Eureka Digital Archive

archim.org.uk/eureka

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 1978 NUMBER 39

EUREKA

THE JOURNAL OF THE ARCHIMEDEANS

NUMBER 39, APRIL 1978

Price 50p

EDITORS: A. J. WASSERMANN
R. C. MASON

BUSINESS MANAGER: D. B. DUVAL

Editorial

In the thirty-third edition of Eureka the editors, all Churchill men, wrote that Eureka had "lost almost all its bias towards Cambridge". This was only the start of a trend which was ultimately to result in the introduction of such new-fangled devices as computers for the production of the magazine. The present editors, however, both Johnians, have sought to revive the old values and once again instil the spirit of Cambridge into Eureka: all the articles this year have been written by people who have studied at Cambridge.

We would like to thank the following for their help with Eureka: the Cambridge Pure and Applied Mathematics Departments for their cooperation; Mrs. Lori Relizani and Mrs. Robin Bringans for typing the manuscript; B. Heydecker and R. Taylor for photographs; O. L. C. Toller and J. Gilby for illustrations; J. Mestel and J. Rickard for problems; and A. N. S. Freeling, N. D. Hooker, M. R. Kipling and the Dragon for their helpful suggestions.

Subscriptions

We regret that, due to continually rising printing costs, future subscriptions will only entitle subscribers to receive copies post free; they will not be entitled to receive copies at below the normal sale price, as has been the case in the past. This will not affect present subscribers until their subscriptions run out.

Cheques, etc., should be made payable to The Archimedeans, "Eureka", and sent to:

The Business Manager, "Eureka",
The Arts School,
Bene't Street,
Cambridge,
CB2 3PY
England.

Infinite-Dimensional Vector Spaces

by Professor J.R.Ringrose

1. Finite systems of linear equations

A simple but important problem, which motivates much of the elementary theory of finite-dimensional vector spaces, is that of analysing a system of finitely many linear equations in finitely many unknowns. Even though this material is probably familiar to most readers, I shall say a little about it, becasue it leads naturally towards some of the questions considered later in this article. It will be sufficient to consider "square" systems

$$\left.\begin{array}{l} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 , \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2 , \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n = b_n , \end{array}\right\} \quad (1)$$

in which there are n equations involving n unknowns, $x_1, \ldots, x_n$. I shall suppose that the given coefficients, a_{ij}, b_i, are complex numbers, and that $x_1, \ldots, x_n$ are to take complex values.

Let $\mathbb{C}$ denote the complex field, and let $\mathbb{C}^n$ denote the vector space consisting of all n-tuples $\underline{c} = (c_1, \ldots, c_n)$ of complex numbers; although, for typographical convenience, such n-tuples are written as <u>rows</u>, I shall in fact treat them as <u>columns</u>, that is $n \times 1$ matrices. Associated with the $n \times n$ matrix $A = [a_{ij}]$, there is a linear operator T_A from $\mathbb{C}^n$ into $\mathbb{C}^n$, defined by $T_A(\underline{c}) = A\underline{c}$. The system (1) can be rewritten in the form

$$T_A(\underline{x}) = \underline{b} \qquad (1')$$

where $\underline{b}$ is a given element of $\mathbb{C}^n$ and $\underline{x}$ is to be found in $\mathbb{C}^n$. We can now make use of the following theorem from elementary linear algebra.

<u>THE "ALTERNATIVE" THEOREM</u>. If T is a linear operator from a finite-dimensional vector space V into itself, then either

(a) T has an inverse operator T^{-1},

or (b) the range $\{T(v): v \in V\}$ of T is not the whole of V, and the null-space $\{v \in V: T(v) = 0\}$ contains non-zero vectors.

If T_A has an inverse then, for every choice of $\underline{b}$, the equation (1') has a unique solution, $\underline{x} = T_A^{-1}(\underline{b})$. When T_A has no inverse, then either $\underline{b}$ lies outside the range of T_A (in which case there is no solution $\underline{x}$ of

(1')) or $\underline{b}$ lies in the range of T_A (in which case (1') has a solution $\underline{x}_0$, and in fact has infinitely many solutions, each obtained by adding to $\underline{x}_0$ a vector in the null-space of T_A).

From the preceding paragraph it follows that, for the system (1), there are two main possibilities. In the "good" case, for every choice of $b_1,\ldots,b_n$ there is a unique solution $x_1,\ldots,x_n$. In the "bad" case, there is either no solution at all, or an infinity of different solutions, depending on the values of $b_1,\ldots,b_n$. I imagine that most readers will be aware that the first possibility occurs when the determinant detA is non-zero, the second when det A = 0.

Of course, it is possible to analyse a system of the type (1) without using either vector space theory or determinants. Observe that, by applying certain "elementary row operations" to such a system, we obtain another system which has exactly the same solutions as the first one. The operations to which I refer are

- (i) the interchange of two equations,
- (ii) multiplication of an equation by a non-zero constant,
- (iii) addition, to one equation, of a multiple of another one.

