Eureka Editor
The Archimedeans
Centre for Mathematical Sciences
Wilberforce Road
Cambridge CB3 0WA
United Kingdom

archim-eureka@srcf.net

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

Thanks to the Betty & Gordon Moore Library, Cambridge

EUREKA

Spring 1984 Number 44

EUREKA

JOURNAL OF THE ARCHIMEDEANS no. 44

ISSN 0071 - 2248

CORRESPONDENCE:

(Clearly marked "Editor" or "Business Manager")

The Arts School
Bene't Street
CAMBRIDGE CB2 3PY
England

Editorial

Editing Eureka is not an easy task , the intended readership should (hopefully) have a wide range of Mathematical experience and it is difficult to find a selection of articles which will appeal to everybody. I hope that I have succeeded in this edition and that neither will the ' professional Mathematicians ' amongst you find the less mathematical articles too boring nor will the remainder of you be put off from reading the others. If anyone has any comments about Eureka or even better some articles for it , these should be addressed to the new editor : Paul Taylor (Trinity) .

The more observant readers will have perhaps noticed the now ' international ' nature of the editorship , both Nigel Boston and myself having deserted Cambridge to find research places elsewhere. In my case one of the reasons for this move was a desire to avoid the very artificial division of Mathematics in Cambridge into ' Pure ' and ' Applied ' each department having its own common room and worse still its own separate library. Surely both ' sides ' have much to gain from talking to one another ! Perhaps this point could be one of discussion in future editions of Eureka (especially from the undergraduate readership) .

Chris Budd

St. John's College Oxford

Acknowledgments

The secretaries at D.A.M.T.P.

Fiona Frank Oxford

Without whose typing skills this edition would have been impossible.

The Society
Maxim Tingley (Secretary)

The Easter term had its usual collection of croquet, punting and other social events aimed at helping members forget about the examinations.

In the new academic year, after a successful recruitment drive, the years programme got rapidly underway. Evening meetings have moved from the Arts School to the Department of Applied Mathematics and Theoretical Physics, and one was aided by computer simulations. Lunchtime meetings, instead of just adding to the number of Cambridge fellows invited by the College Societies, were also addressed by a speaker from a local engineering firm and a graduate.

The Meteorological Office, Bracknell was visited and the University's computer laboratories toured, though both with a lower than expected attendance. The annual mathematical Call my Bluff was organised in two heats, owing to the large number of teams entered, one heat being won by a Cambridge team and the other by an Oxford team though unfortunatly the two winning teams didn't compete in a final. Other social events included an afternoon of games in Oxford and a barn dance.

This year QARCH has come out of hibernation, its annual frequency having dropped from 7 to the more manageable 3. The Computer Group and Puzzles and Games Ring continue to operate and have been joined by a Geometrical Models Group and an Othello Group. In addition, the bookshop, because of its recent financial success, now also buys and sells physics and computer science books.

The puzzles of Lewis Carroll

R. Sabey

Those who have read no more of Lewis Carroll's works than the Alice books may have surmised that he was fond of mathematics and logic, whether because of Humpty Dumpty's finding it difficult to believe that 365-1 is 364, the simple logic of Tweedledum and Tweedledee, or puns like "Ambition, Distraction, Uglification and Derision".

This is in fact correct. Charles Lutwidge Dodgson (to give him his real name) had been interested in mathematics from his teens. He gained first-class honours in mathematics at Christ Church, Oxford, and became a "Student" [i.e. Fellow] at that college. He was later appointed Lecturer, but his lectures were dull and were made worse by the fact that he had a stammer.

Under his real name, Carroll published many mathematical books, but these were nearly all elementary textbooks, "dry as dust", and they contained nothing new. In fact his mathematical ideas were old-fashioned, even for their time: he didn't understand calculus or infinitesimals, and in his "Euclid and His Modern Rivals", where he defends Euclid against moves then modern to update him, he ridiculed those who disputed the parallel postulate, thus denying Riemannian and Lobachevskyan geometry. He did, however, experiment with topology, and, in that cornucopia of Carrollian gems, the "Sylvie and Bruno" books, explains how to sew three square handkerchiefs together to make a Klein bottle. (In fact in the book this never gets completed -- which is just as well, since a Klein bottle cannot be embedded in 3-space without its surface intersecting itself!)

But Carroll often approached mathematics and logic in a lighter frame of mind, and managed to bring humour to these potentially esoteric subjects; it is in this area where Carroll's work is most enlightening. The greatest collections of his mathematical puzzles are contained in his two books "Pillow-Problems" and

"A Tangled Tale". The former consists of 72 mathematical problems, mostly in algebra, geometry and trigonometry; each one was thought out and solved while he lay awake in bed -- hence the title. The latter contains ten chapters, called "knots", each of which is a story, into which Carroll works in a problem or two; the knots were originally published in a magazine, readers of which sent in solutions to the problems. The problems are mostly easy to solve, but most of the interest lies in the Appendix, where he sets up a chatty style with his correspondents, analysing their solutions, showing how the problems should and should not be tackled.

While he was a "Student" at Christ Church, Oxford, he produced several humorous pamphlets relating to topical Oxford matters, of which two parody serious mathematical papers, involving much play with mathematical language. In "The New Method of Evaluation", this is little more than making people's initials from letters representing quantities and points, but in "Dynamics of a Parti-cle", Carroll is more ambitious, and shows how skilful he was at the art of punning, as can be seen from these parodies of Euclid:

PLAIN SUPERFICIALITY is the character of a speech, in which any two points being taken, the speaker is found to lie wholly in regard to those two points.

Let is be granted that a controversy may be raised about any question, and at any distance from that question.

Carroll invented a method of solving syllogisms using a special board and counters, and taught it in his "Game of Logic". His "Symbolic Logic" takes the subject further, and treats it more seriously. An important part of this book is its Appendix, where he reveals great flaws in the classical form of logic; his clearly-put views make a refreshing change from the thinking, too restricted and sometimes muddled, of classical logicians. In both these books, as if to compensate for the tedious descriptions and strange terminology, Carroll rewards the reader, when he has grappled with some logical idea, with a plentiful supply of charming examples and problems, where from sets of premisses he is to draw conclusions like "No kitten with green eyes will play with a gorilla" or "Guinea-pigs never really appreciate Beethoven".

Freed from the strait-jacket of the formal syllogism, Carroll could produce logical paradoxes natural enough to blend into his children's books. For example, in "Alice's Adventures in Wonderland", Alice must distinguish between one side and the other of a round mushroom. In "Through the Looking-Glass", we get even more examples. The White Queen "lives backwards", and so her memory works both ways. So does the Hatter, who is imprisoned for a crime he will commit later -- would it be better, as the Queen said, if he never commits it? The White Queen offers her servants "jam every other day", but they never get any because "today isn't any other day". Plum-cake is offered round, and then cut. Two eggs are cheaper than one. Humpty Dumpty prefers unbirthday presents to birthday presents. And in the "Sylvie and Bruno" books, we find a Professor who thinks barometers affect the weather; he has also invented a way of carrying himself, which will not be tiring, as whatever energy he expends by carrying he saves by being carried. Bruno says he can see "about a thousand and four" pigs in a field because he's sure about the four, even though he isn't sure about the thousand. "Mein Herr" talks of a packing material of negative density that can make parcels weigh less than nothing, so that one receives money for sending them through the post.

Lewis Carroll was not a great mathematician, but he had a passion for setting and solving problems, and a flair for presenting logical paradoxes in disguises that make them much more palatable to the reader than any technical explanation of logical fallacies. And the fact that he, "an obscure writer on logic", as he called himself, could see clearly in the world of elementary logic where logicians were still unsure of themselves is yet another of the puzzles of Lewis Carroll.

for further reading:

1. The Penguin Complete Works of Lewis Carroll, ed. Alexander Woolcott

2. "Pillow Problems" and "A Tangled Tale", ed. Dover

3. "Symbolic Logic" and "The Game of Logic", ed. Dover

4. "The Unknown Lewis Carroll", ed. Dover

5. "The Magic of Lewis Carroll", ed. John Fisher

6. "The Annotated Alice", ed. Martin Gardner

Colson News

We are intending to produce a small newsletter about the use of negative digits. It will contain information about their properties and advantages, comments, problems, puzzles and articles on related topics. We aim to have four issues per year, beginning on New Year's Day 1984, at a subscription of £1 per year (at least initially) to cover the costs of printing, postage and incidentals. Enquiries and subscriptions to:

Cedric A. B. Smith
University College London
4 Stephenson Way
London NW1 2HE

CHROMATIC POLYNOMIALS

An alternative approach to the Four-Colour Theorem?

by D. R. Woodall

Chromatic polynomials were introduced by G. D. Birkhoff [4] in 1912 as a means of tackling the (then) four-colour problem. They have now acquired considerable interest in their own right, and the four-colour theorem has been proved without them; nevertheless, it is the purpose of this article to suggest that there would still be great value in carrying through Birkhoff's original intention.

As most readers will know, the four-colour theorem states that, using a total of at most four different colours, one can colour the regions of every map in the plane (giving just one colour to each region) in such a way that regions that share a length of common border always have different colours. This was conjectured by Francis Guthrie in 1852 and was finally proved by Kenneth Appel and Wolfgang Haken [1] in 1976. The proof relies on an extensive case-by-case analysis by computer, involving the study of well over 1000 separate configurations. It is thus very long, and does not give much insight into why the result is true. Although it was a tremendous achievement by Appel and Haken to prove the result at all, it would certainly be of great interest to have a shorter or more illuminating proof.

Birkhoff's idea was to approach the question of *whether* a given map could be coloured with a specified number of colours by considering, more precisely, *in how many different ways* it could be coloured. However, before explaining his idea, let

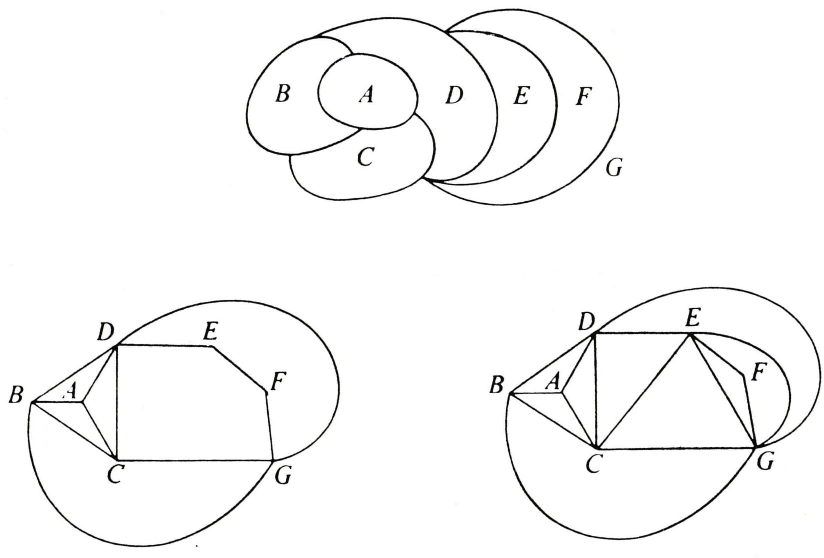


Figure 1

us translate the problem from the terminology of maps, used by Guthrie and Birkhoff, into the more modern terminology of graphs. The four-colour theorem can be formulated as follows: using a total of at most four different colours, one can colour the vertices of every graph drawn 'properly' (that is, without edges crossing) in the plane, giving just one colour to each vertex, in such a way that two vertices joined by an edge always have different colours; such a colouring will be called a **proper 4-colouring**. Moreover, to prove this, it suffices to prove it in the special case when the graph is a **plane triangulation**: that is, when every region formed by the graph, including the unbounded region, is bordered by exactly three edges. It is an elementary exercise to show that these different versions of the theorem are all equivalent. As an illustration, Figure 1 shows a map, a graph equivalent to it, and the same graph made into a triangulation.

Now let G be any graph and t be any positive integer, and define $P(G,t)$ to be the number of different proper t-colourings of G. We shall adopt the convention that $P(G,t) = 0$ if G contains a **loop** (that is, an edge joining a vertex to itself), since we clearly cannot colour a vertex with a different colour from itself. Some simple properties of $P(G,t)$ are easy to see.

Lemma 1.1. *If G has p vertices and no edges, then $P(G,t) = t^p$.*

Proof. Since there are no edges, each of the p vertices can be given any one of the t colours quite independently of the colours that are given to the other vertices, and so the total number of different colourings is t^p.

Lemma 1.2. The Deletion-Contraction Algorithm. *If e is an edge of a graph G, let $G\backslash e$ be the graph obtained from G by deleting the edge e, and let G/e be the graph obtained from G by contracting the edge e; that is, by deleting e and then identifying its former end-vertices (see Figure 2). Then $P(G,t) = P(G\backslash e,t) - P(G/e,t)$.*

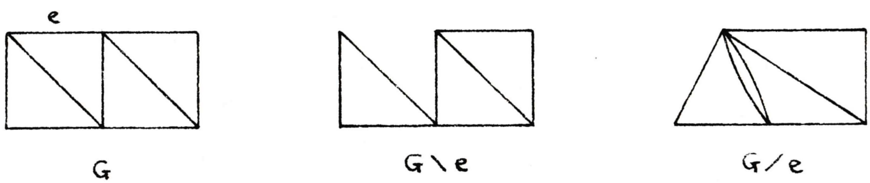


Figure 2

Proof. Let the end-vertices of e be v and w. If $v = w$, then $G\backslash e$ and G/e are the same graph; thus the lemma accords with our convention that $P(G,t) = 0$ if G contains a loop. So suppose $v \neq w$. The number of proper t-colourings of $G\backslash e$ in which v and w have *different* colours is just $P(G,t)$, since the effect of restoring the edge e is precisely to force v and w to have different colours; and the number in which v and w have *the same* colour is just $P(G/e,t)$, since the effect of identifying v and w is precisely to force them to have the same colour. Thus $P(G\backslash e,t) = P(G,t) + P(G/e,t)$, as required.

These two lemmas give an easy proof of the following fundamental theorem.

Theorem 1. *If G has no loops, then $P(G,t)$ is a monic polynomial in t whose degree is equal to the number p of vertices of G.* It is called the **chromatic polynomial**, or (regrettably) the **chromial**, of G.

Proof by induction on the number of edges of G. If G has no edges, then the result follows from Lemma 1.1. If G has an edge, then we can suppose inductively that $P(G\backslash e,t)$ and $P(G/e,t)$ are monic polynomials of degree p and $p-1$ respectively, and then the result follows from Lemma 1.2.

In view of this theorem, it makes sense to evaluate $P(G,t)$ at non-integer values of t, and so to make statements like $P(G,\sqrt{5}) = -\frac{1}{2}(1+\sqrt{5})$. Of course, this is not to be interpreted as saying that there are $-\frac{1}{2}(1+\sqrt{5})$ ways of colouring G with $\sqrt{5}$ colours; it is purely a statement about the value of a certain polynomial function at a particular value of its argument. It is only when t is a positive integer that $P(G,t)$ represents the number of ways of colouring G with t colours; this can be extended to $t = 0$ by the observation that $P(G,0) = 0$ for every graph G, which can easily be proved inductively in the same way as Theorem 1. For the record, R. P. Stanley [7] has found a combinatorial interpretation of $P(G,t)$ whenever t is a *negative* integer; in particular, $P(G,-1)$ is $(-1)^p$ times the number of ways in which one can direct all the edges of G without creating any cycles. As far as I know, nobody has discovered a combinatorial interpretation of $P(G,t)$ for any non-integer value of t.

