extra protection.

... h18–k18

15 k11–i13 Note that White could kill Black's rings, but in doing so would destroy his own ring, and hence would lose.

... k18–k15

16 m3–m13 Disabling the White piece centred on k15.

... b7–e4 Threatening moves such as e3–h3.

17 j3–m3

... r14–o11 Disabling the Black piece centred on m13.

18 r8–r6

... o11–j6 Mate.

A Mathematical Interlude

The Faculty

Here, for those who missed them, is a summary of the lectures given in Cambridge over the last two or so years.

Geometry

"Just as theologians have to ignore the possibility that God might not exist in order to get anywhere, we must ignore these difficulties as an act of faith."

"Even as a baby, when you were scribbling on a bit of paper, you were accustoming yourself to the Euclidean plane."

Methods

"We now equate degrees of awfulness on each side."

"Even an applied mathematician like me feels embarrassment in telling you such a bunch of lies."

"The trick is to interpret what you mean by a limit in a different way."

"Lecturers are told to show respect for their audience in the way they dress. What people don't realise is that wearing a tie and pushing a blackboard up and down, you break out in a sweat. So, now I've shown you my respect, I'm going to take off my tie."

Complex Methods

"We shall assume obvious things are true, and pick up the pieces afterwards when things go wrong."

Games and Logic

"Very few people are born category theorists."

"There aren't many arguments based on definition by contradiction."

"Now that's certainly something to do with *something*."

"Logically, 'and' is the same as 'or'."

Discrete Mathematics

"Why would I be wanting to do that? [pause] Yes, why *would* I be wanting to do that?"

"I've been informed that there are Satanic references in my lectures—if you haven't noticed, don't worry about it."

Lie Groups

"With that remark the following are equivalent:

I've lost my board rubber;

$\{1, v_1, \ldots, v_n\}$ are linearly dependent over $\mathbb{Q}$."

Functional Analysis

"All maths is either analysis or algebra."

"Analysis is great fun and algebra's all right if you like that sort of thing."

Category Theory

"Add brackets to taste."

Electromagnetism

"... because $\mathbf{E}$ is discontinuous only at discontinuities."

Quantum Mechanics and Special Relativity

"Deuterium has the same chemistry as oxygen."

"I never did Chemistry at school because I was too busy doing Greek."

"You could start from axioms and deduce mathematically but then you wouldn't know where you were going."

"We will now have a mathematical interlude."

"Pure mathematicians are still uneasy about calling it a delta *function* but physicists can call anything they want a function."

Knot Theory

"... whereas, believe it or not, the previous section was meant to have proofs in it."

Local Fields

"Until further notice $C = 2$."

Fluid Dynamics

"Now V is arbitrary, so the integrands are equal, assuming the volume is smooth enough, which it always is in Applied Mathematics."

Probability

"Now to work out the variance of the random variable Z, where Z has mean 0 and variance 1."

"A little bit of excitement comes from working it out as I go along ... Is that right?"

Quadratic Mathematics

"They were discovered in the second half of the eighteenth century, that is before the nineteenth century began."

"Above all, Number Theory is an experimental science."

Nonlinear Dynamical Systems

"We express x_n as a binary decimal."

Commutative Algebra

[Describing the course] "Just linear algebra with knobs on."

"Proof in one line—provided your line is long enough."

"I'll leave you to fill in the gaps—not that there *are* any gaps."

"I mean I'm an artist—I can't work with small bits of chalk."

"This 'p' is a Gothic p—sort of curly. It's not just that I can't write p."

"Again I've said nothing really—just words."

"Oh! I've lost all my differentials!"

"For those of you who like category theory: it's not your fault."

"... so you've been studying linear affine algebraic $\mathbf{R}$-varieties since the age of five."

Calculus and Methods

"We'll assume we're in three dimensions, just to make it general."

Differential Geometry

"Before passing on, there are two more things I'd like to talk about in this chapter."

Principles of Dynamics

"You know you're right because you know how to do it in two dimensions."

And finally ...

"Oh dear! I seem to have run out of time too quickly."

Problems Drive 1993

Timothy Luffingham

The Problems Drive is an annual Archimedeans institution. Teams of two enter and have to answer twelve questions; they receive one question every five minutes, and questions are taken away after ten minutes. Teams are given an additional ten minutes to finish off answering questions, make plausible guesses, or just fill the cross-number with random digits. Teams are traditionally represented by a symbol of their own choosing to provide a degree of anonymity.

This year, ten pairs entered, and the questions were easier and the scoring system more rational (in both senses of the word) than the previous year. The wooden spoon was awarded to the appropriately-named 'dustbin', alias one current and one future committee member, both from Clare. The prize for most amusing wrong answer, two creme eggs, was awarded to a pairing of half of the current editorship and a girl from Queens', for the answer 'POISON' for question 12, and the winners, with a score of $7\frac{47}{120}$, were 'stick figure and umbrella stand' (actually intended to be a team self-caricature), otherwise known as Colin Bell and Michael Fryers (respectively) of Trinity College, who will set questions for next year.

Answers can be found on pages 55–56.

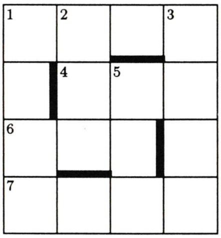
1. What are the next two terms of the following series?

(a) 1, 2, 3, 10, 99, ..., ...
(b) 1, 2, 3, 5, 6, 8, 9, 11, 14, 15, 18, ..., ...
(c) 1, 2, 3, 1, 2, 3, 4, 2, 1, 2, 3, ..., ...
(d) 1, 2, 3, 2, 2, 2, 3, 3, 2, 2, 2, 3, ..., ...

2. The priests of an ancient mystic religion counted, for reasons lost in the mists of time, in base 6. They regarded a number as male if its base 6 representation contained the holy digit 2; if not, it was regarded as female. The most holy numbers, known to the priests as harmonious numbers, were those which perfectly balanced the male and the female (that is, a number was harmonious if the set of integers between 1 and the number, inclusive, contained an equal number of male and female numbers).

The smallest harmonious integer was, therefore, 2. What was the largest harmonious integer? Please give your answer in base 10.

3. Solve the following crossnumber. None of the answers begins with a 0.



1A(base 8) = 3D(base 11)
1D(base 7) = 7A(base 6)
2D(base 4) = 4A(base 5)
2D(base 10) = 6A(base 7)
5D(base 14) = 6A(base 10)

4. I have a 6×6 square board, and I have placed a counter on some of the squares. To each square S without a counter on it, a number $n(S)$ has been assigned. I want to minimise the number of counters used, while making all the numbers $n(S)$ assigned to the uncovered squares different. What is the smallest number of counters I could use if:

(a) $n(S)$ is the total number of squares in a horizontal or vertical line through S which contain counters?
(b) $n(S)$ is the total number of squares in a diagonal line through S which contain counters?

5. A well known mathematics lecturer was murdered last night. You know that one of six suspects was responsible, and they have all made statements to you. You also know that three of the suspects are pure mathematicians, and the other three are applied. Unfortunately, you do not know which is which; you only know that pure mathematicians always give an odd number of true statements, and applied ones always give an even number.

The statements are:

A: I am an applied mathematician.
D and F are applied mathematicians.
B and E are pure mathematicians.

B: I did not commit the murder.
C and F are the same sort of mathematician.
The murderer is an applied mathematician.

C: I am a pure mathematician.
D did not commit the murder.
F did not commit the murder.

D: I did not commit the murder.
B did not commit the murder.
E is an applied mathematician.

E: A killed him.
B killed him.
F killed him.

F: C is an applied mathematician.
B is the murderer.
I was drinking tea with A and D all last night.

Who are the pure mathematicians, who are the applied ones, and who is the murderer?

6. I own a regular polyhedron K, with all its faces numbered differently. You and I play a game: I choose a vertex of K, you name n faces of K, and I then tell you which of those n faces contain the vertex I chose. You then guess which vertex I chose.

What is the smallest value of n which will always enable you to guess correctly if K is:

(a) a dodecahedron?
(b) an icosahedron?

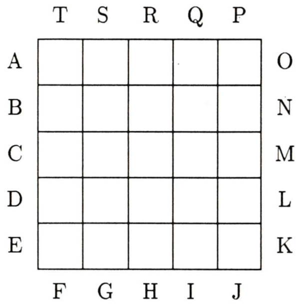
7. What is the length of the side of the smallest regular tetrahedron that can contain two spheres of unit diameter?

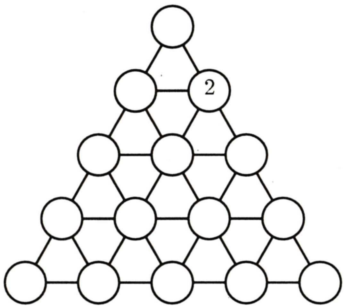
8. Four circles, with centres at $(1,1)$, $(1,-1)$, $(-1,1)$, and $(-1,-1)$, all have radius 2. What is the area of their common intersection?

9. Find a 3×3 matrix M, with entries 1, 2, 3, 4, 5, 6, 7, 8, and 9, which has a determinant of 1.

10. I have taught my pet mouse, Arthur, to run through grids of coloured squares that I set up for him, obeying certain rules. When he runs over a red square he turns left at the centre of the square, when he runs over a blue square he turns right at the centre of it, and when he runs over a white one, he continues straight on.

I have set up a 5×5 grid for Arthur to run through, and I notice that when he enters the grid at the points O, N, M, L, and K (see the diagram below), he leaves it at A, B, C, D, and E respectively. I now send him in at point R. Given that I have coloured the grid with red, blue, and white squares only, at which points would it be possible for Arthur to leave the grid?





11. Four friends are playing Number Boggle, an undeservedly little known variant of Boggle. It is played with the triangular board shown above, with a digit between 0 and 9 in each of the circles, and each player has to make a number by going from circle to adjacent circle along the joining lines, without going into any circle twice.

In one particular game, Alice found the number 36725, Betty found 2000064, Charles found 41821650, and David, who had never played before, only managed to find the number 9. The digit on the right hand side of the second row was a 2. Fill in the rest of the board.

12. Each of the following contains the letters of the surnames of three famous mathematicians. What are the names? (All accents have been ignored.)

(a) AAEEEEFILMMNNRRRTU
(b) CDDEEEGILLNNOOTUW
(c) AAACCCCEHLNOUWYYYY
(d) AAAADEGJMNNORRSSSUU
(e) AAACCEGIILNNOOOPRRST

CUMLL Annual Report

Mark Wainwright

Introduction

The Cambridge University Mathematical Limerick Laboratory has been particularly busy of late, and considerations of space preclude the present paper from giving more than a brief survey of the directions that research has been taking. Readers looking for a more detailed exposition should consult the Laboratory's own journal, and others in the field. A less technical account of some of the Laboratory's earlier work is contained in [**1**].

In this paper, we shall confine ourselves to the following: (1) progress on the Wainwright–Bending hypothesis; (2) the work of the Special Research Unit at CUMLL, set up to investigate thematic pairs of limericks; and (3) a few miscellaneous remarks and results.

1. The Wainwright–Bending Hypothesis

Students of MLs will recall the *Extended Wainwright-Bending Hypothesis*, that there would be a logical equivalence between the two approaches (technical and semantic) to writing limericks. Details will be found in [**1**]. In 1990 a highly technical paper was published in the *Journal of Mathematical Limerick Research* [**2**], proposing a proof of the hypothesis. The paper, which had a number of authors including the present author, was widely acclaimed in the literature; unfortunately four months after its publication a fatal flaw was found in the proof by the distinguished researcher in the field, A. Nomet.* Nomet noticed (in [**4**]) that the Limerick Existence Theorem (LET) was applied where, for rather surprising reasons, its conditions failed to hold; as a direct result he was able to construct an ingenious argument to show that the hypothesis is, in fact, false. However (see [**5**]) Nomet also showed that the original proof could be made to work for a slightly modified form of the hypothesis, and thus that suitable modifications of the two approaches are, indeed, equivalent. Readers may be interested in a graffito recently recorded on the door of one of the toilets in CUMLL:

NOMET'S CAUGHT WAINWRIGHT–BENDING

perhaps giving some idea of the general merriment induced in the Laboratory by Nomet's publication of his results.

2. The Special Research Unit

At the time of publication of [**1**], emphasis at CUMLL was on limericks with a symbolic expression, rather than text; an example given there was

$$\int_0^{\pi/3} \beta\,\mathrm{dc} = (\partial^2/\partial\eta\partial\nu)(|g|).$$

* some of whose results were published in a previous issue of this journal; see [**3**].

Even then, however, there was some interest in limericks expressing well-known or important results; the article ended with a query as to the possibility of expressing the four-colour theorem in limerick form. To put the cart before the horse, we start by giving the following result of Dr J. R. Partington, a valuable corresponding member of the Laboratory:

Remarked Appel to Haken one day,
"Do you know, it is true what they say:
That a graph requires four
Distinct colours, not more,
If it's planar. The proof needs a Cray."

Said the other, "We don't need a Cray—
Just an Apple (as Newton might say).
And so Hearken, old chap,
We shall colour each map
Making do with an Appel a day."

Other results relating to the four-colour theorem are given later in this article; what concerns us here is the felicitous timing of Partington's results. For they came shortly after the institution of CUMLL's Special Research Unit (SRU), to investigate pairs of linked thematic limericks precisely such as Partington's, and thus tied together two hitherto unrelated research projects.

An early problem studied by the SRU was the vexed question of Hardy's Taxi Number. The problem proved especially intractable; no serious results were found, but the following limerick lemmata were thrown up:

Said Hardy, "How dull, I opine,
To have waited so long in that line
And end up in a cab
With a number so drab
As a thousand and seven-two-nine."

Said Ramanujan, sucking at jubes,
"You may as well travel by tubes
As omit to account
For the smallest amount
That is doubly the sum of two cubes."

It will be clear that more work is required in this area if anything of value is to be salvaged.

Other initiatives from the SRU were more successful. In particular, the Unit was able to make excellent use of results from the highly successful Scansion Project, which are still too sensitive to be released. (It is hoped, however, that they will be published before the end of the present academic year.) The question of Fermat's Last Theorem was tackled with some success:

"For positive a, b, and c,"
Remarked Fermat, "raise each to the d.
Then no two of the herd
Will add up to the third,
If d is no smaller than three.

I have really a marvellous proof
And it's not an erroneous goof.
Yet this margin's too small
To contain it at all,
But I promise it isn't a spoof!"

The announcement in 1993, by Prof. Andrew Wiles, that Fermat's Last Theorem had at last been proved, led to a scurry of activity at CUMLL, leading to some brief remarks in the Laboratory journal (see, for example, [**6**]), but this was perhaps excessively hasty, in the light of subsequent analysis apparently showing that Wiles' proof was flawed.*

3. Other results

The Laboratory's Four Colour Project is now being wound up,† its research being essentially complete. Some related results were presented above. Other lines of attack, relying on repetitive case-checking by computer, proved less fruitful than might perhaps have been hoped. Yet another approach was to consider the notion of trying to dapple, or possibly darken, a map. Producing the requisite number of rhymes sometimes led the Project into deep waters:

If you want without clashes to dapple
A chart on a plane, then your map'll
Require only four
Coloured pencils, no more,
As was proven by Haken and Appel.

Other useful work in various fields has been done by Dilip Sequeira; some of his results appeared in [**7**].

References

[1] T. D. Bending, M. A. Wainwright. "Mathematical Limericks". *Eureka* **49** (1989) 56–58.
[2] P. D. B. Dweller, M. Smith, M. A. Wainwright, W. Wanderer. "A proof of the Extended Wainwright–Bending Hypothesis". *J. Math. Lim. Research* **94** (1990) 121–216.
[3] A. Nomet. "Clerihews" and "More Clerihews". *Eureka* **51** (1991) 43 and 54.
[4] A. Nomet. "A note on the DSWW Proof of the Extended Wainwright–Bending Hypothesis". *Bull. Amer. Math. Lim. Soc.* **113** (1991) 145–148.
[5] A. Nomet. "A proof of the Nomet–Wainwright–Bending Hypothesis". *Proceedings of the Second Dublin Mathematical Limerick Congress* (Springer-Verlag 1992) 103–149.
[6] P. G. Ring. "Some conjectures concerning Fermat's Last Limerick". *J. CUMLL* **8** (1993) 111–114.
[7] M. J. Fryers, D. Sequiera, D. J. Sequiera. "Mathematical Verse". *Eureka* **52** (1993) 3.

* This comment may have been slightly hasty, in the light of further analysis apparently showing that Wiles' proof was not as flawed as had previously been thought. —Ed.

† Again, this was written too soon. Apparently a new proof of the FCT has just been announced, much shorter and in principle checkable by a human. CUMLL has set to work once more on the Theorem.

The Gyroscope

Kosuke Odagiri

The complete solution of gyroscopic motion involves elliptic integrals and elliptic functions (which are inverses of elliptic integrals). In this article we derive the standard elliptic integral expressions describing the motion of the symmetric gyroscope.

Elliptic integrals are integrals involving square roots of third or fourth degree polynomials. These can be reduced to combinations of three integrals known as the Legendre–Jacobi standard form:

$$F(k,\phi) = \int_0^\phi \frac{d\psi}{\sqrt{1-k^2\sin^2\psi}} = \int_0^{\sin\phi} \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}},$$

$$E(k,\phi) = \int_0^\phi \sqrt{1-k^2\sin^2\psi}\cdot d\psi = \int_0^{\sin\phi} \sqrt{\frac{1-k^2z^2}{1-z^2}}dz,$$

and

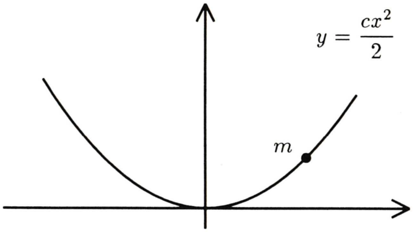
$$\Pi(\phi,n,k) = \int_0^\phi \frac{d\psi}{(1+n\sin^2\psi)\sqrt{1-k^2sin^2\psi}}$$
$$= \int_0^{\sin\phi} \frac{dz}{(1+nz^2)\sqrt{(1-z^2)(1-k^2z^2)}}.$$

Elliptic integrals are so called because the arc length of an ellipse involves an integral of this form. It is easy to show that for an ellipse of major axis $2a$ and minor axis $2b$ the arc length is given by

$$L(a,b) = 4aE\left(\sqrt{1-\frac{b^2}{a^2}},\frac{\pi}{2}\right) = 4aE\left(e,\frac{\pi}{2}\right), \tag{1}$$

where e is the eccentricity of the ellipse.