The four-colour theorem can now be reformulated as follows: 4 is not a zero of the chromatic polynomial of any plane triangulation. This suggests that one might be able to discover a route into the theorem by examining the distribution of these zeros. This was done by Berman and Tutte [3], who used a computer to plot the zeros of the chromatic polynomials of hundreds of plane triangulations. They discovered that there always seems to be a real zero at about 2.61803..., and another near 3.246..., the closeness of approximation to these values seeming to increase as the number of vertices in the triangulation increases. (Some lists of chromatic polynomials of plane triangulations have been published [2,5]; a study of these suggests that, in order to achieve agreement to the number of decimal places quoted above, one should look at 4-connected triangulations without vertices of valency 4 and with at least 20 vertices in total.)

Since the well-known Golden Ratio has the value $\tau = 1.61803...$, Berman and Tutte suggested that the 'true value' of 2.61803... is the larger root of the equation $t^2-3t+1 = 0$, which is $\frac{1}{2}(3+\sqrt{5})$, or $1+\tau$. So the suggestion is that the chromatic polynomial is 'trying to have a zero' at $1+\tau$, but not quite making it. It has become fashionable to refer to this zero as the 'golden root' of the chromatic polynomial. D. W. Hall has suggested that the 'true value' of 3.246... may be the largest root of the equation $t^3-5t^2+6t-1 = 0$, since this polynomial features prominently in various calculations involving chromatic polynomials. Professor Tutte has suggested that the zero that occurs near this value should be called the 'silver root' of the chromatic polynomial.

A further step towards identifying these numbers was taken by S. Beraha, who pointed out that $1+\tau = 2+2\cos 2\pi/5$, and the largest root of the cubic mentioned above can be written as $2+2\cos 2\pi/7$. This suggests that one should define the **Beraha numbers** to be $B_n = 2+2\cos 2\pi/n$. The first few values of B_n are given in the table. For $n = 2$, 3 and 4, we get simply 0, 1 and 2, where there are

n	B_n	n	B_n
2	0	7	3.246...
3	1	8	3.414... $(= 2+\sqrt{2})$
4	2	9	3.532...
5	2.61803... $(= 1+\tau)$	10	3.618... $(= 2+\tau)$
6	3	...	...

always zeros. (By its very nature, a plane triangulation always contains a triangle, and the three vertices of a triangle cannot be coloured with fewer than three colours.) For $n = 5$ we get $1+\tau$, near which there is almost always a zero. For $n = 6$ we get 3, where there is usually a zero. (The only plane triangulations that can be coloured with three colours are those in which every vertex has even valency.) For $n = 7$ we get 3.246..., near which zeros have been observed to occur. For larger values of n we get a sequence of numbers between 3 and 4, and $\lim_{n\to\infty} B_n = 4$. This last fact is probably significant: it suggests that one might be able to approach the four-colour theorem by proving suitable results about values of chromatic polynomials of plane triangulations at the Beraha numbers.

In fact, I know of only two theorems about these values, both proved by Tutte in 1970. The first [8] states simply:

Theorem 2. *If G is a plane triangulation with p vertices, then*

$$|P(G,1+\tau)| \leqslant \tau^{5-p}.$$

The significance of this result is that $\tau > 1$, and so τ^{5-p} is very small if p is large. Thus the value of the chromatic polynomial at B_5 is very small in absolute value. This does not prove that there is a zero nearby—indeed, there may not be—but it makes the proximity of a zero seem quite likely. Nobody has been able to prove that, apart from known exceptions, there is always a zero near B_5; nor has anyone been able to prove a theorem analogous to Theorem 2 for B_7 rather than B_5.

The second result [9] relates the value of the chromatic polynomial at B_{10} $(= 2+\tau)$ to its value at B_5 $(= 1+\tau)$.

Theorem 3. *If G is a plane triangulation with p vertices, then*

$$P(G,2+\tau) = (2+\tau)^{3p-10}[P(G,1+\tau)]^2.$$

This does not ensure that the value of the chromatic polynomial at B_{10} is small; what it does ensure, much more interestingly, is that it is always *positive*. Now, the four-colour theorem asserts that the chromatic polynomial is always positive at 4. Is it possible that it is always positive throughout the range from B_{10} to 4?

Unfortunately not: there are examples of such polynomials that dip down below the axis between B_{10} and 4. However, let us speculate fancifully for a moment. Suppose that by generalizing Theorem 3, or otherwise, one could show that there are infinitely many Beraha numbers at which the chromatic polynomial is always positive. Suppose also that one could show that there are infinitely many Beraha numbers at which the *derivative* of the chromatic polynomial is always positive. Then one would have proved the four-colour theorem. (Incidentally, it is conjectured that not only the polynomial, but all its derivatives up to and including the pth, are positive at 4; nobody has been able to prove this, although it is known [5] to be true at 5.)

The scenario described above may justly be called 'fanciful', since there is not the slightest evidence that results like Theorems 2 and 3 hold for other Beraha numbers, and there is even some evidence that they don't. Moreover, I know of no theorems at all involving derivatives of chromatic polynomials. Nevertheless, it seems very striking that, when one examines the zeros of the chromatic polynomials of plane triangulations—precisely what is relevant to the four-colour theorem—one discovers a sequence of numbers, quite out of the blue, that have 4 as their limit. The scenario described above may be too simplistic; but it is difficult to resist the feeling that there must be *some* way of using these numbers in order to prove the four-colour theorem.

References

References 6, 10 and 11 are to papers that are at least partly expository.

1. K. Appel and W. Haken, Every map is four colorable, *Illinois J. Math.* **21** (1977), 429–567 (plus microfiche supplements).
2. Ruth A. Bari, Regular major maps of at most 19 regions and their Q-chromials, *J. Combinatorial Theory Ser. B* **12** (1972), 132–142.
3. G. Berman and W. T. Tutte, The golden root of a chromatic polynomial, *J. Combinatorial Theory* **6** (1969), 301–302.
4. G. D. Birkhoff, A determinant formula for the number of ways of colouring a map, *Ann. Math.* **14** (1912), 42–46.
5. G. D. Birkhoff and D. C. Lewis, Chromatic polynomials, *Trans. Amer. Math. Soc.* **60** (1946), 355–451.
6. R. C. Read, An introduction to chromatic polynomials, *J. Combinatorial Theory* **4** (1968), 52–71.
7. R. P. Stanley, Acyclic orientations of graphs, *Discrete Mathematics* **5** (1973), 171–178.
8. W. T. Tutte, On chromatic polynomials and the golden ratio, *J. Combinatorial Theory* **9** (1970), 289–296.
9. W. T. Tutte, The golden ratio in the theory of chromatic polynomials, *Ann. New York Acad. Sci.* **175** (1970), 391–402.
10. W. T. Tutte, Chromials, *Springer Lecture Notes in Mathematics* **411** (1974), 243–266.
11. D. R. Woodall, Zeros of chromatic polynomials, "Combinatorial Surveys" (Academic Press, London, 1977), 199–223.

Problems Drive 1981

P. Taylor & C. Feather

The Society has held a Problems Drive in most of the years of its existence: usually in Cambridge but occasionally as the guests of the Invariants in Oxford. Rumour has it that there was a year in which they actually won! Since 1982 there have also been participants from other places such as Warwick and Southampton.

Members compete in pairs: the winning pair receives a bottle of Port and is expected to set the problems for the following year. In accordance with the traditions of the ancient Tripos, the lowest scorers are awarded the coveted Wooden Spoon, although it may be said that in recent years this has been a bit self-defeating since some people have competed directly for it. A teabag has also been offered by the Editor of *Eureka* for the most 'original' solution. The problems are not printed on a single sheet as in any ordinary examination, but written individually on cards: these are circulated amongst the competitors, who see each one for just five minutes.

The following problems were set by Clive Feather and myself in 1981 and have been discussed in *QARCH*, but for various reasons missed *Eureka*. They are probably the most difficult which have been set in recent years: the highest score (about one third of the total available) was achieved by Nick Inglis and Richard Pennington. Problem 2 was the subject of an article in *2-Manifold* 4 (1983) 34-38.

The next competition will be held in Cambridge at approximately the time this *Eureka* is being printed; the problems will be set by Andy Bernoff and Richard Pennington (for the third time!) and will appear in *Eureka* **45**. In my opinion the competition *should* be restricted to *under*graduates.

Paul Taylor

1. Calculate the resistance between each pair of vertices in the following network of 1Ω resistors.

(i) 6 resistors arranged to form the skeleton of a tetrahedron. (one answer) (figure 1)

(ii) 12 to form an octahedron. (two answers)

(iii) 12 to form a cube. (three answers)

(iv) 30 to form an icosahedron. (three answers)

Rider: (v) 30 to form a dodecahedron.

2. In a finite <u>*k*-farm</u> there are k commutative and associative operations, each of which has an identity. The i^{th}operation distributes over the $(i-1)^{th}$, and under the i^{th} operation the elements of the farm excluding the identities of the 1^{st} to $(i-1)^{th}$ form an abelian group. Find a 4-farm with four or more elements. *Hint: the number of elements of a finite field is a prime or a power of a prime.*

Rider: is there a finite 5-farm? Give reasoning.

3. Dissect the shapes in figures 2 and 3 into three congruent parts each.

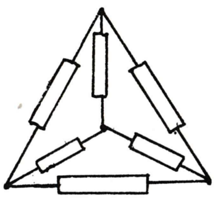


fig.1

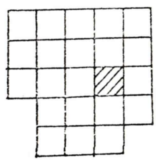


fig.2

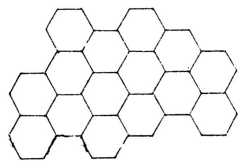


fig. 3

4. A tent is made by hanging a piece of canvass of length 2ℓ and width w symmetrically over a pole of length w suspended a height h above flat horizontal ground. The ens of the canvass are fixed a distance a either side of the line vertically underneath the pole. Find the volume of the tent. Obviously $h^2+a^2<\ell^2$.

5. (i) Factorise $2^{23}-1$. (ii) Find a six-digit palindromic square.

6. A spacecraft consists of a cylinder with conical ends mounted on a delta (Δ) wing. At each wing tip a 10 Mg motor is mounted, and at a distanco x motres from tho noso, on the axis of the cylinder, is a 15 Mg gonorator. The cylinder is 100 m long, including tho onds, which are 5 m long and solid, and 6 m in oxternal diamotor, and in tho main body it is 0.1 m thick. Tho wing is 100 m long, 80 m across and 0.1 m thick. The matorial has density 8 Mg m^{-3}. Whore is tho contre of gravity?

7. What is the area of the ellipse?

$$ax^2 + by^2 + 2hxy - 2fx - 2gy + c = 0$$

8. Find the cube roots of the following numbers; all solutions are integral, although the problem as originally posed included one which wasn't! Calculators are not allowed in the Problems Drive. (i) 50653 (ii) 148877 (iii) 551368 (iv) 102503232 (v) 162771336 (vi) 498677257 (vii) −565609283 Rider: give the fifth roots of (viii) 2476099 (ix) 3486784401 (x) 58602385427607

9. A triangle has sides a, b, c. The radii of the inscribed and circumscribed circles are r, R respectively, and $2s=a+b+c$. Give reasonably neat formulae in terms of R, r, s for $a^n+b^n+c^n$ for $n=0, 1, 2, 3, 4$ and 5.
Rider: $n=6$. You may give this answer in terms of the previous ones.

10. Express $\sin 9^\circ$ in the form $\sqrt{(a\pm\sqrt{(b\pm\sqrt{c})})}$ with a, b, c rational. You may choose either sign for each root. Interpret the other choices of sign.

11. Let $A=1976^{1976}$. Let B be the sum of the digits of A, C the sum of the digits of B and D the sum of the digits of C. What is D?
Rider: make a *guess* at the value of D if $A=1981^{1981}$, stating your assumptions. What is the probability you are wrong?

12. What is the length of the "longest day" (summer solstice) at latitude θ, assuming the earth to be a perfect sphere inclined at angle α to its orbit, which is circular and large compared to the earth's diameter?

The answers will be found on page 71.

Egoritchev's proof of van der Waerden's conjecture. Béla Bollobás

An $n \times n$ matrix (a_{ij}) is doubly stochastic if $a_{ij} \geq 0$ for all i and j, and

$$\sum_j a_{ij} = \sum_j a_{ji} = 1 \quad \text{for every} \quad i .$$

In 1926 van der Waerden conjectured that if $A = (a_{ij})$ is an $n \times n$ doubly stochastic matrix then its permanent is at least $n!/n^n$:

$$\operatorname{per} A = \sum_{\pi} \prod_{i=1}^{n} a_{i\pi(i)} \geq \frac{n!}{n^n} .$$

Furthermore, equality holds iff $A = (1/n)$ that is $a_{ij} = \frac{1}{n}$ for every n.

After several rather weak partial results Friedland proved in 1979 that the permanent is at least $1/n!$ and a year later he improved it to e^{-n}. Thus Friedland proved the conjecture but for a factor of about $(2\pi n)^{1/2}$. Nevertheless, the next improvement did not come by refining the proof of Friedland. Recently two Russians gave independent and different proofs of van der Waerden's conjecture: Falikman and Egoritchev. Here we shall present the proof of Egoritchev, which is based on an inequality concerning "mixed volumes" proved by Alexandrov in 1938.

Given vectors $x_1, x_2, \ldots, x_k \in \mathbb{R}^n$, denote by $(x_1, x_2, \ldots, x_k)$ the $n \times k$ matrix whose i^{th} column is x_i, $1 \leq i \leq k$. Write $\mathbb{R}^n_+$ for the set of vectors $(x_i) \in \mathbb{R}^n$ with $x_i > 0$ for every i.

<u>Theorem 1.</u> (Alexandrov) Let $n \geq 2$, $a_1, \ldots, a_{n-2} \in \mathbb{R}^n_+$ and let

$$\phi(u,v) = \mathrm{per}(a_1,\ldots,a_{n-2},u,v) .$$

Then if $\phi(x,x) > 0$, we have

$$\phi(x,y)^2 \geq \phi(x,x)\phi(y,y)$$

for every $y \in \mathbb{R}^n$.

Proof. Let ψ be a symmetric bilinear form on $\mathbb{R}^n$. The following three assertions are equivalent:

(a) $\psi(x,x) > 0$ for some x and if $\psi(x,x) > 0$ then $\psi(x,y)^2 \geq \psi(x,x)\psi(y,y)$, with equality iff $y = \gamma x$,

(b) $\psi(x,x) > 0$ for some x and if $\psi(x,x) > 0$ and $\psi(x,y) = 0$ then $\psi(y,y) \leq 0$, with equality iff $y = 0$,

(c) ψ is nonsingular and has exactly one positive eigenvalue.

To see the equivalence of these assertions note that (a) implies (b) trivially. Furthermore, if (b) holds then ψ is nonsingular and on no 2-dimensional subspace is it positive definite, so it has exactly one positive eigenvalue. Finally, if (c) holds, $\psi(x,x) > 0$ and $y \neq \gamma$ then $\psi(\mu x+y,\mu x+y) = \mu^2\psi(x,x) + 2\mu\,\psi(x,y) + \psi(y,y) < 0$ for some $\mu \in \mathbb{R}$ so $\mu(x,y)^2 > \psi(x,x)\psi(y,y)$.