A less obvious case is that of the motion of a particle on a parabolic path under gravity.



For a particle on any curve,

$$\frac{m}{2}\left(\frac{ds}{dt}\right)^2 + mgy = \text{total energy} = \text{constant}.$$

If

$$\frac{ds}{dt} = 0 \quad \text{at} \quad y = y_0,$$

$$\frac{m}{2}\left(\frac{ds}{dt}\right)^2 = mg(y_0 - y),$$

and solving this by separation of variables gives

$$t = \int \frac{ds}{\sqrt{2g(y_0 - y)}} \tag{2}$$

as the general solution. In the case of a parabola, substituting

$$y = \frac{cx^2}{2} \quad \text{and} \quad y_0 = \frac{cx_0^2}{2}$$

gives

$$t = \int \sqrt{\frac{1 + c^2x^2}{gc(x_0^2 - x^2)}}dx.$$

Hence the period of oscillation is

$$T = 4\int_0^{x_0} \sqrt{\frac{1 + c^2x^2}{gc(x_0^2 - x^2)}}dx.$$

Substituting $x/x_0 = \cos\phi$ reduces the above expression to

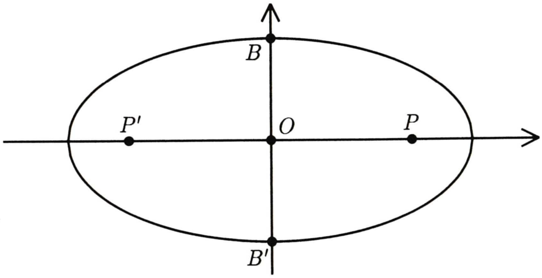
$$\begin{aligned} T &= 4\sqrt{\frac{1 + c^2x_0^2}{gcx_0^2}}\, E\left(\frac{cx_0}{\sqrt{1 + c^2x_0^2}}, \frac{\pi}{2}\right) \\ &= L\left(\sqrt{\frac{c}{g}\left(\frac{1 + c^2x_0^2}{c^2x_0^2}\right)}, \sqrt{\frac{c}{g}\left(\frac{1}{c^2x_0^2}\right)}\right). \end{aligned}$$

The period of oscillation of a particle confined in a parabolic well under gravity has a geometric analogue, as follows:

(1) Plot foci P,P' such that $OP = OP'$ represents time $\sqrt{c/g}$.

(2) Plot B,B' such that $OB = OB' = (1/cx_0)OP = (x_0/2y_0)OP$.

(3) Using a string, or otherwise, draw an ellipse with the foci at P and P' going through B and B'.



The circumference represents the period of oscillation.

It is useful to define complete elliptic integrals as follows:

$$K(k) = F\left(k, \frac{\pi}{2}\right) \quad \text{and} \quad E(k) = E\left(k, \frac{\pi}{2}\right).$$

Hence (1) becomes

$$L(a,b) = 4aE\left(\sqrt{1 - \frac{b^2}{a^2}}\right).$$

These are called the complete elliptic integrals of the first and second kind, respectively.

We now consider the simple pendulum. For this system the length element is $ds = l d\theta$ and the height $y = -l\cos\theta$. Equation (2) now becomes

$$\begin{aligned} t &= \int \frac{l d\theta}{\sqrt{2gl(\cos\theta - \cos\theta_0)}} \\ &= \sqrt{l/g} \int \frac{d(\theta/2)}{\sqrt{\sin^2(\theta_0/2) - \sin^2(\theta/2)}}. \end{aligned}$$

Substituting $k = \sin(\theta_0/2)$ and $kz = \sin(\theta/2)$ reduces the above expression to

$$t = \sqrt{l/g} \int \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}}.$$

Therefore the period of oscillation is

$$\begin{aligned} T &= 4\sqrt{l/g} \int_0^1 \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}} \\ &= 4\sqrt{(l/g)}\, K(k). \end{aligned}$$

Although the above algebraic manipulations are simple the method used in reducing the solution to the standard form is worth attention. Consider an integral of the form

$$\int \frac{d\theta}{\sqrt{\cos\theta - \cos\theta_0}}, \quad \text{with } 0 \leqslant \theta \leqslant \theta_0.$$

Substituting $x = \cos\theta$ transforms this to

$$-\int \frac{dx}{\sqrt{(x - x_0)(1-x^2)}}, \qquad x_0 \leqslant x \leqslant 1.$$

Hence we have established a method for converting an integral of the form

$$\int_0^{x_0} \frac{dx}{\sqrt{(x-x_0)(1-x^2)}} \tag{3}$$

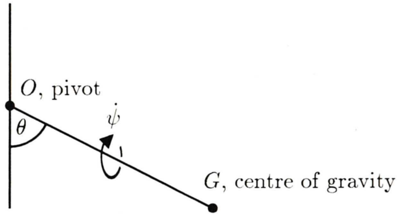
to a complete elliptic integral of the first kind.

The amplitude function $\operatorname{am}(u,k)$ is defined as the inverse of the elliptic integral of the first kind. If

$$u = \int_0^{\phi} \frac{d\psi}{\sqrt{1 - k^2\sin^2\psi}}, \quad \text{then } \phi = \operatorname{am}(u,k).$$

Functions sn and cn are defined as the sine and cosine of the amplitude function.

We are now ready to tackle the gyroscope.



We use standard spherical polars: θ is the angle between OG and the downward vertical and ϕ is the angle between OG and a fixed horizontal as seen from above. The length of the gyroscope is h. The angular rotation of the gyroscope is denoted by $\dot{\psi}$ (positive is anticlockwise as seen from G). It can be seen that although $\dot{\theta}$ is perpendicular to $\dot{\psi}$ and $\dot{\phi}$, $\dot{\psi}$ and $\dot{\phi}$ are not necessarily orthogonal. The component of $\dot{\phi}$ along the $\dot{\psi}$ direction is $-\dot{\phi}\cos\theta$ and that perpendicular to it is $\dot{\phi}\sin\theta$.

A possible solution is that θ will remain constant and that the gravitational couple will be just the right amount to keep the symmetry axis of the gyroscope moving on a cone. This is called *steady precession*, and the rate of precession is the angular speed at which the axis describes a complete cone. In general, however, θ will oscillate. This oscillation is termed *nutation*. If nutation occurs, θ, $\dot{\phi}$, and $\dot{\psi}$ will all oscillate at the nutational frequency. We will derive this frequency by considering the oscillation of θ.

If we define $I_{\|}$ to be the moment of inertia along OG, and $I_{\perp}$ to be the moment of inertia perpendicular to OG about O, then the total kinetic energy is given by

$$T = \tfrac{1}{2}I_{\|}(\dot{\psi} - \dot{\phi}\cos\theta)^2 + \tfrac{1}{2}I_{\perp}\left[\dot{\theta}^2 + (\dot{\phi}\sin\theta)^2\right].$$

The gravitational potential energy is $V = -mgh\cos\theta$ where $h = \overline{OG}$, and m=total mass, so

$$\mathcal{L} = T - V = \tfrac{1}{2}I_{\|}(\dot{\psi} - \dot{\phi}\cos\theta)^2 + \tfrac{1}{2}I_{\perp}\left[\dot{\theta}^2 + (\dot{\phi}\sin^2\theta)^2\right] + mgh\cos\theta.$$

The Euler–Lagrange equations are

$$\frac{\partial\mathcal{L}}{\partial x} - \frac{d}{dt}\left(\frac{\partial\mathcal{L}}{\partial\dot{x}}\right) = 0, \quad (x = \psi, \theta, \phi).$$

Therefore we have the following constants of motion:

1. Total energy,

$$\begin{aligned} E &= \frac{\partial\mathcal{L}}{\partial\dot{\theta}}\dot{\theta} + \frac{\partial\mathcal{L}}{\partial\dot{\phi}}\dot{\phi} + \frac{\partial\mathcal{L}}{\partial\dot{\psi}}\dot{\psi} - \mathcal{L} \\ &= \tfrac{1}{2}I_{\|}(\dot{\psi} - \dot{\phi}\cos\theta)^2 + \tfrac{1}{2}I_{\perp}\left[\dot{\theta}^2 + (\dot{\phi}\sin\theta)^2\right] - mgh\cos\theta. \end{aligned}$$

2. Angular momentum along OG (since $\mathcal{L}$ is independent of ψ),

$$J_{\psi} = \frac{\partial\mathcal{L}}{\partial\dot{\psi}} = I_{\|}\dot{\psi} - I_{\|}\dot{\phi}\cos\phi.$$

3. Angular momentum in the vertical direction (since $\mathcal{L}$ is independent of ϕ),

$$\begin{aligned} J_z = \frac{\partial \mathcal{L}}{\partial \dot{\phi}} &= -I_{\|}\dot{\psi}\cos\theta + +I_{\|}\dot{\phi}\cos^2\theta + I_{\perp}\dot{\phi}\sin^2\theta \\ &= -J_{\psi}\cos\theta + I_{\perp}\dot{\phi}\sin^2\theta. \end{aligned}$$

From these we can derive the familiar equation for gyroscopic motion:

$$E = \tfrac{1}{2}I_{\perp}\dot{\theta}^2 + \frac{J_{\psi}^2}{2I_{\|}} + \frac{(J_z + J_{\psi}\cos\theta)^2}{2I_{\perp}\sin^2\theta} - mgh\cos\theta.$$

We can solve this equation by separation of variables: let $\theta_0 \leqslant \theta \leqslant \theta_1$. Since $\dot{\theta} = 0$ at $\theta = \theta_0$ or $\theta = \theta_1$, we get

$$\frac{(J_z + J_{\psi}\cos\theta_1)^2}{2I_{\perp}\sin^2\theta_1} - mgh\cos\theta_1 = \tfrac{1}{2}I_{\perp}\dot{\theta}^2 + \frac{(J_z + J_{\psi}\cos\theta)^2}{2I_{\perp}\sin^2\theta} - mgh\cos\theta.$$

Let $y = \cos\theta$, $y_0 = \cos\theta_0$ and $y_1 = \cos\theta_1$ (y is minus the height, or the *depth*, of the gyroscope; y_0 and y_1 are the maximum and minimum depths respectively, normalised by h). Then $dy = -\sin\theta d\theta$, so

$$\dot{\theta}^2 = \frac{\dot{y}^2}{\sin^2\theta} = \frac{\dot{y}^2}{1-y^2}.$$

Substituting this into the previous equation we get

$$\tfrac{1}{2}I\frac{\dot{y}^2}{1-y^2} = mgh(y-y_1) + \frac{(J_z + J_{\psi}y_1)^2}{2I(1-y_1^2)} - \frac{(J_z + J_{\psi}y)^2}{2I(1-y^2)},$$

or, rearranging:

$$\begin{aligned} I^2\dot{y}^2(1-y_1^2) &= 2mghI(y-y_1)(1-y_1^2)(1-y^2) \\ &\quad + (J_z + J_{\psi}y_1)^2(1-y^2) - (J_z + J_{\psi}y)^2(1-y_1^2) \end{aligned}$$

where the subscript on $I_{\perp}$ is omitted for simplicity. To get this in the form (3) this expression must be factorised. Clearly $y - y_1$ is a factor.

$$\begin{aligned} I^2\dot{y}^2(1-y_1^2) &= 2mghI(y-y_1)(1-y_1^2)(1-y^2) + J_z^2(y_1^2 - y^2) \\ &\quad + 2J_{\psi}J_z\left[y_1(1-y^2) - y(1-y_1^2)\right] + J_{\psi}^2(y_1^2 - y^2) \\ &= 2mgh(y-y_1)(1-y_1^2)(1-y^2) + (J_z^2 + J_{\psi}^2)(y_1^2 - y^2) \\ &\quad + 2J_{\psi}J_z(y_1 - y)(1 - y_1 y) \\ &= (y-y_1)\left[2mghI(1-y^2) - (J_z^2 + J_{\psi}^2)(y+y_1) - 2J_{\psi}J_z(1-y_1 y)\right]. \end{aligned}$$

Now, to make the remaining factorisation easier, define

$$\begin{aligned} P^2 &= 2mghI(1-y_1^2) \geqslant 0, \\ 2PQ &= J_{\psi}^2 + J_z^2 - 2J_{\psi}J_z y_1 \\ &= (J_{\psi} - J_z y_1)^2 + J_z^2(1-y_1^2) \geqslant 0, \\ \text{and}\quad R^2 &= P^2 + Q^2 - 2J_{\psi}J_z - (J_{\psi}^2 - J_z^2)y_1 \geqslant 0, \end{aligned}$$

with R and P positive. This gives

$$\begin{aligned} I^2\dot{y}^2(1-y_1^2) &= (y-y_1)\left[P^2(1-y^2) - 2PQy - 2J_\psi J_z - (J_\psi^2 + J_z^2)y_1\right] \\ &= (y-y_1)\left[R^2 - (Py+Q)^2\right] \\ &= (y-y_1)(R+Py+Q)(R-Py-Q). \end{aligned} \tag{4}$$

We now want to find y_0, the maximum depth. Clearly $\dot{y}=0$ at y_0.

$$(y_0-y_1)(R+Py_0+Q)(R-Py_0-Q)=0,$$

so $R+Py_0+Q=0$ or $R-Py_0-Q=0$, i.e.,

$$y_0 = -\frac{Q+R}{P}, \quad \text{or} \quad y_0 = \frac{R-Q}{P}.$$

It can be shown that

$$-\frac{Q+R}{P} \leqslant y_1,$$

but y_0 and y_1 were chosen to be the maximum and minimum depths respectively, and so if y_0 were this, we would have $y_1 > y_0$, which is clearly impossible. Therefore

$$y_0 = \frac{R-Q}{P}.$$

Returning to (4), we can rewrite it as

$$I^2\dot{y}^2(1-y_1^2) = R^2(y-y_1)\left(1+\frac{Py+Q}{R}\right)\left(1-\frac{Py+Q}{R}\right).$$

Define

$$Y = \frac{Py+Q}{R} \quad \text{and} \quad Y_1 = \frac{Py_1+Q}{R}.$$

Since $y_1 \leqslant y \leqslant y_0$, $y_0 = (R-Q)/P$ and $P, R \geqslant 0$,

$$Y_1 = \frac{Py_1+Q}{R} \leqslant \frac{Py_0+Q}{R} = 1.$$

Now

$$\dot{Y} = \frac{P}{R}\dot{y}.$$

Therefore

$$\begin{aligned} I^2\frac{R^2}{P^2}\dot{Y}^2(1-y_1^2) &= R^2\frac{R}{P}(Y-Y_1)(1+Y)(1-Y), \\ I^2\dot{Y}^2(1-y_1^2) &= RP(Y-Y_1)(1-Y^2), \end{aligned}$$

which is of the form (3). Using the same method we used for the simple pendulum, let $Y = \cos\Phi$ and $Y = \cos\Phi_1$. Since $Y_1 \leqslant Y \leqslant 1$ we have $0 \leqslant \Phi \leqslant \Phi_1$. It follows that $\dot{Y} = -\sin\Phi\dot{\Phi}$, so $\dot{Y}^2 = \sin^2\Phi\dot{\Phi}^2 = (1-Y^2)\dot{\Phi}^2$.

$$\begin{aligned} I^2(1-Y^2)\dot{\Phi}^2(1-y_1^2) &= RP(\cos\Phi - \cos\Phi_1)(1-Y^2); \\ I^2\dot{\Phi}^2(1-y_1^2) &= RP(\cos\Phi - \cos\Phi_1) \\ &= 2RP\left(\sin^2\frac{\Phi_1}{2} - \sin^2\frac{\Phi}{2}\right). \end{aligned}$$

Let $k = \sin(\Phi_1/2)$, $kz = \sin(\Phi/2)$, with $0 \leqslant \Phi \leqslant \Phi_1$, so $0 \leqslant z \leqslant 1$. This implies that

$$\dot{\Phi}^2 = \frac{4k^2\dot{z}^2}{1-k^2z^2}.$$

Therefore

$$I^2\frac{4k^2\dot{z}^2}{1-k^2z^2}(1-y_1^2) = 2RP(k^2-k^2z^2).$$

Solving for $\dot{z}^2$ we have

$$\dot{z}^2 = \frac{RP}{2I^2(1-y_1^2)}(1-z^2)(1-k^2z^2),$$

which gives us the following solution to our differential equation:

$$\int dt = \int \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}}\sqrt{\frac{2I^2(1-y_1^2)}{RP}}. \tag{5}$$

Now let

$$\nu = \sqrt{\frac{RP}{2I^2(1-y_1^2)}} = \sqrt{\frac{mghR}{IP}};$$

then the solution can be expressed as

$$\begin{aligned} \sin^{-1}z &= \mathrm{am}[\nu(t-t_0)] \\ \text{or} \qquad z &= \mathrm{sn}[\nu(t-t_0)]. \end{aligned}$$

The period of the oscillations can be defined as two nutational cycles, so that in the limiting case of a simple pendulum the period is the period of oscillation of a simple pendulum. Therefore

$$\begin{aligned} T &= 4\int_0^1 \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}}\sqrt{\frac{2I^2(1-y_1^2)}{RP}} \\ &= 4\sqrt{\frac{2I^2(1-y_1^2)}{RP}}K(k) \\ &= 4\sqrt{\frac{IP}{mghR}}K(k). \end{aligned}$$

Hence the timing of nutational oscillation of a gyroscope is equal to that of a simple pendulum with length of string l and amplitude α such that

$$l = \frac{2I^2 g\sin^2\theta_1}{RP} = \frac{IP}{mhR}$$

and

$$\cos\alpha = \frac{P\cos\theta_1 + Q}{R},$$

$$\sin\frac{\alpha}{2} = \sqrt{\frac{P}{2R}(\cos\theta_0 - \cos\theta_1)},$$

where

$$P = \sqrt{2mghI}\sin\theta_1 ,$$
$$R = \sqrt{P^2 + Q^2 - 2J_\psi J_z - (J_\psi^2 + J_z^2)\cos\theta_1} ,$$
$$Q = \frac{1}{zP}\left((J_\psi - J_z)^2 + J_z^2\sin\theta_1\right) ,$$

and θ_0 and θ_1 are the minimum and maximum values of θ respectively.