Now with $j = (1,1,\ldots,1)^t$ the symmetric bilinear form ϕ satisfies $\phi(j,j) > 0$ so our theorem states that ϕ has the equivalent properties (a), (b), and (c). The matrix of ϕ is easily determined. Denote by $A(i_1,\ldots,i_k|j_1,\ldots,j_\ell)$ the matrix obtained from a matrix A by omitting rows $i_1,\ldots,i_k$ and columns $j_1,\ldots,j_\ell$. For $n \geq 3$ and $1 \leq k,j \leq n$ set

$$q_{ij} = \begin{cases} \operatorname{per}(a_1,\ldots,a_{n-2})(i,j|\emptyset) & \text{if} \quad i \neq j \\ 0 & \text{if} \quad i = j \end{cases}$$

For $n = 2$ put $q_{11} = q_{22} = 0$ and $q_{12} = q_{21} = 1$. Then $Q = (q_{ij})$ is a symmetric $n \times n$ matrix and for $x,y \in \mathbb{R}^n$ we have

$$\phi(x,y) = \operatorname{per}(a_1,\ldots,q_{n-2},x,y) = \sum_i \operatorname{per}(a_1,\ldots,a_{n-2},x,y)(i|n)y_i$$

$$= \sum_i y_i \sum_{j\neq 1} \operatorname{per}(a_1,\ldots,a_{n-2},x,y)(i,j|n-1,n)x_j$$

$$= \sum_i t_i \sum_{j\neq 1} \operatorname{per}(a_1,\ldots,a_{n-2})(i,j|\emptyset)x_j$$

$$= \sum_{i,j} y_i\, q_{ij}\, x_j = y^tQx .$$

After all these preparations let us proceed to the proof of Theorem 1. We apply induction on n. For $n = 2$ the eigenvalues of Q are 1 and -1 so (c) holds. Assume now that $n \geq 3$ and the theorem holds for $n - 1$.

Suppose that Q is singular, say $Qu = 0$ for some $u \in \mathbb{R}^n$, $u \neq 0$. Then for every $v = (v_i) \in \mathbb{R}^n$ we have

$$v^tQu = \phi(u,v) = \sum_i \operatorname{per}(a_1,\ldots,a_{n-2},u,v)(i|n)v_i$$

$$= \sum_i \operatorname{per}(a_1,\ldots,a_{n-2},u)(i|\emptyset)v_i$$

so for $1 \leq i \leq n$ we have

$$\operatorname{per}(a_1,\ldots,a_{n-2},u)(i|\emptyset) = 0 .$$

Hence by the induction hypothesis

$$\operatorname{per}(a_1,\ldots,a_{n-3},u,u)(i|\emptyset) \le 0 , \tag{1}$$

with equality iff $u_j = 0$ for every $j \ne i$.

Since we also have

$$\sum_i a_{i,n-2} \operatorname{per}(a_1,\ldots,a_{n-3},u,u)(i|\emptyset)$$

$$= \sum_i a_{i,n-2} \operatorname{per}(a_1,\ldots,a_{n-3},u,u,a_{n-2})(i|n)$$

$$= \operatorname{per}(a_1,\ldots,a_{n-3},u,u,a_{n-2}) = \operatorname{per}(a_1,\ldots,a_{n-2},u,u)$$

$$= u^t Qu = 0 ,$$

equality holds in (1) for every i with $a_{i,n-2} > 0$. Hence $u_j = 0$ for every i. This shows that Q is <u>non-singular</u>.

A simple homotopy argument shows that Q has exactly one positive eigenvalue. For $0 \le \theta \le 1$ let Q_θ be the $n \times n$ matrix defined by

$$y^t Q_\theta x = \operatorname{per}(\theta a_1 + (1-\theta)\underline{j},\ldots, a_{n-2} + (1-\theta)j, x,y) .$$

Then, as we have just shown, no Q_θ is singular and $Q_1 = Q$ so Q has exactly as many positive eigenvalues as Q_0. Clearly

$$Q_0 = (n-2)!(J_n - I_n) ,$$

where J_n is the $n \times n$ matrix all whose entries are 1s and I_n is the identity matrix. Since the spectrum of J_n is $\{n,0,0,\ldots,0\}$, the spectrum of Q_0 is $(n-2)!\{n-1,-1,-1,\ldots,-1\}$. Therefore Q has indeed exactly one positive eigenvalue. This completes the proof of (c) and so that of our theorem. $\square$

It is worth noting the following immediate consequence of Theorem 1 and the continuity of the permanent: if $a_1,\ldots,a_{n-2}$, x and y are non-negative vectors then

$$\operatorname{per}(a_1,\ldots,a_{n-2},x,y)^2 \geq \operatorname{per}(a_1,\ldots,a_{n-2},x,x)\ \operatorname{per}(a_1,\ldots,a_{n-2},y,y)\ . \qquad (2)$$

Let us turn now to the proof of van der Waerden's conjecture. Denote by Ω_n the set of doubly stochastic $n \times n$ matrices. We say that $A \in \Omega_n$ is a <u>minimizing matrix</u> if per attains its minimum on Ω_n at A. It is easily seen that a minimizing matrix is fully indecomposable, that is no rearrangement of its rows and columns is the direct sum of two matrices. This is particularly easy to show if we assume van der Waerden's conjecture up to $n - 2$, say. The following important property of a minimizing matrix was proved by London in 1971.

<u>Lemma 2.</u> If A is a minimizing matrix then

$$\operatorname{per} A(i|j) \geq \operatorname{per} A$$

for all i,j, and equality holds if $a_{ij} > 0$.

<u>Proof.</u> Ω_n is the set of matrices $A = (a_{ij})$ satisfying

$$\sum_i a_{ij} = 1 \quad \text{and} \quad \sum_i a_{ji} = 1 \quad \text{for every } j$$

and

$$a_{ij} \geq 0 \quad \text{for all} \quad i \text{ and } j\ .$$

For simplicity we write $A_{ij} = A(i|j)$. Clearly

$$\frac{\partial}{\partial a_{ij}} \operatorname{per} A = \operatorname{per} A(i|j) = \operatorname{per} A_{ij}\ .$$

Therefore by the Kuhn-Tucker theorem of optimization theory there are constants λ_i, μ_j and ρ_{ij} such that

$$\rho_{ij} \geq 0 \quad \text{and} \quad \rho_{ij}a_{ij} = 0 \quad \text{for all} \quad i,j\ ,$$

and

$$\text{per } A_{ij} + \lambda_i + \mu_j - \rho_{ij} = 0 \quad \text{for all} \quad i,j \, . \tag{3}$$

Consequently

$$\sum_j a_{ij} \text{ per } A_{ij} + \sum_j \lambda_i a_{ij} + \sum_j \mu_j a_{ij} = \sum_j \rho_{ij} a_{ij}$$

and so

$$\text{per } A + \lambda_i + (A_\mu)_i = 0 \, ,$$

and

$$(\text{per } A) \, \underline{j} + \underline{\lambda} + A\underline{\mu} = 0 \tag{4}$$

where $\underline{\mu} = (\mu_i)$, $\underline{\lambda} = (\lambda_i)$ and $\underline{j} = (1,1,\ldots,1)^t$.

Similarly we obtain

$$(\text{per } A)\underline{j} + \underline{\mu} + A^t\underline{\lambda} = 0 \, .$$

Solving this equation and substituting it into (4) we find that

$$\underline{\mu} = -A^t\underline{\lambda} - \text{per}(A)\underline{j} \, ,$$

and

$$(\text{per } A)\underline{j} + \underline{\lambda} = AA \, \underline{\lambda} + (\text{per } A)\underline{j} \, ,$$

that is

$$AA^t\underline{\lambda} = \underline{\lambda} \, .$$

Since A is fully indecomposable, AA^t is a fully indecomposable symmetric doubly stochastic matrix. Hence the multiples of $j = (1,1,\ldots,1)$ are the only eigenvectors with eigenvalue 1 . Therefore $\lambda_1 = \lambda_2 = \ldots = \lambda_n = \lambda$ and , analogously, $\mu_1 = \mu_2 = \ldots = \mu_n = \mu$.

With this new information (3) becomes

$$\operatorname{per} A_{ij} + \lambda + \mu = \rho_{ij} \geq 0 \tag{3'}$$

so

$$\sum_{i,j} a_{ij} \operatorname{per} A_{ij} + \lambda \sum_{i,j} a_{ij} + \mu \sum_{i,j} a_{ij} =$$

$$n \operatorname{per} A + n\lambda + n\mu = 0 \ . \tag{5}$$

Comparing (3') and (5) we find that

$$\operatorname{per} A_{ij} \geq \operatorname{per} A \quad \text{for all} \quad i,j$$

and equality holds whenever $\rho_{ij} = 0$ so, in particular, when $a_{ij} > 0$. □

The gist of Egoritchev's proof of van der Waerden's theorem is that the Alexandrov inequality can be used to boost London's lemma to a stronger result.

Lemma 3. If A is a minimizing matrix then

$$\operatorname{per} A(i|j) = \operatorname{per} A \quad \text{for all} \quad i,j \ .$$

Proof. Suppose $\operatorname{per} A(1|1) > \operatorname{per} A$. Since A is fully indecomposable, we may assume that $a_{12} > 0$. Then by (2) we have

$$(\operatorname{per} A)^2 = \operatorname{per}(a_1,\ldots,a_n)^2 \geq \operatorname{per}(a_1,a_1,a_3,\ldots,a_n) \ \operatorname{per}(a_2,a_2,a_3,\ldots,a_n)$$

$$= (\sum_i a_{i1} \operatorname{per} A(i|2)) \ (\sum_i a_{i2} \operatorname{per} A(i|1))$$

$$\geq \operatorname{per} A(\operatorname{per} A + a_{12}(\operatorname{per} A(1|1) - \operatorname{per} A)\}$$

$$> (\operatorname{per} A)^2 \ .$$

This contradiction completes the proof. □

As another immediate consequence of Theorem 1 we find that van der Waerden's conjecture follows if we show that a minimizing matrix has only strictly positive entries.

Lemma 4. If $A = (a_1,\ldots,a_n)$ is a minimizing matrix and $a_i \in \mathbb{R}^n_+$ for every $i \geq 1$ then $A = \frac{1}{n} J$.

Proof. The sequence of inequalities in the proof of Lemma 3 implies that

$$\operatorname{per}(a_1,\ldots,a_n)^2 = \operatorname{per}(a_1,a_1,a_3,\ldots,a_n)\operatorname{per}(a_2,a_2,a_3,\ldots,a_n) .$$

Hence by Theorem 1 $a_1 = a_2$. Similarly we find that all column vectors are equal so $A = \frac{1}{n} J$. □

We are about to complete our preparations: the final lemma shows that a rather simple operation transforms a minimizing matrix into a minimizing matrix.

Lemma 5. Let $A = (a_1,\ldots,a_n)$ be a minimizing matrix and let A' be obtained from A be replacing a_i and a_j by $\frac{1}{2}(a_i+a_j)$. Then A' is also minimizing.

Proof. The assertion follows immediately if we expand $\operatorname{per} A'$ and apply Lemma 3:

$$\operatorname{per} A' = \frac{1}{2} \operatorname{per} A + \frac{1}{4} \operatorname{per}(a_1,\ldots,a_i,\ldots,a_i,\ldots,a_n) + \frac{1}{4} \operatorname{per}(a_1,\ldots,a_j,\ldots,a_j,\ldots,a_n)$$

$$= \frac{1}{2} \operatorname{per} A + \frac{1}{4} \sum_k a_{ki} \operatorname{per} A(k|i) + \frac{1}{4} \sum_k a_{kj} \operatorname{per} A(k|j)$$

$$= \frac{1}{2} \operatorname{per} A + \frac{1}{4} \operatorname{per} A + \frac{1}{4} \operatorname{per} A . \quad \square$$

Now we are all set to prove van der Waerden's conjecture.

Theorem 6. $\frac{1}{n} J$ is the only minimizing matrix.

Proof. Suppose $A \neq \frac{1}{n} J$ is a minimizing matrix. Then by Lemma 4 A has at least two columns containing 0's. We can find a sequence of matrices $(A_i)_1^k \subset \Omega_n$ such that $A_1 = A$, A_{k+1} is obtained from A_k by an operation described in Lemma 5 and exactly one column of A_ℓ contains 0's. Indeed if, say $a_{11} = 0$, take consecutively the averages of the columns 2 and n, 3 and $n, \ldots, n-1$ and n, 2 and n, 3 and $n, \ldots, (n-1)$ and n. Then we end up with a matrix A_ℓ in which only the first column contains a 0. By Lemma 5 $A_2, \ldots, A_\ell$ are all minimizing matrices. However, this contradicts Lemma 4 since in A_ℓ exactly one column contains 0's. □

References

A.D. Alexandrov, To the theory of mixed volumes of convex bodies (in Russian), Mat. Sbornik 3 (1938) 227-251.

G.P. Egoritchev, A solution of a problem of van der Waerden for permanents (in Russian), Inst. Fiziki im. L.V. Kirenskovo, USSR Akad. Nauk, Siberian Branch, preprint IFSC-13M, Krasnoyarsk, 1980.

D.I. Falikman, A proof of van der Waerden's conjecture on the permanent of a doubly stochastic matrix (in Russian), Mat. Zametki 29 (1981) 931-938.

S. Friedland, A lower bound for the permanent of a doubly stochastic matrix, Ann. of Math. (2) 110 (1979) 167-176.

D. London, Some notes on the van der Waerden conjecture, Linear Alg. Appl. 4 (1971) 155-160.

B.L. van der Waerden, Problem, Jahresberichte der Deutschen Math.-Verein. 25 (1926) 117.

Voidology Theory

W.J.R.Mitchell

Introduction

Recently there have been reports that exciting new work (the so-called Voidology Theory) is being done in the remote and inaccessible Laputan Institute. Thanks to the co-operation and kindness of many members of the Department of Pure Mathematics and Mathematical Statistics, Cambridge University, the author has been able to reconstruct two documents which show the power and scope of this theory.

The first document is a transcript of a discussion between senior members of the Institute. Clearly many keen and acute scholars are contributing to a serious co-operative research effort. The second document is a graduate level examination paper, which suggests that the Institute is destined to become the leading centre of teaching in Voidology Theory.

These documents need little introduction. It suffices to say that the consideration of them has led the author to think again about basic mathematical definitions, and to question the reliability and correctness of various standard works of reference; and that similar reactions are likely in anyone who reads what follows with care.

Alpha: I've just been caught out by one of my pupils.

Beta: So what's new?

Alpha: No, listen, if A, B, C are sets and $A \times B = A \times C$, then prove that $B = C$.

Beta: Well, C and B have the same cardinality, so ...

Gamma: False! If A is empty then $Z \times A = A$ always.

Beta: Oh, I see. There is a lot of scope here. For instance, there is only one equivalence relation on the empty set, because $\phi \times \phi = \phi$ ....

Sigma: And it's certainly the union of its equivalence classes.

Alpha: True, that dosen't take too long to check. Indeed all unions of empty sets are empty ...

Epsilon: ... including the empty one ...

Alpha: ... and so are all products.

Gamma: Are you sure? What if the indexing set is empty?

Beta: Well, without loss of generality we may assume that all sets in the indexed family are empty.

Kappa: We need $\mathrm{Map}(X, \prod_{\phi} A_i) = \prod_{\phi} \mathrm{Map}(X, A_i)$ for any X.

Epsilon: That's a big help — an empty product defined in terms of an empty product!

Kappa: We are given an empty set of maps $f_i : X \to A_i$ $(i \in \phi)$. Indeed by Beta's remark we may assume each A_i is empty, and so that X is empty.

Sigma: (aside) False!

Kappa: Then there is just one way of combining this data into a map of X such that the appropriate diagrams commute.

Lambda: ... but the A_i are empty ...

Gamma: ... but there aren't any of them ...

Sigma: (aside) I think someone's just lost some generality.

Epsilon: Surely it's better to regard the product as a function $f: I \to \cup A_i$ such that for each $i \in I$ we have $f(i) \in A_i$. Then by Beta's remark there is a unique function f (from ϕ to ϕ) ...

Sigma: (aside) ... Called the logarithm ...

Epsilon: ... satisfying the stated condition.

Lambda: I've been thinking about algebra. Sadly most algebraic objects are required to contain at least one element. We can consider empty quasigroups, however.

Delta: Notice that Lagrange's theorem is false for quasigroups, since every quasigroup contains a subobject of order zero. But that's hardly an extensive remark.

Sigma: There is always the zero ring S. That's our nearest approach to an empty ring.

Gamma: It must have all properties.

Delta: Well, either find me a pair of zero-divisors, or tell me what the characteristic of its field of fractions is. But I have a wonderful (empty) theorem about S ...

Kappa: (aside) "... which the margins of this paper are too wide to contain ..

Delta: ... namely —

Theorem The following are equivalent and characterise S among rings, i.e. if R satisfies one of the properties, then $R = S$.

1. $\operatorname{Spec} R = \phi$.
2. X is a unit in $R[X]$.
3. R and $R[X]$ are isomorphic as R-algebras.

Alpha: You should add

4. All R-modules are isomorphic to R.

Sigma: Ah yes, since $m = 1.m = 0.m = 0$.

Tau: Can I have all your cups, please?

Nu: How about some topology? Of course ϕ has a unique topology, with one open set ...

Delta: Known as the usual topology.

Alpha: It is clearly connected.

Beta: Then show me a component! Besides it says in H----- and W----- that if X is connected, then $H_0(X) \cong Z$.

Alpha: Wait ... I see, we should define connectedness as an equivalence relation on the space, with the components being equivalence classes. Thus by an earlier remark, ϕ has zero components, and in general $H_0(X)$ is free of rank equal to the number of path components. As it hasn't got a component, ϕ isn't connected.

Gamma: So $H_*(\phi) = 0$?

Tau: Can I have all your cups please!

Mu: Yes, but $\tilde{H}_{-1}(\phi) \cong Z$, since the suspension of ϕ is clearly the 0-sphere.

Delta: Why?

Mu: Well, since $H_0(\phi) = 0$, the augmentation homomorphism must have a non-zero cokernel.

Lambda: Clearly ϕ is an oriented PL n-manifold for all n — but shouldn't there be a fundamental class?

Sigma: No, because although oriented, we know that ϕ cannot have any orientations ...

Mu: ... because an orientation is an equivalence class of ... etc.

Nu: Notice that ϕ is locally homeomorphic to the Hawaiian ear-ring ...

Mu: What?!

Nu: ... since each point has indeed got a neighbourhood with a homeomorphism of it ...

Sigma: Change the subject!

Lambda: What subject?

Tau: Can I have all cups now please !!

Beta: If E is an empty set of reals, then any real is an upper bound ...

Gamma: ... so $\sup E = -\infty$ ...

Beta: ... and by a similar argument $\inf E = +\infty$.

Alpha: How embarrassing!

Tau: Ladies and gentlemen, will you please stop talking about nothing and give me your tea-cups at once!

Diploma Examination

Paper 0 : Topics in Empty Topology

In the following questions X denotes the empty space with the usual topology.

1. Is X locally connected? Contractible? Locally contractible? Arcwise connected?

2. Enumerate all sheaves on X. What are their sheaf spaces?

3. Show $1_X : X \to X$ is a Hurewicz fibration. What is the fibre?

4. Give as many examples as you can of countable connected metric spaces.

5. Give an example of a map $f : A \to B$ such that A and B are non-homeomorphic compact Hausdorff spaces, and f is a fibration and a cofibration.

6. a) Define the Whitehead groupoid $Wh(G)$ of a groupoid G.

   b) Calculate $\tau(f) \in Wh(\pi(X))$, for any map $f : X \to X$, where $\pi(X)$ is the fundamental groupoid of X. Deduce that X is a finite simple Poincaré complex.

   c) Describe the Spivak stable normal fibre space of X, and write down its classifying map.

7. Calculate the algebraic K-theory of the zero ring, and compare it with the topological K-theory of X.

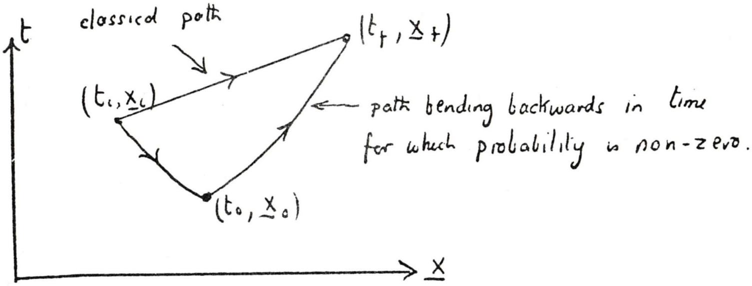
Candidates are advised to write on more than 0 sides of the paper.

Quantum Gravity

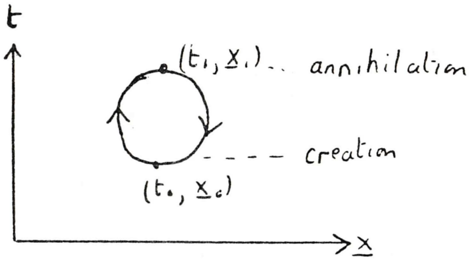
by G.W.Gibbons

The beginning of this century saw two profound changes in our understanding of the basic laws that govern the physical world. Firstly, it was found that the basic laws must be formulated in such a way as to appear the same to observers moving with different velocities. This is Einstein's Principle of Special Relativity. The Principle of Relativity has the consequence that one can no longer regard space and time as essentially different but rather one must combine them into a 4-dimensional spacetime with a geometrical structure which is similar to, but distinct from, ordinary Euclidean geometry. This geometry is called Minkowski geometry. The role of straight lines is played by the paths of light rays and massive particles. In the presence of gravitating bodies these paths are bent and the flat geometry of Minkowski spacetime becomes a curved 4-dimensional psueudo-differential geometry governed by Einstein's equations of General Relativity (G.R.) which tell us by how much spacetime is curved in the presence of massive bodies. Changes in the positions of such bodies can produce changes in their gravitational fields and hence in the geometry of spacetime which propagate in a wave-like manner, at the speed of light. This phenomenon is called Gravitational Radiation.

The second major change in our picture of the world was brought about by the advent of Quantum Mechanics. According to Quantum Mechanics the basic laws of physics are not deterministic as was previously thought and as is the case for example with the Einstein Equations but rather they are statistical in nature. For example, if you know that a particle started at a given position $\underline{x}_i$ in space at a given time t_i and arrived at some different position $\underline{x}_f$ at a later time t_f (i and f stand for initial and final), and that it was influenced only by a known gravitational field, you could (at least if t_f-t_i is small enough) according to classical G.R. say exactly which path it must have followed. According to Quantum Mechanics however this is not possible. You can associate a complex number called an amplitude with any possible continuous path and the probability that the particle followed the path or proportional to the (modulus)^2 of the complex number. The probability will be greatest for the classical path but it will be non-zero for paths which travel backwards and then forwards in time like this



Paths like these may be interpreted as being due to the creation at the earlier time t_0, at position $\underline{x}_0$ of a particle-antiparticle pair. The particle has a path (or world line) with a future directed arrow and runs from $(t_0, \underline{x}_0)$ to $(t_f, \underline{x}_f)$. The antiparticle has a path with a past directed arrow and runs from $(t_i, \underline{x}_i)$ to $(t_0, \underline{x}_0)$. If the particle carried a charge the antiparticle will carry the opposite charge because it moves backwards in time. If the particle has mass m the antiparticle has the same (positive) mass m and to create a pair permanently would cost at least $2mc^2$ in energy. Any experiment to follow a particle's path will take some energy and so the measurement of a single particle's position may itself produce more particles. Even if the measurement does not require more than $2mc^2$ there can, by Heisenberg's Uncertainty Principle, $\delta E \delta t \geq \hbar$ ($\hbar$ is Planck's constant divided by 2π), be fluctuations in the energy for short times. These quantum fluctuations may be thought of as the creation and subsequent annhilation of a pair of particles. They thus correspond to closed-loop world lines like this:



Fluctuations like this can, by the Uncertainty Principle, last for no more than a time $\hbar/2mc^2$. Their existence means that if one makes a relativistically invariant theory which is consistent with quantum mechanics one must allow for the existence of not just one particle but rather of infinitely many particles. Such theories can be constructed. At present the most successful theories which combine special relativity and quantum mechanics are called Quantum Field Theories. For example, Quantum Electrodynamics combines Maxwell's theory and quantum mechanics in an extremely accurate theory describing the interaction between electrons and light which according to one of its creators, Dirac, "explains most of physics and all of chemistry". It is generally referred to, not exclusively for brevity, as Q.E.D.

One of the consequences of Q.E.D. is light has a particle aspect; accelerating electrons emit and absorb light in discrete packets called photons. Indeed, the forces between electrons can also be though of as being due to the emission and absorbtion of photons. These are massless particles and their number is also subject to quantum fluctuations. If one resolves the electromagnetic field into a series of normal modes which behave like simple harmonic oscillators, a mode of angular frequency ω has energy levels $(n+\frac{1}{2})\hbar\omega$, $n=0,1,2...$ The $\frac{1}{2}\hbar\omega$ term is called zero-point energy. It is due to the quantum fluctuations in the number of photons.

Now we turn to Quantum Gravity (Q.G) which tries to combine General Relativity with Quantum Mechanics to give a theory which incorporates the curved geometry of spacetime with the statistical laws of quantum mechanics. So far no completely successful synthesis has been achieved. We know how to construct theories in which the gravitational field remains classical and unaffected by the quantum fluctuations of, for example, electrons and photons which are themselves affected by gravity, but we do not know how to incorporate properly the quantum fluctuations in the gravitational field. The problem is that, just as in Q.E.D. electromagnetic radiation consists of massless particles called photons, so in Quantum Gravity gravitational radiation consists of massless particles called gravitons. While photons carry one unit of angular momentum (in units of $\hbar$) the graviton carries two. Just as in Q.E.D. we must allow for fluctuations in the number of photons, so in Q.G. we must allow for fluctuations in the number of gravitons. As with Q.E.D., we resolve the gravitational field into normal modes, each one of which behaves like a simple harmonic oscillator of angular frequency ω, and obtain energy levels of $(n+\frac{1}{2})\hbar\omega$. The zero-point energy $\frac{1}{2}\hbar\omega$ gives the energy of each fluctuation. However, since the number of oscillators is infinite, the total energy diverges.

Divergences, and other similar ones of this sort, are encountered in Q.E.D., but special techniques have been developed to handle them. However, these techniques do no appear to work in Quantum Gravity. In Q.E.D., if you put the electron mass to zero there are no natural length scales in the problem. Theories of this type are called "renormalizable". Because no natural scale is picked out the divergences occur in certain rather specific expressions in the theory and it is possible to make adjustments to cancel them. However, in Quantum Gravity, even if the masses of all the particles are set to zero, there is a natural length scale which appears - the Planck length $(\hbar G/c^3)^{\frac{1}{2}} \approx 10^{-33}$cm. Note that this expression contains all 3 of the basic constants of physics. One consequence of this natural scale is that the divergences enter into more expressions in the theory and the cancellation procedure mentioned above does not work. In fact, at distance scales comparable with $(\hbar G/c^3)^{\frac{1}{2}}$ quantum fluctuations of the metric become so strong that, at present, reliable calculations are impossible. One possibility is that quantum fluctuations of the metric and hence of the geometry become so large that even the topology of spacetime fluctuates. Such a situation is sometimes referred to as Spacetime-foam in analogy with the surface of the sea which on large scales may appear quite smooth while on small scales may have a foamy structure whipped up by the wind. Perhaps the idea of a smooth manifold which is central to the usual differential geometric description of spacetime breaks down. It may also be that the recent ideas of Fractal and Stochastic geometry, which seem to play an important role in other parts of physics, will come into play.

A different and perhaps complementary approach to the difficulty of quantum fluctuations is provided by the recent discovery that incorporating into Quantum Field Theories a symmetry called Supersymmetry can lead to theories with fewer and in some cases no divergences. According to Quantum Mechanics the amplitude for 2 identical particles to be at positions $\underline{x}_1$ and $\underline{x}_2$, $\Psi(\underline{x}_1,\underline{x}_2)$, must either remain unchanged or simply change its sign when the positions of the two particles are interchanged. In the first case the particles are called bosons and in the second they are called fermions. It is known that particles with interger spin (in units of $\hbar$) are bosons while particles with half integer spin (i.e. $\frac{1}{2}\hbar, \frac{3}{2}\hbar, ..$) must be fermions. Clearly for fermions the amplitude for two particles to be at the same position is zero which is Pauli's Exclusion Principle. The electron has spin $\frac{1}{2}\hbar$ and is therefore a fermion while the photon and graviton with spin $\hbar$ and $2\hbar$ are bosons. Now the zero-point energies of fermion fluctuations of angular frequency ω are $-\frac{\hbar\omega}{2}$. This may be understood intuitively as follows.

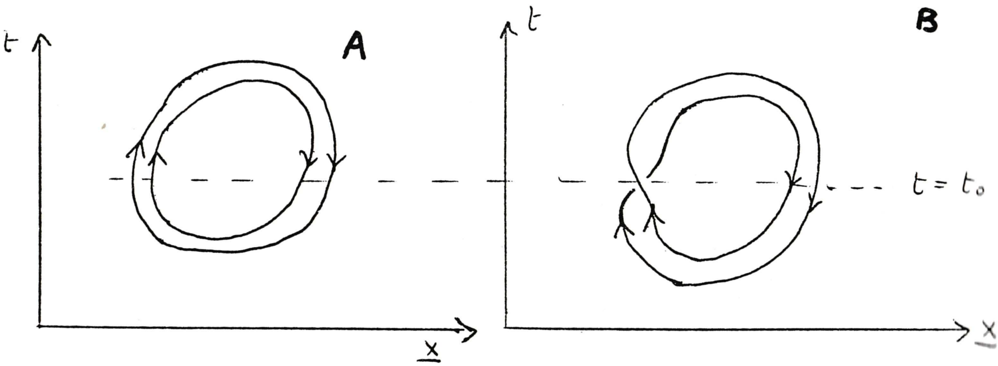


Diagram A differs from B solely by the interchange of two fermions at $t=t_0$. Thus the amplitudes have opposite signs. However, diagram A represents the creation and annhilation of two fermions. But two fermions behave exactly like a boson. Thus we expect, and a detailed calculation confirms that, the zero-point energies of fermions are opposite in sign to those of bosons.