The rate of precession can be calculated from the equation

$$J_z = -J_\psi\cos\theta + I_\perp\dot{\phi}\sin^2\theta .$$

This gives

$$\dot{\phi} = \frac{J_z + J_\phi\cos\theta}{I_\perp\sin^2\theta} = \frac{J_z + J_\psi y}{I_\perp(1-y^2)} .$$

We know that $y = y_0 - (2R/P)k^2z^2$, so we get

$$\dot{\phi} = \frac{J_z + J_\psi(y_0 - (2R/P)k^2z^2)}{I_\perp[1 - (y_0 - (2R/P)k^2z^2)^2]}$$

and z is given by $\mathrm{sn}[y(t-t_0)]$. ϕ can also be found as a function of z by multiplying the above expression by

$$dt = \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}}\sqrt{\frac{2I^2(1-y_1^2)}{RP}}$$

and integrating. This will be a combination of elliptic integrals of the third kind.

(5) reduces to the compound pendulum solution for $J_\psi = J_z = 0$, i.e., $P = R$. If $J_\psi = 0$, we have the general motion of a paricle on a string which remains taut. The extension to an asymmetric gyroscope would involve a more tedious derivation of the equation of motion.

How to be a Good Lecturer

Jonathan Partington

"Well hello and welcome to the first lecture of the course er look I said hello look I'd like to start now will you shut up SHUT UP PLEASE oh thank you I don't mind you talking if you do so quietly I didn't ask to do this course you know I wanted to do algebra I told them I didn't know any analysis ...

"... Now this course is all about complex numbers and I've got a list of recommended books here er well no in fact I seem to have left it behind never mind they're all out of print anyway now let me write up a definition where's the chalk gone ah here it is [SNAP] ah let me take another piece [THUD] not very big these platforms are they I keep falling off them ...

"... Now definition 1.1 is ah um of course I haven't said what this section's called yet oh it doesn't seem to have a name anyway it's all about convergence of power series you did something like it in real analysis didn't you don't you remember well he should have done it in his lectures I don't have time to go into it now ...

"... Now definition 1.1 [scribble scribble] can you read that at the back no oh well sit further forward then can you read it at the front ah come to think of it I can't read it either perhaps if I turn on this light ah no not that one another one oh well the cord was a bit frayed I suppose well look that symbol is a capital sigma yes what's the problem yes well green seems to be the only colour they have left in the box probably because nobody in his right mind uses it so they leave it for me ...

"... Well look perhaps if I explain it in words it's all in the textbooks anyway I can't help it if they're missing from the library people eat them or something well now I'll draw a diagram you don't have to copy this exactly because it's slightly wrong anyway this is diagram 2 good question I think I forgot to draw diagram 1 anyway as I say it doesn't help much phew let me take my jacket off a bit [rip] oh well I sewed that button on myself you can tell can't you ...

"... Now let me digress a minute about the history of the subject here it was discovered by Cauchy or do I mean Gauss one of those people and he sent a copy of his paper to someone else who well anyway it's very important and has a lot of applications such as er such as well anyway you will see applications in your other courses I expect of course they don't use the same notation but then they don't have the same ideas of rigour as we do and now let's write down the first result lemma 1.2 ...

"... Lemma 1.2 oh I haven't actually defined radius of convergence yet have I still let me write it up and we can decide what it means later well I still seem to have a few minutes left so I'd better start the proof let n be this and r be this and v be that and n be that no on second thoughts I'm already using n now so I'll call it nu pardon no it's a nu a greek letter you must have seen it before you know greek letters alpha etcetera no this one is nu all right call it v if you like but we're already using v still it won't cause confusion ...

"... Now multiply this out and obviously what we get is er clearly um oh that can't be right what have I done wrong here can you see the mistake maybe I lost a minus sign somewhere search me oh dear it's time to finish isn't it well give me just 5 more minutes and I'll finish this off and oh maybe I should do this bit again more carefully next time ah that should have been a nu maybe no it should be a v oh it's an t is it oh well look I'll finish this next time I'm sure I've got most of the details right it's really very elementary after all I haven't done anything nontrivial yet ..."

How to be a good member of a lecture audience

"Aaaachoooo! Cough. Splutter. Wheeze. Yes I've got a cold. There's a lot of it about. No I don't use a handkerchief. Sniff. Sniff. Cough. Oh thanks, now I've sneezed on your notes I might as well blow my nose on them. Zurrrrrkkk! Hoooossssh! Now what lecture is this? ...

"... Do you think he's got this bit wrong? Well I'm sure you can prove it quickly using matrices. Shall I ask him whether you can? No. Something wrong? No, nothing wrong. I was just wondering if you could prove it more quickly with matrices. Oh I see. Stick my head in a bucket of WHAT? Oh right. Yes ...

"... God this is so boringly obvious. I think I'll do the crossword instead. Mixed-up caterpillar in tribal religion, we hear? Hmm. Can you think of an anagram of caterpillar? Oh I'm SORRY. I didn't realise you were listening to the lecturer. Oh I thought he was proving a different theorem. Excuse me, how do you get x-squared there? You just explained that. Sorry, I didn't realise ...

"... Can I borrow a bit of paper? Have I really borrowed one every day this week? Ah thanks. I don't suppose you have a pen I could use? Yes I'll take care of it. Ooops, it's on the floor. [SCRUNCH.] Ah well at least we know where it is now ..."

And how to be a good exam invigilator

"O.K. you can start writing as soon as you get to your places. Look would you mind sitting down? What do you mean there isn't a desk for you? You must be in the wrong room. What's your name? Oh. Well there don't seem to be enough desks. Perhaps you could sit on the floor this time. Come on, let's get started ...

"... Ha ha ha ha ha! Oh sorry. I've just seen the joke in question 5. I don't know how they think of them. What a laugh exams are, eh? Anyway don't let me disturb you. Sorry about that ...

"... What do you want? Well why didn't you go beforehand? Honestly, the incontinents you get round here. Well why didn't you bring a pottie with you? Oh all right I'll find someone to escort you. Can't have you stinking the place out, can we? Though maybe you should have a doctor's certificate [rustle rustle]. No, it doesn't mention that. O.K. get a move on ...

"... [Creak, creak, creak, crash!] Bloody hell, they don't make chairs like they used to, do they. I bet Chippendale's chairs never gave way when you leant back on them. Oh well, now I've nowhere to sit down. [Tramp, tramp, tramp.] (God what a useless answer that chap's writing. Even I know that 2+2 is 4 not 5. Must be nerves, poor chap.) Oh sorry, am I putting you off? I'll go and breathe down somebody else's neck ...

"... Ah, this one looks calm—he's writing away nineteen to the dozen. A-a-a-a-SHOOOO!!! Oh sorry. Yes we can pick up all the sheets of paper. And I'll try and find you a clean question paper. What was that sheet that went through the window? Question 2? Oh well, maybe somebody will pick it up and hand it in to us. You wouldn't have got many marks on it anyway, it's quite tricky ...

"Right, all writing must cease now. In fact if you knew your stuff it would have ceased 20 minutes ago. Look I told you to stop. Well you'll have to hand it in anonymously then, won't you? I don't suppose it'll make much difference to your result ...

"... Honestly the students of today just can't cope the way we had to ..."

On Postulates of a Group

Jingcheng Tong

First a definition; A *semigroup* is a set G together with an associative binary operation (written multiplicatively) $G \times G \to G$.

Therefore a semigroup G with an identity is a group if every element a in G has an inverse. It is well known that this statement can be weakened to: a semigroup G with a left identity is a group if every element a in G has a left inverse. Can this weakened statement be weakened again? The following theorem gives an affirmative answer:

THEOREM 1. *A semigroup G is a group if and only if*

(i) *there is an element e in G such that, for each element a in G, there is a positive integer m_a with $e^{m_a}a = a$;*

(ii) *for each element a in G, there is an element a' in G and a positive integer n_a with $a'a = e^{n_a}$.*

PROOF. The necessity is trivial since if G is a group we can let $m_a = n_a = 1$ for any element a in G.

To prove sufficiency, let $a = e$ in (i). Then

$$e^{m_e+1} = e.$$

It is easily seen that, for any positive integer s,

$$e^{s(m_e+1)} = e^s.$$

Let a be an arbitrary element in G. Then there is a positive integer m_a such that $e^{m_a}a = a$. There are obviously many positive integers satisfying this equality. Denote the smallest such by $\overline{m}_a$. Then $e^{\overline{m}_a}a = a$. We shall now prove that $\overline{m}_a \leqslant m_e$ for any element a in G.