Consider now a theory with equal numbers of fermions and bosons. The zero-point energies will exactly cancel. If the theory is exactly symmetrical between fermions and bosons other divergences will also cancel. In 1976, Ferrara, Freedman, Van Nieuwenhuizen, Deser and Zumino succeeded in constructing a supersymmetric generalization of Einstein's theory called Supergravity. In this theory, one not only has a graviton of spin $2\hbar$ but a fermionic partner called the gravitino, which has spin $\frac{3}{2}\hbar$. The graviton, like the photon, is massless and its own antiparticle so by supersymmetry the gravitino is massless and its own antiparticle. Supergravity is completely symmetrical between the graviton and the gravitino - they should really be thought of as two different aspects of the same entity just as in ordinary quantum mechanics the electron has both wave and particle properties which are complementary aspects of the same entity.

Since 1976, theorists have constructed more elaborate supersymmetric theories of gravity. The most symmetrical and elaborate is called "N=8 extended supergravity". This theory has 1 graviton, 8 gravitini (that is the 8 of "N=8"), 28 spin $\hbar$ particles, 56 spin $\frac{\hbar}{2}$ particles, and 35 spinless particles. Readers of Eureka should have no difficulty in verifying that the number of bosons equals the number of fermions. These 128 particles should all be thought of as different manifestations of the same basic entity since they occur in the theory on a completely symmetrical footing.

Supergravity theories have a beautiful but complicated structure. It is not yet known whether one can handle the quantum fluctuations in these models or whether they can be made to fit with observation. Naively, they do not since nature does not contain equal numbers of fermions and bosons. However, it is known that the symmetries of a field theory can be hidden from immediate view so this need not rule them out. it is too early to say whether supergravity theories will provide satisfactory theories of Quantum Gravity. They are by no means the only approach being tried at present. Nevertheless, the idea of supersymmetry is so appealing that it seems not unlikely Nature makes use of it in some way.

<u>Further Reading</u>

Popular: D.Z. Freedman and P. Van Nieuwenhuizen, Scientific American, <u>238</u>, No. 2 (1978), p. 126;

More technical: P. Van Nieuwenhuizen, "Supergravity", Physics Reports, <u>68</u>, No. 4 (1981)

A Construction for Geometrical Optics

P. Taylor

In Geometrical Optics, as taught at O-level, it is assumed that a lens may be treated as a disc of infinitessimal thickness with a characteristic focal length which may be negative. The points at this distance to the left and right are called the focal points: note that if the focal length is negative the left focal point is on the right and vice versa. With rays of light produced appropriately,

(i) a family of rays is parallel or has a common point iff the deflected family has the same property,

(ii) a ray parallel to the axis is deflected through the focal point on the other side,

(iii) a ray passing through the centre does so undeflected.

These axioms clearly admit an easy classical construction which is no doubt familiar to readers:

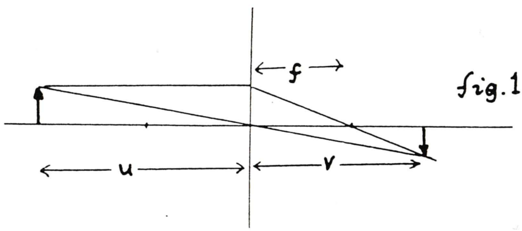


fig. 1

Here $-u$, v, f denote the distances to the right of the object, right focal point and image. By considering two pairs of similar triangles we obtain the usual formula:

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$$

A Tan in Eureka 43 (1983) 31-35 gives another construction with the same relationship between its lengths. However both of these suffer from the disadvantage of requiring parallel or perpendicular lines to be constructed; this article offers a novel construction which, whilst it looks less like the intuitive physical picture than fig. 1, requires far fewer construction lines.

It is literally child'splay to draw a pencil of lines 60° apart. Two applications of the sine theorem for plane triangles will convince the reader that u, v, f are in the required relationship:

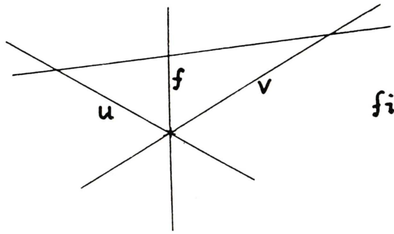


fig.2

Now suppose we have a compound lens such as a microscope or telescope. This is represented by a figure with a number of vertical lines, one for each lens:

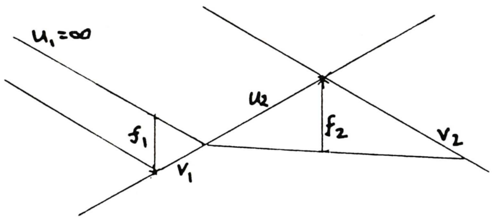


fig.3

An infinite object distance is, as before, represented by a line parallel to the object axis; however this is easier to draw than before since the line will cut both the lens and image lines at distance f.

Fig. 3 demonstrates the simplifications which may be effected. If it is known that the object will always be on the same side of the lens (as is usual), the object line need only be drawn in one direction, and indeed the lens line need only consist of a single point. The distance along the common image/object line between the intersections is equal to that between the corresponding lenses.

What happens if the intermediate image lies beyond the second lens? One might at first think that a (real) image should be treated as another illuminated object, so that one should use (for a positive or converging lens) the nearer side of the lens line, ie (in therms of the classical construction) the far focal point. However the following shows that this is wrong:

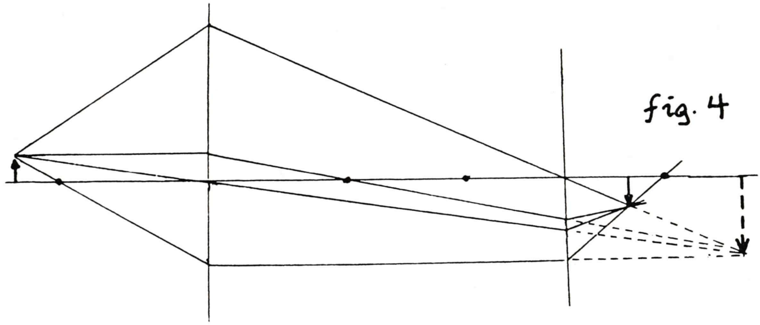


fig. 4

The reason for this is of course that, once a ray has got to the real image it doesnt turn round and go back again! The significant thing is not where the object or image is, but from which side the ray approaches or leaves the lens.

Consequently, irrespective of the positions of the intermediate images, if the object is always on the same side of the first lens, each lens line need only be drawn in one direction. Thus this construction has the further advantage of being unambiguous, since it distinguishes properly between positive and negative lenses.

Once the lens system has been drawn, each calculation of image position requires only the positions of the intermediate images to be marked, which can be done with ruler alone. That this must be possible is clear from the theory since, given the focal lengths and separations, the image position is a rational function of the object position.

Euclid's Algorithm – since Euclid (*)

By P M Cohn

It has been said that every good idea has been thought of before (**). This may not seem true in Mathematics, where so much spectacular progress has been made in our day, but nevertheless there is some truth in it, in the sense that nearly every new idea has its ancestor somewhere among the great old ideas. Here I want to trace one idea down the ages: Euclid's algorithm.

1. Consider the natural numbers:

$$\mathbb{N}: \quad 1,2,3,\ldots$$

The two operations that can be performed on them, addition and multiplication, show very different behaviour. Using $+$ we can get all positive integers by starting from 1: $2 = 1 + 1$, $3 = 1 + 1 + 1$, ...; we say that $\mathbb{N}$ is generated by 1, using $+$. For multiplication we need an infinite generating set, consisting of all the prime numbers: $p_1 = 2$, $p_2 = 3$, $p_3 = 5$, ... Every number a is uniquely expressible in the form

$$a = p_1^{\alpha_1} p_2^{\alpha_2} \ldots, \tag{1}$$

where the α_i are non-negative integers and all but a finite number of them are 0. This is often called the fundamental theorem of arithmetic. Nowadays one usually states it for the ring $\mathbb{Z}$ of all integers (got from $\mathbb{N}$

(*) Based on a talk to the Adams Society on October 19, 1982

(**) Alles gescheite ist schon gedacht worden; man muss nur versuchen es noch einmal zu denken (Goethe: Sprüche in Prosa) by throwing in 0 and the negative integers). Thus we can say that in $\mathbb{Z}$ every non-zero element is either a unit (i.e. invertible) or a product of unfactorable elements, which are unique except for the order in which they occur and for unit factors, a fact expressed more briefly by saying: $\mathbb{Z}$ is a unique factorization domain, or more briefly still, a UFD.

In a UFD it is easy to describe the highest common factor (HCF) and least common multiple (LCM) of two numbers. Instead of (1) we can more briefly write $a = \pi p_i^{\alpha_i}$. If $b = \pi p_i^{\beta_i}$ is another number, then

(2) HCF: $(a,b) = \pi p_i^{\delta_i}$, where $\delta_i = \min\{\alpha_i, \beta_i\}$,

(3) LCM: $[a,b] = \pi p_i^{\mu_i}$, where $\mu_i = \max\{\alpha_i, \beta_i\}$.

If we know one of (2), (3), the other is obtained by the formula: $(a,b)[a,b] = ab$, but this is of no help in finding the HCF and LCM themselves, unless all factorizations (1) are known.

To find the HCF of two numbers a,b without factorizing them, Euclid [2] uses the division with remainder: if $b > 0$, then there exist numbers q,r such that

(4) $$a = bq + r, \qquad 0 \le r < b.$$

We now repeat the process with a,b replaced by b,r and get a series of diminishing remainders, of which the last is the desired HCF. If we number the quotients and remainders as $q_1, q_2, \ldots,$ and $r_1, r_2, \ldots,$ we can write the chain of equations constituting the Euclidean algorithm in matrix form as follows. Write $P(x) = \begin{pmatrix} x & 1 \\ 1 & 0 \end{pmatrix}$, then $P(x)^{-1} = \begin{pmatrix} 0 & 1 \\ 1 & -x \end{pmatrix}$ and now

$$(a,b) = (b,r_1)P(q_1),\ (b,r_1) = (r_1,r_2)P(q_2), \ldots, (r_{n-1},r_n) = (r_n,0)P(q_{n+1}).$$

Hence on writing $C = P(q_{n+1})\ldots P(q_1)$, the matrix C has an inverse, because each P has, and we have

$$(a,b) = (r_n,0)C, \qquad (r_n,0) = (a,b)C^{-1}.$$

This shows that for any d, $d \mid r_n$ if and only if $d \mid a$, $d \mid b$ (where $d \mid a$ means that $a = md$ for some m).

2. It is not difficult to obtain explicit formulae for a,b from the Euclidean algorithm. We define polynomials p_n in n variables recursively by $p_o = 1$, $p_1(t_1) = t_1$, and for $n \geq 2$,

$$p_n(t_1,\ldots,t_n) = p_{n-1}(t_1,\ldots,t_{n-1})t_n + p_{n-2}(t_1,\ldots,t_{n-2}) .$$

This definition shows incidentally that $p_n(1,1,\ldots,1)$ is the nth Fibonacci number, cf. [3]. The first few p's are 1, t_1, $t_1t_2 + 1$, $t_1t_2t_3 + t_3 + t_1$, $t_1t_2t_3t_4 + t_3t_4 + t_1t_4 + t_1t_2 + 1$. In general p_n is formed by the "leapfrog rule": write down $t_1t_2\ldots t_n$ and add to it all products obtained by omitting one or more pairs t_it_{i+1}. Alternatively p_n may be described as the polynomial part of the rational function

$$(t_1 + t_2^{-1})(t_2 + t_3^{-1})\ldots(t_{n-1} + t_n^{-1})t_n .$$

The p_n occur as numerators and denominators of continued fractions [3], and so are called continuant polynomials, and they can also be described as determinants (continuants), but our interest in them here stems from the fact that (by an easy induction),

$$(5) \qquad P(t_1)\ldots P(t_n) = \begin{pmatrix} p(t_1,\ldots,t_n) & p(t_1,\ldots,t_{n-1}) \\ p(t_2,\ldots,t_n) & p(t_2,\ldots,t_{n-1}) \end{pmatrix}$$

Since $P(x)$ has determinant -1, the inverse of (5) follows from the formula

$$(6) \qquad \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = (-1)^n \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

If we now apply (5) to the Euclidean algorithm we find that for any numbers a,b with HCF d, we have

(7) $a = dp(q_{n+1},\dots,q_1)$, $b = dp(q_{n+1},\dots,q_2)$,

and

(8) $au - bv = d$, where $u = (-1)^{n-1}p(q_2,\dots,q_n)$, $v = (-1)^{n-1}p(q_1,\dots,q_n)$.

3. One can ask similar questions about the ring of polynomials $a_o x^n + a_1 x^{n-1} + \dots + a_n$. The answer is now part of most second year courses, but it was not always so easy. Pedro Nuñez [5] (inventor of the Vernier scale) writing in 1567 tries to find the HCF of two polynomials, but without success; he does not get beyond some generalities. Yet only 18 years later Simon Stevin [6] sets it as a problem and says: the answer is obtained by applying the Euclidean algorithm, as for integers. Of course this is not quite true, we need to use the degree of the polynoliam in place of $|a|$. But why did Nuñez find it so hard? My guess is that he used integer coefficients; then the Euclidean algorithm does not apply and in fact the problem is quite difficult. Stevin (who among other things introduced decimal notation) would be more likely to use rational coefficients and so make the problem more tractable.

4. Guided by these examples one can define a Euclidean domain as an integral domain R with an integer-valued function v on R such that

E.1. $v(a) \geq 0$ with equality for $a = 0$,

E.2. $v(ab) \geq v(a)$ for all $a,b \in R$, $b \neq 0$,

E.3. for any $a,b \in R$, if $b \neq 0$ and $v(a) \geq v(b)$, there exists $c \in R$ such that $v(a - bc) < v(a)$.

This is not quite the usual form of the division algorithm, but it is easily seen to be equivalent to the latter. Examples:

(i) $\mathbb{Z}$, $v(a) = |a|$, (ii) $k[x]$, $v(a) = \deg a$, (iii) $\mathbb{Z}[i]$ or more generally the integers in $\mathbb{Q}(\sqrt{d})$, $v(a) = |a|$, for $d = -1, -2, -3, -7, -11$, and $2,3,5,6,7,11,13,17,19,21,29,33,37,41,55,73$ (cf. [7], p. 95).

5. By a more elaborate method (using Gauss's lemma) one can show that the polynomial ring in several variables $k[x_1,\ldots,x_n]$ over any field k of coefficients is a UFD, but for $n > 1$ the Euclidean algorithm seems to have got lost. This may well be connected with the fact that whereas $k[x]$ is a principal ideal domain (which I shall not stop to define as it plays only a tangential role here), the ring $k[x_1,\ldots,x_n]$ for $n > 1$ is not principal. Some efforts to find a Euclidean algorithm or a substitute were made (cf. e.g. [4]), but without much success. Nevertheless there is an analogue applying to polynomial rings which enables us to prove almost everything the usual algorithm does (as far as it is true). It applies to polynomials in any number of variables over any field k , but with the proviso that the variables do not commute. This polynomial ring in non-commuting variables is called the <u>free associative algebra</u> on $x_1,\ldots,x_n$ over k , written $k\langle x_1,\ldots,x_n\rangle$.

To find this general algorithm, let us take two variables x,y for simplicity. Given two elements of $k\langle x,y\rangle$, say $f = x^2y + yx + 1$ and $g = xyx + yxy$, in general neither can be divided by the other, but for a very good reason: f and g have no common right multiple at all, apart from zero. This is a new feature which did not appear in the commutative case, where any two non-zero elements f,g have the common multiple $fg = gf$. If we now restrict attention to pairs with a non-zero common right multiple, the Euclidean algorithm is restored, e.g. if $f = xyz + z + x$, $g = xy + 1$, then $f = gz + x$, $g = x.y + 1$, $x = 1.x$.

What we need is a kind of n-term algorithm which applies whenever an appropriate right multiple condition is satisfied. This is the <u>Weak algorithm.</u> Given $a_1,\ldots,a_m$, if there exist $b_1,\ldots,b_m$ such that

$$v(\Sigma a_i b_i) < \max\{v(a_1 b_1),\ldots,v(a_m b_m)\} ,$$

and if the a's are numbered such that $v(a_1) \leq v(a_2) \leq \ldots \leq v(a_m)$, then there exists j in the range $1 \leq j \leq m$ and $c_1,\ldots,c_{j-1} \in R$ such that

$$v(a_j - \Sigma_1^{j-1} a_i c_i) < v(a_j), \quad v(a_i c_i) \leq v(a_j) \quad (i = 1,\ldots,j-1) .$$

This is satisfied by the free algebra $k\langle X\rangle$ in any set X of non-commuting variables, taking v to be the usual degree. It enables one to prove a unique factorization property. E.g.

$$(9) \qquad xyx + x = (xy + 1)x = x(yx + 1) .$$

Any two complete factorizations of a given element have the same number of factors and the factors on the two sides can be paired off so that corresponding factors differ not just by a unit factor, but are 'similar' (in (9), $xy + 1$ is similar to $yx + 1$, cf. [1], Ch.3). Moreover, for any two elements a,b which have a common non-zero right multiple, we have a highest common left factor d , which can again be written as a linear combination of a and b , as in (8). In fact, exactly the same formulae (7), (8) apply, where the q's are the quotients obtained from the weak algorithm (cf. [1], Ch.2). Here the p_n are defined as before, and (5) still holds. The formula (6) for the inverse matrix cannot now be used, but the inverse exists and is easy to write down, bearing in mind that $P(x)$ has an inverse.

Of course $k\langle X\rangle$ is not a principal ideal domain, but one can show that it is a 'free ideal ring' (fir for short), i.e. a ring in which every left, or right ideal is free, as a module over the ring, with a well-defined rank. One can work out a theory of firs which in many respects parallels the theory of principal ideal domains, and one finds that many important rings are firs. Better still, some have a weak algorithm; this enables one to take over most of the formulae of No. 2, but in spite of their very explicit form, many questions can be asked about these rings which are still unanswered ([1] contains 86 unsolved problems, of which 71 are still open).

References

1. P.M. Cohn, Free rings and their relations, LMS Monographs No. 2 (Academic Press, London 1971)
2. Euclid, Elements, -300
3. G.H. Hardy + E.M. Wright, An introduction to the theory of numbers, (Clarendon Press, Oxford 1938)
4. H. Levi, A characterization of polynomial rings by means of order relations, Amer. J. Math. 65 (1943) 221-234
5. P. Nuñez, Libro de Algebra (Antwerp 1567)
6. S. Stevin, Arithmétique 1585 (vol. II of collected works, 1958)
7. I.N. Stewart + D.O. Tall, Algebraic number theory, (Chapman + Hall, London 1979)

Are You a Pure Mathematician?

C. J. Budd

A very commonly asked question in Cambridge is 'are you a pure or an applied mathematician?' There are those amongst us who would argue that if you are unsure then you can't really be a Pure mathematician at all. But this is unfair, if your abiding passion in life is the application of Lie Group theory to the study of Newt psychology what do you reply? To help you make this important decision I have constructed a series of questions for you to answer. When you have completed the test add up all your scores and then find the Manifold with the nearest Poincaré index. Construct one of its submanifolds (out of Leggo) and then throw it out of the window. Use the time it takes you to do this against the score chart.

Questions

1. 'Life the Universe and Everything'

The 'Hitch Hiker's Guide to the Galaxy' claims that the ultimate answer to life et cetera is the number 42. This is quite a challenge to accepted knowledge. Is your reaction to

(a) Dismiss it immediately on the grounds that the answer should surely be a prime number?

(b) Agree more or less, you think that us 'Green Furry things from Alpha Centauri' ought to stick together.

(c) Disagree. The answer is already known - it is one of the Sub-groups of the Monster.

(d) Agree to within about 10 orders of magnitude (note that infinity=4 *)

---

* See Eureka 42

2. 'Applied Mathematicians'

You are approached in Kings Parade by someone claiming to be an Applied Mathematician. Do you

(a) Disbelieve him/her

(b) Attempt to deprogram(me) him/her

(c) Run screaming into Fitzbillies

(d) Shake their hand and say how good it is that there are at least two of you in Cambridge.

3. 'Politics'

Sooner or later someone in Cambridge will ask you what your politics are. As a 'Pure mathematician' you reply:

(a) Now let's think - today is Tuesday...

(b) I'm not sure - isn't it one of the Applied Part III courses?

(c) It's a Branch of Catastrophe Theory

(d) $\exists f : \forall X \quad f : X \longrightarrow \emptyset$

4. 'Exams'

It's a fact that even the best of us have to face an exam. As you are a Mathematician you have spent your whole year finding a new completion of the rationals and have done no relevant work at all. You are thus faced with a question paper on which no question is possible. Do you:

(a) Realise that you have accidently walked into an Exam on 'Ancient Arabic'

(b) Decide to change to 'Ancient Arabic'

(c) Shut your eyes and think of England

(d) Shut your eyes and think of Ancient Arabia

(e) Copy off the pretty girl next door

5. '<u>Fashion</u>'

Given that even 'Pure Mathematicians' have to wear clothes sometimes - at least when they are attending Supervisions. Do you think that the highest point in fashion is:

(a) An Archimedeans' scarf

(b) A twelve foot Archimedeans scarf

(c) A Klein Bottle *with handles

(d) An anorak, College scarf and old school briefcase.

6. '<u>Supervisions</u>'

A common scenario: You have done no work at all for your supervision through no fault of your own i.e. the overwhelming pressures of your social life, Archimedeans' committee meetings, drinking, et cetera; you survive the supervision by:

(a) Not turning up or turning up precisely one hour late

(b) Turning up, but wear a Viking costume to distract the Supervisor

(c) Grunting

(d) Boning up on some very obscure topic and asking questions like 'Will the course involve Non Euclidean Compactifications of a Semi Simple Banach Lattice?'

(e) Telling the supervisor about your interest in Ancient Arabic.

7. '<u>Computing Software</u>'

'A man needs a computer like a fish needs an exercise wheel'. Knowing this your favourite computing language is:

(a) Basic

(b) BASIC

(c) Anything provided that it doesn't have real arithmetic or a GOTO statement

(d) A trivial application of the results of Category Theory.

---

* For the daring only

8. '<u>Computing Hardware</u>'

If you are forced to do any computing do you do it

(a) On the latest model from Sinclair

(b) At 3 a.m.

(c) In your head

(d) As a trivial part of your PhD. Thesis on Category Theory.

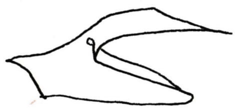
9. '<u>Sex</u>'

A difficult one this. You are often required to fill in application forms where one of the questions asks you to tick a box to say which sex you are. Do you

(a) Tick both to be absolutely certain

(b) Take a more sophisticated approach and tick each box 50% of the time

(c) Tick neither box and write 'Pure Mathematician'

(d) Draw  or similar in the boxes.

10. '<u>The role of Mathematics</u>'

Do you think that Mathematics is:

(a) A golden thread in life's rich tapestry

(b) The whole tapestry

And finally, would you agree that

(a) Classical Analysts do it with small epsilons

(b) Algebraists do it in Groups

(c) Logicians do it without choice

(d) Numerical Analysts do it by crunching

(e) Boundary values do it on the side

Scoring ...

So what are you:

1. Forgot to open window.

Very good, a high sense of priorities demonstrated here. You would be likely to do well in most Part III pure courses.

WARNING You may be asked whether you go to Trinity.

2. Under 30 seconds.

Also good. You are what is known in the trade as a Gerbil. Much of your time is likely to be spent discussing Cohomology Groups with your Girl/Boyfriend and spending dirty weekends classifying the diffeomorphic topologies of the real line.

3. $30+\varepsilon$ seconds to 5 minutes.

Not bad. This group comes under the heading of the 'Sainsbury'. For you 'working in L^2' really means something (Man). You don't, however, think that Functional Analysis is as much fun as riding a motorbike.

4. 5 minutes to 2 hours.

A lot of people fall into this group, perhaps it is just part of growing up. You may have been seen in the DAMTP common room and are probably a member of the 'Captain Kirk fan club'. Weak topologies aren't your overwhelming passion in life but on some days you agree that they might be quite important.

5. 2 hours to 5 years.

We call these people Hobbits. You are not too keen on Maths but still prefer Eureka to reading the Beano.

6. 5 years $<$ time $<$ infinity.

The Toad. Pretty bad but you may well be a possible candidate for the Archimedeans Committee

7. Time = infinity.

Congratulations, you are a normal human being! Given this assumption, why are you reading Eureka?

The Decline and Fall of Euclid. by D. Haskell

"yet can no humayne science saie thus, but I (geometrie) onely, that there is no spark of untruthe in me: but all my doctrine and workes are without any blemishe of errour that mans reason can discerne."

Preface to Pathway to Knowledge

Robert Recorde 1551

For over two thousand years, geometry has been studied for many reasons: particularly for its use in navigation, surveying, and construction, and as an academic discipline to develop logical reasoning abilities. It has rarely been taught purely for its own sake; the other reasons have dominated the thinking of educators in this country since the Renaissance. In particular, the public schools have focussed on the development of reasoning skills as both a reason for teaching geometry and a rationalisation of their method of teaching. Specifically, in the centuries following the Middle Ages, the teaching of geometry meant the teaching of Euclid.

Euclid wrote his Elements of Geometry in approximately 300 BC, bringing together much work that had already been done and fitting it into a logical framework. He aimed to produce a model of mathematical thinking, and certainly set a high standard for other mathematicians to emulate. However, when it comes to teaching geometry, this highly theoretical approach is clearly not appropriate. The Elements was never intended for schoolchildren: it does not present concepts in a way that makes them easily understandable to the beginner. Also, it is not directly applicable to any practical use of geometry.

Nevertheless, Euclid was considered the geometer for gentlemen, and Elements continued to be used as a school textbook until the beginning of this century. This was partly due to its traditional place in the education of the elite, but even more so to the gradual establishment of a national examination system. Since every examinee would have studied Euclid, the questions were predictable which made it easy for students to prepare for examinations, and the possible answers were extremely limited which made marking relatively objective. It may thus be said that the prominence of Euclid's Elements in the teaching of geometry had less to do with its intrinsic merits than with other factors. So it is, perhaps, surprising that the process of replacing it took so long.

Euclid started with he smallest possible unit, the point, which he defined as "position without magnitude". Increasing in complexity, a line, he says, is "breadthless length" and the two are related by "the extremities of a line are points". It was to this that all later writers reacted.

By the Edwardian Statutes of 1549, all freshmen at Cambridge were required to be taught mathematics. One of the recommended books was Euclid's Elements. The learning of mathematics had been encouraged by some individuals, but had never before been compulsory. This was changed again when new statutes were issued by Elizabeth I in 1570. In the meantime, the first translation of Euclid was completed by Henry Billingsley in 1570, and the first geometry textbook in English, Pathway to Knowledge by Robert Recorde, was published in 1551.

Recorde also starts his definitions with the smallest possible unit and works up, but in such a way that everything is defined in terms of what came before. As he was writing for the practical user of geometry as well as the scholar, he gives both a theoretical definition and explanatory notes to clarify the concept. Thus he says:

"a poynt or a prycke is named of Geometricians that small and unsensible shape, which has in it no partes, that is to say: nother length breadth nor depth. But as this exactness of definition is more meeter for onlye Theorike speculacion, then for practice and outward worke (consideringe that myne intente is to applye all these whole principles to woorke) I think meeter for this purpose, to call a poynt or prycke, that small point of penne, pencyle, or other instrumente, which is not moved, nor drawn from his fyrst touche, and therfore has no notable length nor breadth"

Of the line he says:

"Nowe of a great number of these prickes is made a Lyne .... and this lyne is called of Geometricians, Lengthe withoute breadth."

The straight line he defines in terms of one of its properties:

"A Straight lyne, is, the shortest that maye be drawenne betweene two prickes".

Recorde's imposition of practicality on Euclid's purely philosophical definitions may have been considered heretical, as much in social as in intellectual terms.

Despite being considered a more vocational subject, mathematics began to gain a foothold in the universities. In 1619, a professorship in geometry was established at Oxford by Sir Henry Savile and in 1662, the Lucasian chair in mathematics was established at Cambridge. By the end of the seventeenth century, mathematics was firmly in place at the universities, but largely ignored by the upper-class schools.

The period after the reformation was one of intolerance nationally, and stagnation in the universities. The Act of Uniformity in 1662 led to over one hundred and fifty school teachers and professors from the universities losing their appointments for religious reasons. Many of them carried on teaching and established "dissenting academies", often of no fixed address. Initially, their curricula were still firmly grounded in the classics, but as the academies became more settled after the passing of the Toleration Act in 1689, the second generation dissenters were less affected by the Oxbridge traditions. In the first half of the eighteenth century, mathematics, and especially Euclid, was taught to strengthen the students' abilities for logical thinking. By this time there were several different translations of Euclid into English, e.g. R Simson 1723. The dissenting academies developed into theological colleges which discarded mathematics as being conducive to scepticism, but a way had opened for it to take up a more important position in secular academies and the universities. By the beginning of the nineteenth century, all Cambridge graduates would have studied, and been examined in, mathematics. Indeed, W. W. R. Bell in his History of Mathematics at Cambridge (1889) quotes John Jebb: "as the highest academic distinctions are invariably given to the best proficients in mathematics and natural philosophy, a very superficial knowledge in morality and metaphysics will suffice."

At the same time, applied mathematics was developing. Charles Hutton's A Course of Mathematics (1798) was written specifically for use at the Royal Military Academy. As the author says in the preface, it retains:

"only such parts and branches, as have a direct tendency and application to some useful purpose in life, especially in the military profession ... the author hopes he will not be too severely criticised if, through a desire of rendering this branch more easy and simple, he has in some instances deviated a little from the tedious and rigid strictness of Euclid."

His definitions, however, are pure Euclid.