If $\overline{m}_a \geqslant m_e + 1$, then there are two positive integers q_a and r_a such that

$$\overline{m}_a = q_a(m_e + 1) + r_a, \qquad \text{where } q_a \geqslant 1, \text{ and } \quad 0 \leqslant r_a \leqslant m_e.$$

This implies

$$e^{\overline{m}_a} = e^{q_a + r_a}.$$

So $q_a + r_a$ satisfies the requirements for m_a and is less than $\overline{m}_a$ which is a contradiction. Therefore $\overline{m}_a \leqslant m_e$. Similarly we can prove that the least positive integer $\overline{n}_a$ satisfying $a'a = e^{\overline{n}_a}$ cannot exceed m_e.

If s is a positive integer and $e^{m_a}a = a$, then

$$e^{sm_a}a = e^{(s-1)m_a}(e^{m_a}a) = e^{(s-1)m_a}a = \ldots = e^{m_a}a = a.$$

If $k = m_e!$ then k is a constant and, for any element a in G, $\overline{m}_a \mid k$. Therefore $e^k a = a$ for all a in G.

From $a'a = e^{\overline{n}_a}$, we have

$$e^{k-\overline{n}_a}(a'a) = e^{k-\overline{n}_a}e^{\overline{n}_a}$$

which we rearrange to

$$(e^{k-\overline{n}_a}a')a = e^k.$$

Therefore for each element in a in G, $e^{k-\overline{n}_a}a'$ is a left inverse to the identity e^k. G has a left identity and left inverses. Therefore by the result stated at the start of the article G is a group. □

In Theorem 1 we weakened the requirements on a one sided identity. It is natural to ask whether we can weaken the two sided form. The following theorem gives this:

THEOREM 2. *A semigroup G is a group if and only if*

(i) *there is an element e in G such that, for any element a in G, there are two positive integers l_a, r_a such that $e^{l_a}ae^{r_a}$;*

(ii) *for each element a in G, there is an element a' in G and a positive integer n_a such that $a'a = e^{n_a}$.*

PROOF. The necessity is trivial since if G is a group we may let $l_a = r_a = 1$.

To prove the sufficiency, let $a = e$ in the first equality. Then we have

$$e^{l_e+1+r_e} = e.$$

Let $k = l_e + r_e + 1$. By similar methods to before, it is easily seen that for any positive integer s

$$e^{sk} = e^s.$$

Let $\overline{l}_a, \overline{r}_a$ be a pair of positive integers with minimal sum such that $e^{\overline{l}_a}ae^{\overline{r}_a} = a$. Then, as in Theorem 1, dividing $\overline{l}_a$ and $\overline{r}_a$ by $k+1$ implies $\overline{l}_a \leqslant k-1$ and $\overline{r}_a \leqslant k-1$. Similarly if n_a is the smallest positive integer such that $a'a = e^{\overline{n}_a}$, then $\overline{n}_a \leqslant k-1$. Now k is a constant and

$$a = e^{\overline{l}_a}ae^{\overline{r}_a},$$

so

$$\begin{aligned} e^{k-1}a &= e^{k-1}ee^{\overline{l}_a-1}ae^{\overline{r}_a} \\ &= ee^{\overline{l}_a-1}ae^{\overline{r}_a} \\ &= a. \end{aligned}$$

Therefore e^{k-1} is a left identity in G.

Now we show that for each element a in G has a left inverse. From $a'a = e^{\overline{n}_a}$ we have

$$e^{k-\overline{n}_a-1}(a'a) = e^{k-1}.$$

Hence $e^{k-\overline{n}_a-1}a'$ is a left inverse of a. So G has a left identity and left inverses and as before G is a group. □

Solving Polynomials

Jamie Gabbay

A complex polynomial p is a map $\mathbb{C} \to \mathbb{C}$ defined by

$$p(x) = \sum_{i=0}^{n} t_i x^i = t_n \prod_{i=1}^{n} (x - p_i) \qquad t_i, p_i, x \in \mathbb{C}.$$

In this article, p will be monic throughout (normalising if necessary) and we shall also impose the condition that the roots are distinct. Given this, a degree n polynomial determines and is uniquely determined by its roots (a set of n complex numbers). Using this isomorphism, we can treat p and the set $\{p_i\}$ more or less interchangeably.

The reason for imposing distinct routes is that, for example, a cubic with a double root behaves for our purposes as a quadratic. This special case needs careful treatment but in almost all cases the formulae we obtain work anyway through a limiting argument, so in this article multiple roots are simply going to be forbidden. We shall write $\mathcal{P}$ for the set of finite sets of complex numbers (or the set of monic polynomials in $\mathbb{C}$ with distinct roots) and $\mathcal{P}_n \subset \mathcal{P}$ for the subset of sets with n elements. We shall also write

$$p(x) = \{p_1, p_2, \ldots, p_n\}(x) = \prod_i (x - p_i). \tag{1}$$

For example, $\{1,5\} \subset \mathcal{P}_2 \subset \mathcal{P}$, and $\{1,5\}(x) = x^2 - 6x + 5$

We next define the action of a Möbius map on elements of $\mathcal{P}$. A Möbius map is of the form

$$T(x) = \frac{ax+b}{cx+d}, \qquad a, b, c, d \in \mathbb{C}.$$

The action of T on a set is as usual defined to be that of applying T to each of the elements, thus

$$T\{v, w, x, y, z\} = \{Tv, Tw, Tx, Ty, Tz\}.$$

As is well known,a unique T exists which maps three given distinct complex numbers to any three other given complex numbers. If we wish to find a T such that $Tp_i = q_i$, where $\{p_i\}$ and $\{q_i\}$ are in $\mathcal{P}_3$, we must solve the three linear equations

$$ap_1 + b = q_1(cp_1 + d), \qquad ap_2 + b = q_2(cp_2 + d), \qquad ap_3 + b = q_3(cp_3 + d),$$

and this can be done up to multiplication by a scalar (a bit of algebra shows that linear independence is equivalent to the members of the first triple being distinct).

Generally T is an isomorphism of $\mathbb{C}$ and well behaved except for two loose ends. If $x = -d/c$, then the denominator is zero, so we shall call $T(x)$ "∞" and define $T(\infty) = a/c$, a standard device when dealing with Möbius maps. Note that we do not bother enforcing the standard normalisation ($ad - bc = 1$); we will be multiplying through by a (different) normalisation constant anyway to keep our polynomials monic. If $ad - bc = 0$, however, the map is not an isomorphism; it is *singular* and maps everything onto $b/d = a/c$. This is obviously unhelpful, for our aim is to transform 'nasty' polynomials to 'nice' ones in order to find their roots, and we will exclude this case. Note that singularity almost always corresponds to the roots not being distinct, and we shall assume all our maps are non-singular without proof.

Consider $p \in \mathcal{P}_n$ and T Möbius. From (1) we have

$$(Tp)(x) = \{Tp_1, \ldots, Tp_n\}(x) = \prod(x - Tp_i).$$

This polynomial is zero when $x = Tp_i$, or, put another way, when $T^{-1}x = p_i$. We can write T^{-1} as

$$T^{-1}x = \frac{dx - b}{a - cx}.$$

If we now apply p to this, we get

$$p(T^{-1}x) = \prod(T^{-1}x - p_i) \tag{2}$$

This as it stands is not a polynomial in x, but we can make it one by multiplying it by $(a-cx)^n$, the denominator of (2), and then make it monic by dividing by $K = \prod(b+cp_i)$. Neither operation changes the roots, which is the set of x such that $T^{-1}x = p_i$, so it must be the same as $(Tp)(x)$. We hence have an explicit form for $(Tp)(x)$:

$$(Tp)(x) = pT^{-1}(x)(a - cx)^n/K.$$

In terms of the roots this can be simplified further to $(Tp)x = pT^{-1}(x)$ since the remaining terms on the right hand side are just the normalisation factors to turn it into a monic polynomial and do not affect the roots. It should be noted that $(Tp)x$ is not the same as $T(px)$ where in the latter case we are applying the original transformation T to the number $p(x)$.

We can now solve the general quadratic. Take $p \in \mathcal{P}_2$, say

$$p(x) = c_0 + c_1x + x^2.$$

As it stands, this is hard to solve (using the standard formula is off-limits here), so we seek a T such that

$$Tp = \{\alpha, -\alpha\}, \quad \text{i.e.,} \quad p(T^{-1}x)(a - cx)^n/K = x^2 - \alpha^2.$$

We have a fair amount of freedom here. We are specifying only two points to be moved, and not specifying α gives us a further degree of freedom. So we may set $a = d = 1$ and $c = 0$, which leaves us with $T(x) = x + b$. Then

$$\begin{aligned} x^2 - \alpha^2 = (Tp)(x) = p(T^{-1}x)(a - cx)^2/K &= t_0 + t_1(x - b) + (x - b)^2 \\ &= (t_0 - t_1b + b^2) + x(t_1 - 2b) + x^2. \end{aligned}$$

Equating coefficients gives us $b = t_1/2$ and $\alpha^2 = t_1^2/4 - t_0$. This gives us T, and so we know $p = T^{-1}\{\alpha, -\alpha\}$ or, in other words,

$$p = \left\{\frac{-t_1}{2} + \sqrt{\frac{t_1^2}{4} - t_0}, \frac{-t_1}{2} - \sqrt{\frac{t_1^2}{4} - t_0}\right\}$$

This should be recognisable as the usual formula for solving a monic quadratic. What is really going on here? Algebraically we have found a map from an arbitary quadratic into a more restricted family. Symbolically, we have completed the square. Geometrically, if p is real, we have shifted the axis of the parabola so that it passes through the origin, and more generally we have mapped the average of the two roots to the origin.