Mathematics was now an important subject in grammar schools, but it was not until the 1830's that the first mathematics master was appointed at a major public school. The pressure of open scholarship examinations to the universities, and to the Royal Military Academies, led to mathematics finally being firmly established in public schools in the 1860's. Euclid, of course, was the text used for geometry, despite its unsuitability for schools, and throughout the first half of the nineteenth century, numerous new editions were published. Some of these were (nearly) direct translations e.g. Bonnycastle 1803, whilst others "endeavoured carefully to retain the spirit of the original" and at the same time "enlarging the basis and disposing the accumulated materials into a regular and more compact system" as John Leslie says in the preface to his Elements of Geometry (1820).

Bonnycastle's definitions are the same as Euclid's, but Leslie's attempt to provide additional explanations is somewhat baffling:

"Body divested of all its essential characters, presents the idea of mere surface; a surface, considered apart from its peculiar qualities, exhibits only linear boundaries: and a line, omiting its continuity, leaves nothing in the imagination but the points which form its extremities."

He does do rather better when discussing the straight line, cleverly disguising the fact that he has avoided giving a direct definition:

"The uniform tracing of a line which through its whole extent is stretched in the same direction gives the idea of a straight line. No more than one straight line can therefore join two points; and if a straight line be conceived to turn like an axis about both extremities, none of its intermediate points will change their position."

Thus, by the first half of the nineteenth century, Leslie could present this adaptation of Euclid without the apologies Hutton had felt obliged to make thirty years earlier. Nevertheless, Hutton's Course of Mathematics was still tainted with its association of practical use.

The next step for the displacement of Euclid from the canon occurred in 1871 when the Association for the Improvement of Geometrical Teaching was established by a group of teachers.* They felt that two things were needed: text books with a different approach and a change of attitude on the part of public examiners. The former were readily available: James Wilson's Elementary Geometry (1868) was one of the first to be written to provide a scholarly alternative to Euclid.

Wilson's book was generally not well received. This was due in no small part to the attitude of the university examiners. The examination system was essentially controlled by Cambridge, which refused to change. So schools would teach geometry

---

* Now the Mathematical Association

for a few years, then revert to drilling students in Euclid as examinations approached. It was not until 1903 that Cambridge finally capitulated with the acceptance of the Senate that "Any proof of a proposition shall be accepted which appears to the examiners to form part of a systematic treatment of the subject."

That same summer, C. W. Godfrey and A. W. Siddons wrote their famous text book Elementary Geometry which remained in print until 1973. They were now free to leave the discussion of points and lines to the end of the practical section of their book, and the formal definitions appear only in an appendix. You are encouraged to deduce for yourself that the boundary of two different solids belongs to neither and therefore has no thickness, and so on. Having described a point by working down from a solid, they then turn around and present a line as generated by a moving point, and a surface as generated by a moving line. However, they do not attempt to justify the equivalence of these definitions. A moving point is, of course, the same as Recorde's line made up of a great number of pryckes. W. J. Dobbs' A School Course of Geometry (1913) treats the subject similarly, though his discussion is less object-oriented than that of Godfrey and Siddons. Interestingly, Dobbs is the only one who distinguishes between a line and a line-segment. Both of them say outright that the idea of a straight line is intuitively clear, but very difficult to define.

It can be seen that both Godfrey and Siddons, and Dobbs tried to present the material in a way that encouraged school children to think for themselves. Euclid and Recorde had equally tried to stimulate the minds of their readers, but had a different, older, reader in mind. It was the misapplication of Euclid that was at fault, not the Elements itself. This misapplication can be blamed on the inertia of the examination system with its requirement of uniformity. Examinations should not, however, provide the motivation for either teaching or studying geometry. Of course, there will always have been individual teachers who were genuine educators and could overcome the imperfections of their material, just as there have always been individuals who transcended the limitations of their education. Nevertheless, for the large majority of students, geometry when taught through Euclid was a boring subject that had to be memorized. It has now been freed from the constraints of Euclid's _Elements_, but the examination system still reigns. It is no longer the _Elements_, but other books which are memorized by bored students, and will continue to be as long as there are examinations for them to sit. Thinking, after all, is not only so much harder, but is so seldom rewarded.

References

A. G. Howson _A History of Mathematics Education in England_ (1982)

R. Simson ed. Euclid's _Elements of Geometry_ (1810)

Text books consulted:

Robert Recorde _Pathway to Knowledge_ (1551)

Charles Hutton _A Course of Mathematics_ (1798)

John Bonnycastle ed. Euclid's _Elements of Geometry_ (1803)

John Leslie _Elements of Geometry_ (1820)

James Wilson _Elementary Geometry_ (1868)

C. W. Godfrey and A. W. Siddons _Elementary Geometry, Practical and Theoretical_ (1903)

W. J. Dobbs _A School Course of Geometry_ (1913)

Sampled Functions.

C. Budd

The increasing use of discrete techniques to solve mathematical and engineering problems - for example digital computation and digital signal processing - usually requires that the data we work with consists of samples taken from an input function. The questions thus arises - given a sequence of sample points what can we say about the original function?

Clearly in the general case we can say very little, a sine wave if sampled at the wrong points will yield the not very revealing sequence: 0 , 0 , 0 , ... However a lot more may be deduced from sampling at three times this frequency to give the sequence 1 , $\frac{1}{2}$, $-\frac{1}{2}$, -1 .... .

Traditional techniques for studying a sequence a_0 , a_1 , ... rely on the notions of forward, backward and central difference which should be familiar to anyone who has studied numerical analysis.

Defining $\Delta a_0 = a_1 - a_0$, $\nabla a_1 = a_1 - a_0$ $\delta a_{\frac{1}{2}} = a_1 - a_0$ their effect on a function y which has a_i as its values at ih, are $y = y(x + h) - y(x)$ et cetera.

On the assumption that y is analytic (i.e. it has a Taylor series representation) we obtain the following relations:

$$\Delta y = (e^{hD} - 1)y \ , \quad \nabla y = (1-e^{-hD})y \ , \quad \delta y = 2\sinh(\tfrac{1}{2}hD)y \qquad (Dy = dy/dx).$$

So far so good, the above being formal expressions of the Taylor Series for y. These ideas may be used to interpolate y between the values given in particular:

$$y(x+\theta h) = (1 +\Delta)^{\theta} \ y = y(x) + \theta\Delta y + \ .... \qquad (1)$$

(Gregory Newton formula)

We may also make the following calculation to obtain the value of the derivative of $y(x)$. Given $\Delta y = (e^{hD} - 1)y$

$e^{hD} = (1 + \Delta)$ thus $hD = \log(1 + \Delta)$

i.e. $D = 1/h\,(\Delta - \Delta^2/2 + \Delta^3/3 - \ldots\ldots)$ (2)

(Note that the first term is $(y(x+h) - y(x))/h$ which looks as though it's on the right lines.)

Ignoring the fact that (2) is a difficult formula to use numerically - frequently involving the subtraction of large values and thus prone to rounding error, how valid are they as descriptions of the behaviour of the function? Given the function $\sin \pi x$, the tabulated values for $h = 1$ are 0 , 0 , 0 , ... and our two formulae yield $y(\theta) = 0$ all θ and $dy/dx|_0 = 0$ - both invalid results. On first inspection it would appear that there is no real reason why (1) and (2) should tell us very much at all - both interpolation and differentiation being results local to the point of evaluation. If we allowed an unrestricted range of functions as 'data', the formulae would be useless - it is possible to construct an infinitely differentiable function, i.e.  which has all possible derivatives vanishing at the data points and yet having arbitrary values between them! Hence the amount of information we can gain from the sequence depends heavily on the spaces in which we allow our functions to lie.

One of the weaknesses of (1) and (2) is that we have no well defined notion of what we mean by $(1 + \Delta)^\theta$ and $\log(1 + \Delta)$.

This is provided by the operational calculus where

$$\begin{aligned} (1+\Delta)^\theta &= \int_\gamma (1+w)\ (wI - \Delta)^{-1}\, dw/2\pi i \\ \log(1+\Delta) &= \int_\gamma \log(1+w)(wI - \Delta)^{-1} dw/2\pi i \end{aligned} \qquad (3)$$

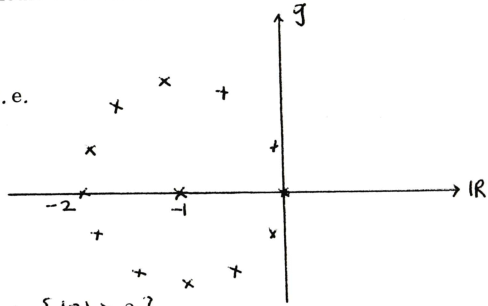
γ is any contour lying in $\{ |w| < e \}$ $e = \sup\{ |z| : (zI - \Delta)^{-1}$ does not exist $\}$ and $(1 + z)^\theta$ and $\log(1 + z)$ are analytic in the interior of γ.

Suppose we restrict our function space to $P = \{$ 'all periodic fns.' on $[0, 1]$ $\}$

and $\Delta y = y(x+1/n) - y(x)$. A little calculation shows that

$\{w : (wI - \Delta)^{-1}$ does not exist $\}$

is $\{w : (1+w)^n = 1\}$ i.e.



Hence γ is any contour in the set $\{|z| > 2\}$

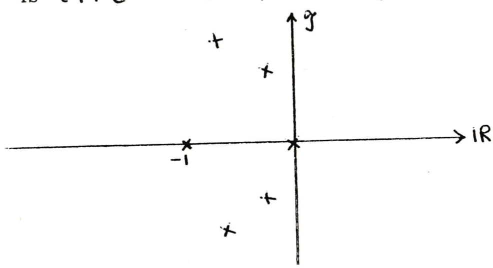
But in (3) both $\log(1+z)$ and $(1+z)^{\theta}$ have branch points at -1 which γ must circle. Thus (3) breaks down and hence (1) and (2) fail even on this very restricted set of functions P as our $\sin \pi x$ example shows. These formulae can however work in some cases. If instead of P we use the set

$P_m = \{$ Polynomials of degree m in $e^{\pm i x}\}$ then

$\{w : (\Delta - wI)^{-1}\}$ does not exist is $\{-1 + e^{I 2\pi i r/n}$ $r = 0, 1 \ldots, m\}$

e.g. for m = 2 this is

for some n



For (3) to succeed we want to enclose this set by a contour which does not encircle -1, this will be possible if each element of the set has modulus less than one. A little calculation shows that this occurs if $2\pi m/n < \pi/3$ i.e. $\underline{m/n < 1/6}$ (4)

This result can be seen more clearly is we examine the sequence ar directly. Suppose that $U^o_r = a_r$ and $U^n_r = \Delta^n a_r$ then $U^{n+1}_r = U^n_{r+1} - U^n_r$. Assuming that $U^n_r = c_n d^r$ obtain: $c_n = (d-1)^n$ i.e. $\Delta^n a_r = (d-1)^n d^r$ and that $\Delta^n a_0 = (d-1)^n$.

Thus (2) gives $dy/dx \Big|_0 \text{'='} \sum_{n=1}^{\infty} (d-1)^n (-)^{n+1}/n = \log(1+(d-1)) = \log(d) = 2\pi ih$.

This is perfectly correct if $d = e^{2\pi ih}$ and the calculation is valid provided that

$|d-1| < 1$ i.e. $|e^{2\pi imh} - 1| < 1$ or $\underline{mh < 1/6}$ the result obtained earlier.

A similar calculation with (1) gives $y(\theta h) \doteq (1+(d-1))^{\theta} = d^{\theta} = e^{\theta 2\pi imh}$

which is again correct.

If we take the limiting case mh=1/6 letting m=1 and h=1/6 the samples of the sequence for $\cos 2\pi x$ are:

Δ^0	1	$\frac{1}{2}$	$-\frac{1}{2}$	-1	$-\frac{1}{2}$	$\frac{1}{2}$	1	$\frac{1}{2}$ ......
Δ^1	$-\frac{1}{2}$	-1	$-\frac{1}{2}$	$\frac{1}{2}$	1	$\frac{1}{2}$	$-\frac{1}{2}$	
Δ^2	$-\frac{1}{2}$	$\frac{1}{2}$	1	$\frac{1}{2}$	$-\frac{1}{2}$	-1		
Δ^3	1	$\frac{1}{2}$	$-\frac{1}{2}$	-1	$-\frac{1}{2}$			

$\cos \pi x/3$

A sequence with a pleasing degree of regularity.

In engineering applications a sampled function is frequently derived from a continuous signal by observing it at a set of equally spaced time intervals. The functions corresponding to our trigonometric polynomials are 'band limited' i.e. their Fourier Transform is identically zero except at a finite set of frequencies. These signals are 'low pass' if their spectrum contains zero and are otherwise bandpass. Although such signals cannot exist in actuality, in many applications there is a range of frequencies outside of which the spectrum is small enough to be assumed negligible. For these signals we have the Sampling Theorem.

A low pass, band limited function with frequency components outside -m, m Hertz can be described uniquely for all time by a set of sample values at time instants separated by 1/(2m) seconds or less. i.e. for $\underline{mh \leq \frac{1}{2}}$

The frequency and other information can then be extracted using some form of digital filter.

This upper bound for h is much better than our previous one and clearly represents the problems of using formulae (1) and (2).

For example, let m = 1 and h = 1/2, samples of cos $2\pi x$ are:

$$
\begin{array}{lccccccccccccccc}
 & 1 & & 0 & & -1 & & 0 & & 1 & & 0 & & -1 & & 0 \ldots\ldots \\
\Delta^1 & & -1 & & -1 & & 1 & & 1 & & -1 & & -1 & & 1 & \\
\Delta^2 & & & 0 & & 2 & & 0 & & -2 & & 0 & & 2 & & \\
\Delta^3 & & & & 2 & & -2 & & -2 & & 2 & & 2 & & & \\
\vdots & & & & & 4 & & 0 & & 4 & & 0 & & & &
\end{array}
$$

with the leading term growing as $2^{n/2}$ offering no hope of using (1) or (2) in a meaningful fashion.

Of course, had we known that our sequence was a trigonometric polynomial, used the above sequence to extract the Fourier cfts., and then differentiated we would have obtained the correct answer! The differentiation routine (1) was not told this beforehand and hence gave a meaningless result.

I suppose the moral to all of this is - be very careful when using any result based on limited information of a function that the methods which you use use the information wisely!

The Uniqueness of the Quaternions

P. Taylor

Abstract

Assuming the so-called "Fundamental Theorem of Algebra", that **C** is algebraically closed, it is possible to show that **R**, **C** and **H** are the only finite-dimensional real associative division algebras, using only some basic linear algebra.

Introduction

Most undergraduates will have heard, even if they dont know how to state or prove the result precisely, that there's nothing else associative beyond **R**, **C** and **H**. Some will also have heard of the Cayley Numbers (**O**), a non-associative eight-dimensional algebra with the following multiplication (see, e.g., [13]):

1	i	j	k	ij	jk	ki	$i(jk)$
i	-1	ij	$-ki$	$-j$	$i(jk)$	k	$-jk$
j	$-ij$	-1	jk	i	$-k$	$i(jk)$	$-ki$
k	ki	$-jk$	-1	$i(jk)$	j	$-i$	$-ij$
ij	j	$-i$	$-i(jk)$	-1	$-ki$	jk	k
jk	$-i(jk)$	k	$-j$	ki	-1	$-ij$	i
ki	$-k$	$-i(jk)$	i	$-jk$	ij	-1	j
$i(jk)$	jk	ki	ij	$-k$	$-i$	$-j$	-1

However I suspect most readers will assume (particularly if they've attended one of Prof. Adams' college society talks such as [2, 3, 4]), that the result is difficult to prove: this I propose to refute.