One might ask why we do not try to map the roots directly to $\{1, -1\}$. The trouble with doing this is that we would then have one degree of freedom fewer and hence could only fix two of a, b, c, and d. Whichever pair we choose to fix, it turns out we have to solve a quadratic to find what at least one of the other two are, which defeats the point of the exercise. There are two intuitive reasons why this should be: looking at the formulae we obtain shows that at least one of the terms in the Möbius map contains a square root (of course we don't know, a priori, that this doesn't come from a quadratic of the form $x^2 \pm k$ but this does not happen). Alternatively, there are two possible T, one which takes p_1 to 1 and p_2 to -1, and the other which takes p_1 to -1 and p_2 to 1, and both solve $(pT^{-1}) = \{1, -1\}$: this again forces a quadratic (again we don't necessarily know it is a nasty one). By our original choice of mapping, we are effectively only mapping one point to one point, so this problem doesn't arise.

We'll now consider the cubic, and map the three roots to the set $\{\lambda, \omega\lambda, \omega^2\lambda\}$, where $\omega^3 = 1$. At first sight, there are six ways of doing this, which would seem to imply a sextic. Remember though that λ is arbitary up to multiplication by ω, so there are in fact only two ways: if p_1 maps to λ, p_2 can map to either $\omega\lambda$ or $\omega^2\lambda$. The choice of λ leaves only one degree of freedom; we shall start with two free variables,(b and d), and normalise later. We set $a = 1$ and $c = -1$, so $T^{-1}(x) = (dx - b)/(x + 1)$. We are left to solve the system $Tp = \{\lambda, \omega\lambda, \omega^2\lambda\}$, which by our above identity becomes

$$p(T^{-1}x)(x+1)^3/K = x^3 - \lambda^3,$$

K being our normalisation factor as before. We shall start by assuming $t_2 = 0$. This can be achieved by the mapping $x \mapsto x - t_2/3$, and simplifies the algebra somewhat. Multiplying the left hand side out gives us

$$(t_0 + dt_1 + d^3)x^3 + (3t_0 + 2dt_1 - bt_1 - 3d^2b)x^2 + (3t_0 + dt_1 - 2bt_1 + 3db^2)x + (t_0 - bt_1 - b^3).$$

If we equate the x^2 and x coefficients and then rearrange we get $b = t_1/(3d)$ and $3d^2t_1 + 9dt_0 - t_1^2 = 0$, the quadratic we were expecting and which we can solve for d. (It can be noted that the two roots of this equation multiply to $t_1/3$ so the other one is in fact b.) λ is the cube root of the coefficient of 1 divided by the x^3 coefficient. From this it is easy to derive the standard solution to the cubic.

At this point it is worth pausing to compare what we have done with the more standard Galois theory method of solving polynomials, if somewhat sketchily. For more details of the Galois theory see any standard book on the subject.

Consider a field of the form $\mathbb{Q}(i, a_1, \ldots, a_n)$, which is the algebraic closure of $\mathbb{Q}$ and the set $\{i, a_i, \ldots, a_n\}$. In particular, consider a field of this type where $a_i = p_i$, the roots of some polynomial. This is known as the *splitting field* of the polynomial (over the field $\mathbb{Q}(i)$, to be precise). The *Galois group* of this field is for our purposes the group permuting the roots of the polynomial, leaving the *ground field* $\mathbb{Q}(i)$ untouched. For the polynomials with distinct roots under consideration, this is just a subgroup of S_n, the symmetric group on n elements. The general idea used in solving polynomials via Galois theory methods is to find quantities which are both invariant under the Galois group and which we can express in a simple form in terms of the roots. For example, for the quadratic, the Galois group consists of the identity and an element swapping p_1 and p_2. Hence we know that $p_1 + p_2$ and $(p_1 - p_2)^2$ are invariants, and from elementary considerations $p_1 + p_2 = t_1$ and $(p_1 - p_2)^2 = t_1^2 - 4t_0$, yielding the standard solution.

The cubic is somewhat more complicated. A standard theorem in Galois theory states that a polynomial is soluble in radicals if and only if it has a *soluble* Galois group. Soluble means that it has a finite series of subgroups $1 = G_0 \lhd G_1 \lhd \ldots \lhd G_n = G$, with G_{i+1}/G_i abelian. An appropriate series for S_3 is $1 \lhd A_3 \lhd S_3$. Armed with this information, it turns out to be a good idea to look for invariants of A_3 first. Define $y = p_1 + \omega p_2 + \omega p_3$; then since $A_3 = C_3$, the group just permutes the roots cyclically and an element of it multiplies y by a power of ω. Hence y^3 is an invariant. Similarly, if $z = 1 + \omega^2 p_3 + \omega p_2$, then z^3 is also an invariant. Moving up to S_3 we find that odd permutations swap y^3 and z^3, and so $y^3 z^3$ and $y^3 + z^3$ are invariants under S_3. A bit more algebra shows $y^3 + z^3 = -27p_0$ and $y^3 z^3 = -27p_1^3$ (assuming $p_2 = 0$ again), and from these a solution to the cubic can be calculated.

Having looked at this, what we did with the Möbius maps seems rather familiar. λ was effectively our invariant under the group A_3 and our Möbius map was reducing the symmetry among the roots from S_3 to A_3.

Solving the general quartic requires a little more theory. We define the *cross-ratio* C on $p \in \mathcal{P}_4$, mapping it to $\mathbb{C}$.

$$Cp = \frac{(p_1 - p_2)(p_3 - p_4)}{(p_1 - p_3)(p_2 - p_4)}.$$

C has the nice property that for $p, q \in \mathcal{P}_4$, $Cp = Cq$ if and only if $q = Tp$ for some T. C is useful because it tells us which quartics can be mapped onto each other by a Möbius transformation and which cannot. (Note that we can only specify three points for a Möbius transformation, and hence have no control over whether the fourth goes, so unlike the previous cases, all quartics are not equivalent to each other.) Suppose we were, naïvely, to try solving for T in

$$Tp = \{\lambda, i\lambda, i^2\lambda, i^3\lambda\} = q,$$

p being some quartic: we would get nowhere. This is because $Cq = \frac{1}{2}$. So we know that the general quartic cannot be converted into some q. Alternatively, there is no T such that $(Tp)(x) = x^4 - \lambda^4$.

Though the algebra in some cases can be a nightmare we can tell a lot of things about what we have to do in general when solving $Tp = q$ for T. There are two aspects we have to consider. The first is finding T, which is more subtle than previously. The second is solving the q once we get it, but through careful choice of method we can cut down the work needed in this step.

So let us solve $Tp = q$. We start with $q = \{1, 2, 3, \lambda\}$. It is reasonably obvious from the definition of C that there is a unique λ for each value of Cq. The value of Cp may vary, depending on the ordering of the roots. Any value of Cp will produce a corresponding λ and hence q such that $Cp = Cq$, and so a T such that $Tp = q$. Let

$$\alpha = (p_1 - p_2)(p_3 - p_4); \qquad \beta = (p_1 - p_3)(p_2 - p_4); \qquad \gamma = (p_1 - p_4)(p_2 - p_3).$$

If we permute p_1 to p_4, a short consideration of cases tells us that Cp takes one of the values $\{\alpha/\beta, \beta/\alpha, -\alpha/\gamma, -\gamma/\alpha, \gamma/\beta, \beta/\gamma\}$. There would appear to be 24 possible

T_i: mapping the set $\{p_i\}$ to $\{1, 2, 3, \lambda\}$ in every combination. However, again we do not particularly care which order the roots come out in: we will be given a polynomial $(x-1)(x-2)(x-3)(x-\lambda)$ which is trivial to solve, and merely need to map the roots back under T to find the roots of the original polynomial. An alternative way of looking at it is as follows: although there are 24 possible T_i, we can find Möbius transformations which permute $\{1, 2, 3, \lambda\}$ for a fixed λ by considering appropriate expressions of the form $T_i T_j^{-1}$. It turns out that all of these are their own inverse, and this means that the symmetry group for a given λ is $S_2 \times S_2 = V_4$. This is also the first term in a series for S_4: $1 \lhd V_4 \lhd A_4 \lhd S_4$. What we have done is to break down S_4 again: we have written it as $V_4 \times S_3$. By factorising in this way, we avoid any number larger than three and can solve the quartic.

The problem here is actually finding T. It can be done by a similar method to that we used in the cubic, but the algebra is rather more complicated and it takes some care to get a cubic and quadratic out.

A brief aside. Surely it might be possible to arrange for some other function, J say, with properties similar to C but insensitive to rearrangment of the roots. This would reduce the number of cases further by a factor of 6. When one looks at the mechanics of rearranging roots and what happens when JTp is simplified it becomes clear that any prospective J would have to be a rational polynomial in α^2, β^2, and γ^2 or something closely equivalent, with nominator and denominator homogenous and of the same degree. $\alpha^2 + \beta^2 + \gamma^2$ would be an ideal start, but has no partner. The next simplest candidate is

$$J = \frac{\alpha^4 + \beta^4 + \gamma^4}{\alpha^2\beta^2 + \beta^2\gamma^2 + \gamma^2\alpha^2}$$

but, considering this as a polynomial in p_1, we quickly discover that $J \equiv 2$. The next class of possibilities are all of degree 6, and therefore gain us nothing.

A somewhat simpler way of doing the same job turns out to be to transform the polynomial into one with roots $\{0, \alpha, \beta, \gamma\}$. Not only is this polynomial ultimately easier to solve but the algebra is rather simpler as we can express the roots more directly in terms of the roots of the original equation. Once again the Galois theory is doing something very similar, considering the same terms but with plus signs replacing our minus signs. These turn out to be the roots of the *resolvent cubic*, whose coefficients are relatively easy to find in terms of the coefficients of the original. (If the original has been simplified to $x_4 + t_2x^2 + t_1x + t_0$ then the resolvent is $x_3 - 2t_2x^2 + (t_2^2 - 4t_0)x + t_1^2$.)

The next step is to try to solve the general quintic, a worthy if ill-fated ambition, since the Galois theoretic methods tell us it is impossible in radicals (which are the only thing our theory can produce). We have analogues to C, say the pair C_1 and C_2, which operate on $p \in \mathcal{P}_5$.

$$C_1p = C(p\backslash\{p_1\}) \qquad C_2p = C(p\backslash\{p_2\})$$

These have all the properties one would expect, and they assure us that we can solve $Tp = q$, where $q = \{1, 2, 3, \lambda, \mu\}$ and obtain T. However, we now have $5! = 60$ solutions. Whichever way we factor 60 we cannot avoid a 5 which would imply a quintic. There are no obvious simplifications as occurred in the quadratic and cubic cases through having extra degrees of freedom, and so our method fails. The basic structure of the problem is, as the Galois theory reveals, the fact that S_5 (or more particularly A_5) represents an immovable block sitting in any larger system no matter what the precise nature of the transformation.

A Result in Metrical Space Theory

Michael Fryers

THEOREM: (Dilip Sequeira, 1993, [**1**])

If M's a complete metric space,
And non-empty, we know it's the case
That if f's a contraction
Then under its action
Just one point remains in its place.

PROOF: (Michael Fryers, 1994)

First suppose to the contrary two points don't move:
$f(Q)$ equals Q, $f(P)$ equals P.
Then consider the distance PQ: we can prove
That this distance is less than itself, which can't be.

Thus uniqueness; existence takes longer to get:
We'll construct such a point, fixed by f, as is sought.
Let us first take a point, say Y_0, in the set
And let Y_n be $f^n(Y_0)$.

By the triangle law, given t less than s,
Summing lengths from Y_{b-1} to Y_b
Over b more than t up to s, gives not less
Than the length to Y_s all the way from Y_t.

Now suppose that f's constant is capital C,
And the distance Y_0 to Y_1 is called k;
Then this sum is not more than the sum, a from t
Up to $(s-1)$, of kC^a.

This is less than or equal to k times the sum
Of all C^a for which a's at least t
And this last sum, by easy summation, will come
To $kC^t/(1-C)$.

So the sum tends to zero, and (Y_n) is then
Clearly Cauchy, and so it converges. Now see
That by f's continuity, $f(Y_n)$
Tends to f of the limit of Y's—call it P.

But now $f(Y_n)$ equals Y_{n+1},
So this P's fixed by f as required, so we're done. □

Reference

[1] D. J. Sequeira, *Eureka* **52** (1993) 3.

Density and Tides

Duncan Cochrane

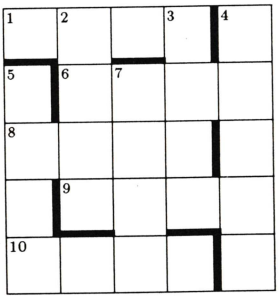
It is well known in mathematics and physics that quantities are often proportional to one another, but one particularly surprising instance of this is the following: given two bodies of the same apparent size as seen from the Earth, their density is proportional to the size of the tides they cause. A particular example of this is the Moon and the Sun, which are both approximately half a degree in diameter. But how, if we know the tides caused by the Moon are about 2·4 times larger, can we deduce the same fact about their density? For a body of a given apparent size, its radius r is proportional to its distance d. Then its mass is proportional to ρr^3, where ρ is its density. The gravitational attraction on the Earth is thus proportional to $(\rho r^3)/d^2$, or just $d\rho$. It remains to work out what the tidal effect of this is on the Earth: tides are caused by the differential pull of the other object on the two sides of the Earth. Since the Earth is small compared to the distance to either the Sun or the Moon, the pull may be assumed equal to the derivative of the gravitational attraction at the centre of the Earth multiplied by the diameter of the Earth, and this is just proportional to the density, so we are done. □

NOTE. The idea above is not original to the author; it is believed to originate with John Rickard.

A Cross-No.

Negipnu the Scribe

I wrote this crossnumber some months ago, and I can't remember where I put the across clues. I do remember however that all lights are written in standard decimal notation (with no leading zeros) and that a zero should be omitted from each light before entry in the grid. It should be noted that when a clue is referred to, the value before removal of the zero should be taken.



2	1A + 6A + 9A + 10A
3	10A/4
4	A prime power
5	$\dfrac{4\text{D} \times 8\text{A}}{7\text{D}}$
7	A factorial

Solutions to the Problems Drive

Timothy Luffingham

1. (a) 9802, 96079203. (Each term is obtained by squaring the previous one, and alternately adding or subtracting one.)

(b) 20, 21. (Numbers m such that $2m+1$ is prime.)

(c) 3, 2. (The nth term is the smallest number of squares which add to exactly n, so $12 = 4+4+4$, $13 = 9+4$.)

(d) 4, 5. (Number of consonants in ONE, TWO, THREE, etc.)

2. 1748. (The number of female numbers up to 6^n is 5^n. Work from there.)

3.

^{1} 3	^{2} 2	7	^{3} 1
0	^{4} 1	^{5} 2	3
^{6} 4	2	2	2
^{7} 4	5	2	5

(4A(base 5) = 2D(base 4), so 4A $\leqslant$ 223. Suppose 4A starts with a 2. Then 2D $\geqslant$ 320, so 6A $\geqslant$ 335. But then 2D = 323, so 6A = 641: contradiction. Therefore, 4A starts with a 1, so 121 $\leqslant$ 2D $\leqslant$ 301, so 2D starts 21$*$. Then, 6A will be 42$*$, so 2D = 212. The rest of the 3-digit answers follow. Now 7A $\leqslant$ 5555, so 1D $\leqslant$ 3531. But 3D $\geqslant$ 1300, so 1A $\geqslant$ 3236. So 1A starts with 32$**$. So 3D starts with 13$**$. Now it is a matter of checking cases.)

4. (a) 35.

(b) 28.

(For (a), for each square let $q(S)$ be the number of squares in lines through S which do not contain a counter; we want to make $q(S)$ different for each uncovered square. If there are x squares not containing counters, then $q(S)$ will take values between 0 and $x-1$. If $x > 1$, we cannot have squares with $q(S) = 0$ and $x-1$ in the same grid, however, so there will be only be $x-1$ possible values of $q(S)$. Therefore, x cannot be greater than 1. For (b), trial and error is necessary. One possible grid is:

*	*	*	4	*	*
*	*	*	*	*	*
*	*	8	*	7	2
*	6	9	*	5	*
*	*	*	*	*	*
*	*	3	*	*	*

5. Applied: A, C, and F.
Pure: B, D, and E.
Murderer: B.

6. (a) 9. (You can leave out three mutually adjacent faces.)
(b) 6. (There are lots of ways of doing this.)

7. $1+\sqrt{6}$. (The smallest tetrahedron is, in fact, one which can contain four spheres, one in each 'corner'.)

8. $4+4\pi/3-4\sqrt{3}$.

9. There are many possible matrices, for example,

$$\begin{pmatrix} 2 & 5 & 1 \\ 3 & 6 & 7 \\ 4 & 8 & 9 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 2 & 6 & 5 \\ 4 & 8 & 9 \\ 3 & 9 & 7 \end{pmatrix}.$$

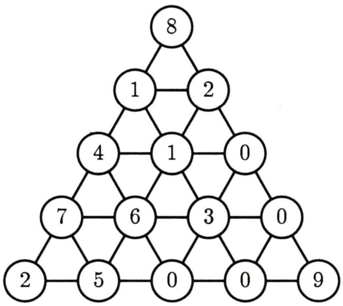
10. H, I, J, L, M, N, O, P, and Q. Possible exits are given by grids of these forms:

	T	S	R	Q	P	
A						O
B		R	R			N
C		B	B			M
D						L
E						K
	F	G	H	I	J	

or

	T	S	R	Q	P	
A						O
B	R			R		N
C	B	R	R	B		M
D		B	B			L
E						K
	F	G	H	I	J	

11.



12. (a) RIEMANN, FERMAT, EULER.
(b) EUCLID, GODEL, NEWTON.
(c) CAUCHY, CAYLEY, CONWAY.
(d) ERDOS, GAUSS, RAMANUJAN.
(e) POINCARE, GALOIS, CANTOR.