An algebra is a real vector space A together with a map $* : A \times A \longrightarrow A$ which is linear in each variable separately; it's associative if $a * (b * c) = (a * b) * c$ for all $a, b, c \in A$ and a division algebra if $a * b = 0 \Rightarrow a = 0$ or $b = 0$. If associativity is weakened, requiring equality only when a, b, c are linearly dependent (eg if two of them are equal), it can be shown that **R**, **C**, **H** and **O** are the only finite-dimensional alternative real division algebras.

Frank Adams' contribution [1] to the subject, requiring a substantial amount of Algebraic Topology, was to show that if we drop even alternativity, whilst we do get more division algebras, they all have dimension 1, 2, 4 or 8.

In this paper associativity will be assumed, as will be the fact that **C**=**R**(i) is algebraically closed, so every nontrivial polynomial equation with complex coefficients in one variable has a complex root. This was first (genuinely) proved by Gauss in 1799 [7], using what we now know as Topology, and by pure Algebra in 1815 [8]. The latter proof subsequently evolved into one using Galois and Sylow Theory, which were of course not available to Gauss, so it contains a rather hairy calculation! There will be a discussion of it in *Eureka* 45.

Gauss' second proof used only the following three elementary facts about **R**:

(i) -1 cannot be expressed as a sum of squares in **R**.

(ii) If $c \in \mathbf{R} \setminus \{0\}$ then exactly one of $c, -c$ has a square root.

(iii) Every polynomial equation of odd degree over **R** has a root in **R**.

A field satisfying these properties is called real closed, and we shall assume nothing more of **R**. Hence there is a corresponding result for the real algebraic numbers.

Some Lemmas

There are three important properties of fields which we shall use.

(i) A polynomial equation of degree n can have at most n distinct roots.

(ii) Any homomorphism between fields (i.e. a map preserving 0, 1, +, −, * and /) is injective.

(iii) A field can be uniquely extended by a root of an irreducible polynomial.

The first two are obvious but perhaps the third is not as elementary as my abstract would have you believe.

The reader is invited to work out for itself what the irreducible polynomial x^2+1 has to do with the construction of **C** from **R** and how this is to be generalised to any irreducible polynomial over any field. The precise result which we shall use is that if f is an irreducible polynomial over a field K then there is a field $L = K/(f) = K(\alpha)$ containing (a field isomorphic to) K as a subfield and also some element $\alpha \in L$ satisfying $f(\alpha)=0$, and if $M \supset K$ is another field with $\beta \in M$ s.t. $f(\beta)=0$ then there is a unique field homomorphism $\phi: L \longrightarrow M$ fixing K (i.e. $\phi(k)=k$ for all $k \in K$) and taking α to β.

If H is an associative finite-dimensional (real) division algebra and $a \in H \setminus \{0\}$, then $a \mapsto xa$ is a linear map from a finite-dimensional vector space to itself with trivial kernel and hence an automorphism. Hence a itself is the image of some unique element which we may call 1, and this is the image of some a^{-1}. It's easy to check that these have the required properties. We shall identify **R** with the subspace spanned by $1 \in H$.

If $K \subset H$ is a field (i.e. a commutative division subalgebra) and $a \in H$ commutes with everything in K (we say a is in the <u>centraliser</u> of K), then (K, a) together generate a subfield since sums, products and inverses are given to exist within H and any two expressions obtained in this way from $K \cup \{a\}$ must commute by hypothesis (and an obvious induction if you're fussy.) In particular, if $a \in H \setminus \mathbf{R}$, there is a subfield $C = \langle \mathbf{R}, a \rangle \subset H$ of dimension (as a real vector space) at least 2.

Now H is a finite-dimensional vector space, so the powers $1, a, a^2, \ldots$ of some $a \in H \setminus \mathbf{R}$ cannot be linearly independent; in other words, a satisfies a nontrivial polynomial equation $x^n + r_{n-1}x^{n-1} + \ldots + r_1 x + r_0 = 0$ with real coefficients. wlog, n is minimal, whence the polynomial is <u>irreducible</u> (cannot be split into factors with real coefficients). Such an equation has a root $\alpha \in \mathbf{C}$. The subfield of $\mathbf{C}$ generated by $(\mathbf{R}, \alpha)$ must be a real vector space of dimension at least two, but then of course it must be the whole of $\mathbf{C}$ since that only has dimension two, so $\mathbf{R}(\alpha) \cong \mathbf{C}$. Then by the theorem of the previous section, there is a field homomorphism (which is always injective) $\mathbf{C} \longrightarrow \mathbf{R}(a) \subset H$. Hence we may identify $\mathbf{C}$ with a subfield of H.

The main proof

Suppose H is a finite-dimensional real associative division algebra which is not isomorphic to either $\mathbf{R}$ or $\mathbf{C}$. Clearly, then, its dimension is at least 2, so let $a \in H \setminus \mathbf{R}$. By the linearity of multiplication in H, a must commute with $\mathbf{R}$, so it generates a subfield (isomorphic to) $\mathbf{C}$ by the previos section. Having shown that $\mathbf{C} \subset H$, we may pick $i \in \mathbf{C}$ with $i^2 = -1$.

Now I claim that $\mathbf{C}$ is its own centraliser in H. For if $b \in H \setminus \mathbf{C}$ were to commute with the whole of $\mathbf{C}$ they would together generate a larger subfield inside H, in which b would satisfy a nontrivial irreducible polynomial equation over $\mathbf{C}$, which is impossible since $\mathbf{C}$ is algebraically closed.

Now consider the following subsets of H, which may easily be seen to be vector subspaces since $\mathbf{R}$ is central:

$$U = \{u \in H:\ ui = iu\} \qquad V = \{v \in H:\ vi = -iv\}$$

Clearly $U \cap V = 0$ and $U = \mathbf{C}$ is the centraliser of $\mathbf{C}$ in H. I claim $H = U \oplus V$. For given $h \in H$ let $h_\pm = \frac{1}{2}(h \mp ihi)$; then $h \mapsto h_+$ and $h \mapsto h_-$ are linear maps with $h = h_+ + h_-$ and, as may easily be verified, $h_+ \in U$, $h_- \in V$. Thus dim U = 2 and dim $V \geqslant 1$.

Now let $b \in V \setminus \{0\}$. Then $u \mapsto ub$, $v \mapsto vb^{-1}$ give vector-space isomorphisms between U and V, so dim V = 2 and dim H = 4. Also $ib^2 = -bib = b^2 i$ so $b^2 \in U = \mathbf{C}$. But clearly b centralises $\langle \mathbf{R}, b^2 \rangle$, so $\langle \mathbf{R}, b^2 \rangle \neq \mathbf{C}$ whence $b^2 \in \mathbf{R}$.

Clearly $b^2 \neq 0$, but can $b^2 > 0$? No, because then b would be a root in $\langle \mathbf{R}, b \rangle \cong \mathbf{C}$ (although this is not the same subspace as $\langle \mathbf{R}, a \rangle$) of $x^2 - b^2 = 0$, and b would have to be one of $\pm\sqrt{(b^2)} \in \mathbf{R}$ contrary to the hypothesis that $b \in V \setminus \{0\}$ (and hence $b \notin \mathbf{R}$. So we may put $j = b/\sqrt{(-b^2)} \in V$ and $k = ij$, which are easily seen to be linearly independent in V.

Then $\{1, i, j, k\}$ is a basis of $H = U \oplus V$ over $\mathbf{R}$ and the usual multiplication table may easilybe verified, so $H \cong \mathbf{H}$ as required.

The result was first proved by Frobenius in 1878, but I cant find the reference.

Bibliography

[1] Adams, J.F., On the non-existence of elements of Hopf-invariant one *Bull. Am. Math. Soc.* 64 **(1958) 279-282.** **and** *Ann. Math.* **72** (1960) 20-104.

[2] Adams, J.F., *No more Division Algebras?* Address to the Trinity Mathematical Society, 3rd February 1959.

[3] Adams, J.F., *Real Numbers, Complex Numbers, Quaternions - and then what?* Address to the New Pythagoreans, 3rd November 1981.

[4] Adams, J.F., *Real Numbers, Complex Numbers, Quaternions, Cayley Numbers and ...?* Address to be given to the Trinity Mathematical Society on 7th May 1984.

[5] Cohn, P.M., *Algebra* Vol II (1977) Wiley

[6] Dickson, L.E., *Linear Algebras* Hafner, New York, (1914, 1960)

[7] Gauss, C.F., Demonstratio nova theorematis omnem functionem algebraicam rationalem integram unius variabilis in factores reales primi vel secundi gradus resolvi posse. *Gauss Werke*, dritter Band, p.p. 1-32.

[8] Gauss, C.F., Demonstratio nova *altera* theorematis omnem functionem algebraicam rationalem integram unius variabilis in factores reales primi vel secundi gradus resolvi posse. *Gauss Werke*, dritter Band, p.p. 33–56.

[9] Hyland, J.M.E., *Who's afraid of a proof of Gauss?* Address to the Archimedeans, 7th October 1983.

[10] Hyland, J. M. E., and Taylor, P., Gauss' Second Proof, to appear in *Eureka* **45**.

[11] Jameson, G. J. O., *A First Course in Complex Functions* Chapman & Hall (1970) p.p. 83–85

[12] Korner, T.W., *Topological Groups* D.P.M.M.S. lecture notes (1983)

[13] Porteous, I.R., *Topological Geometry* (1969, 1981) C.U.P.

A new operation?

by N. Boston

If addition and multiplication are thought of as arising from $x+y+z$ and xyz respectively, then what sort of group operation does the remaining elementary symmetric polynomial $yz+zx+xy$ produce? In fact nothing essentially new, although there is a subtle twist as we shall see.

Let S be the set of all monic cubic polynomials $t^3+a_1t^2+a_2t+a_3$ with all coefficients and roots in a given field F of characteristic $\neq 2,3$. We are interested in subsets of S containing polynomials with roots of the form $x, y, (x*y)^{-1}$, where $*$ is some group operation, inverses denoted $(.)^{-1}$.

For example, if $T = \{p(t) \in S: a_1 = 0\}$, then $*$ is addition, whilst if $U = \{p(t) \in S: a_3 = 1\}$, then $*$ is multiplication. $V_k = \{p(t) \in S: a_2 = k\}$ $(k \in F)$ will be investigated in this article.

With this choice of subset, if $z = (x*y)^{-1}$, then $yz+zx+xy = k$ is satisfied, i.e. $(x*y)^{-1} = \dfrac{k-xy}{x+t}$ (1), defined if $y \neq -x$ (so x and $-x$ cannot both belong to any group obtained). What we ultimately seek is some largest group with elements in F and operation $*$. The existence of an identity element e is therefore desirable, i.e. $3e^2 = k$ should be soluble $(e = (e*e)^{-1})$. For simplicity we take $k=3, e=1$.

The inverse of x will be $(x*e)^{-1} = \dfrac{3-x}{1+x}$ (2), defined if $x \neq -1$, which when combined with (1) gives $x*y = \dfrac{3x+3y-3+xy}{x+y+3-xy}$ (3), which <u>is</u> associative!

There is now a surprisingly simple trick, which clarifies this whole business. Enlarging F to F' containing $\sqrt{-3} = c$, we can define a bijective transformation $f: F'-\{c\} \to F'-\{1\}$ by $f(x) = \dfrac{x+c}{x-c}$. In particular $f(1)$ is a nontrivial cube root of 1; call it ω.

The usefulness of f lies in the key identity $f(x)f(y) = \omega f(x*y)$ (4), which is verified by simply expanding both sides and using (3). Thus $\dfrac{f(x)}{\omega} \cdot \dfrac{f(y)}{\omega} = \dfrac{f(x*y)}{\omega}$ (where defined), implying that there exists a bijection between $x \in F'-\{c\}$ and $\dfrac{f(x)}{\omega} \in F'-\{\omega^2\}$, which preserves the group structures (on the LHS $*$, on the RHS multiplication in $F'-\{0\}$). It follows that the groups we may define under $*$ correspond to subgroups of $F'-\{0\}$ (under multiplication) not containing ω^2 ; call these admissible subgroups.

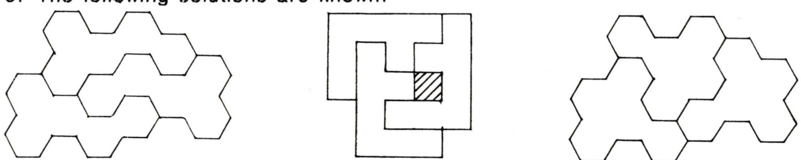
As an example let $F=F'$ be a finite field of prime order $p=1 \pmod 3$. Its multiplicative group is cyclic of order $p-1$, and its admissible subgroups are those of order not divisible by 3. In particular the maximal group G definable with operation $*$ is cyclic of order n , where $p-1 = 3^r n$ and $(n,3) = 1$. Finally note that the set $-G = \{x: -x \in G\}$ is another maximal group under $*$ with identity -1 .

Answers to Problems Drive

1. (i) 1/2 (ii) 5/12 1/2 (iii) 7/12 3/4 5/6 (iv) 11/30 7/15 1/2 (v) 19/30 9/10 16/15 17/16 7/6

2. There is only one 4-farm (finite or infinite) and no 5-farm. There are $(2^p-2)/p$ 3-farms of order 2^p, where 2^p-1 is prime, and also a trivial 3-farm of order 3. These were classified in *QARCH* 7 and 2-*Manifold* 4. We do not know whether there is an infinite 3-farm.

3. The following solutions are known:



4. The volume is $2clw + alw - alw\sqrt{(4c^2+l^2-h^2)}/\sqrt{(l^2-h^2)}$ where $x=a/2c$ is the unique solution of $\sinh x/x = \sqrt{(l^2-h^2)}/a$

5. $2^{23}-1 = 8388607 = 47 . 178481$ $\quad$ $698896 = 836^2$

6. $97905/(32250+11288\pi)$ m below the axis.
$(792500+225x)/(48375+16932\pi)$ m behind the geometrical centre.

7. $\pi(f^2/a + g^2/g - c)/\sqrt{(ab-h^2)}$

8. (i) 37 (ii) 53 (iii) 82 (iv) 468 (v) 546 (vi) 793 (vii) −827 (penalty marks for forgetting the minus sign!) (viii) 19 (ix) 81 (x) 567

9. Using $a+b+c=2s$ $ab+bc+ca=s^2+\ell^2+4rR$ and $abc=4rsR$ (see *QARCH* 1 & 2) in the method outlined in *Eureka* **42** (1982) 30 you get the results.

10. $\sin 9^\circ = \sqrt{(\frac{1}{2}-\frac{1}{8}\sqrt{(10+2\sqrt{5})})}$ with ++, −− and +− signs corresponding to the sines of 81°, 27° and 63°.

11. Using arithmetic modulo nine and logarithms to provide a bound, in the first case D=7. In the second case we're left with either 1 or 10, with D=1 only if C=10. Assuming the digits of 1981^{1981} to be independent and uniformly distributed, their sum has mean 29395 and standard deviation 984.89, so the normal distribution gives a probability of 0.0035 that the sum is outside (28000, 30007).

12. (12 hrs)/(π rad) $(\pi+2\cos^{-1}(\tan\alpha \tan\theta))$.