It is not difficult to verify that, after applying a finite number of such operations, and then renumbering (that is, changing the order of) the unknowns $x_1,\ldots,x_n$, the original system (1) can be replaced by an equivalent system, having the "reduced" form

$$\left.\begin{aligned} \varepsilon_1(x_1 + k_{12}x_2 + k_{13}x_3 + \ldots + k_{1n-1}x_{n-1} + k_{1n}x_n) &= y_1 \\ \varepsilon_2(x_2 + k_{23}x_3 + \ldots + k_{2n-1}x_{n-1} + k_{2n}x_n) &= y_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ \varepsilon_{n-1}(x_{n-1} + k_{n-1n}x_n) &= y_{n-1} \\ \varepsilon_n x_n &= y_n \end{aligned}\right\} \quad (2)$$

in which (for some integer r such that $0 < r \le n$)

$$\varepsilon_i = 1 (1 \le i \le r), \quad \varepsilon_i = 0 (r < i \le n).$$

<u>If $r = n$</u>, then ε_i is 1, and it is easy to calculate $x_n, x_{n-1},\ldots,x_1$ (in that order), by using the equations of the system (2) in reverse order; in this way we obtain a unique solution to the system.

<u>If $r < n$</u>, then $\varepsilon_{r+1} = \varepsilon_{r+2} = \ldots = \varepsilon_n = 0$, and the system has no solution if any one of the given complex numbers $y_{r+1}, y_{r+2},\ldots, y_n$ is non-zero; if, however,

$$y_{r+1} = y_{r+2} = \ldots = y_n = 0,$$

then $x_{r+1},\ldots,x_n$ may be chosen <u>arbitrarily</u>, and we can then obtain unique values for $x_r, x_{r-1},\ldots,x_1$ by using the first r equations of the system (2), in reverse order. The "good" case occurs when r = n, the "bad" case when $r < n$.

In the "good" case in which each ε_i is 1, there is a different method of solving the system (2); and although it is clumsy in the present context, it suggests a useful approach to one of the infinite-dimensional problems discussed below. The matrix of the system (2) (in the "good" case) is $I_n + K$, where I_n denotes the $n \times n$ identity matrix and K is the $n \times n$ matrix which has the given coefficients k_{ij} above the main diagonal, and has zero in each entry on or below that diagonal. The system (2) can be replaced by the equation

$$\underline{x} + T_K(\underline{x}) = \underline{y} \qquad (2')$$

where $\underline{y}$ is a given element of $\mathbb{C}^n$, and $\underline{x}$ is to be found in $\mathbb{C}^n$. Elementary calculation shows that K^2 has zero in each entry on, below, or just one step above the main diagonal, K^3 has zero in each entry on, below, or at most two steps above the main diagonal, and so on. In particular, by proceeding in this way we can show that $K^n = 0$, and from this it follows that the operator T_K satisfies $T_K^n = 0$.

If $\underline{y}$ is a given element of $\mathbb{C}^n$, we can rewrite (2') in the form $\underline{x} = \underline{y} - T_K(\underline{x})$, and iteration yields

$$\begin{aligned} \underline{x} &= \underline{y} - T_K(\underline{y} - T_K(\underline{x})) = \underline{y} - T_K(\underline{y}) + T_K^2(\underline{x}) \\ &= \underline{y} - T_K(\underline{y}) + T_K^2(\underline{y} - T_K(\underline{x})) = \underline{y} - T_K(\underline{y}) + T_K^2(\underline{y}) - T_K^3(\underline{x}) \\ &= \ldots = \underline{y} - T_K(\underline{y}) + T_K^2(\underline{y}) + \ldots + (-1)^{n-1}T_K^{n-1}(\underline{y}) + (-1)^n T_K^n(\underline{x}). \end{aligned}$$

Since $T_K^n = 0$, this shows that the only possible solution $\underline{x}$ of (2') is given by

$$\underline{x} = \underline{y} - T_K(\underline{y}) + T_K^2(\underline{y}) - \ldots + (-1)^{n-1}T_K^{n-1}(\underline{y}); \qquad (3)$$

and (again using the fact that $T_K^n = 0$), it is immediately verified that this vector $\underline{x}$ does indeed satisfy (2'). Accordingly, for each $\underline{y}$ in $\mathbb{C}^n$, (2') has a unique solution $\underline{x}$, which is given by (3).

2. Infinite systems of linear equations

Having set the scene in finite dimensions, I shall now turn to some infinite-dimensional analogues of the problems considered above. Perhaps the most obvious thing to try first is a system

$$\sum_{j=1}^{\infty} a_{ij}x_j = b_i \quad (i = 1,2,3,\ldots) \qquad (4)$$

of infinitely many linear equations in infinitely many unknowns. To deal with such a system by vector space methods, it will be necessary to replace $\mathbb{C}^n$ by another space, in which each vector $\underline{c} = (c_1, c_2, \ldots)$ has infinitely many coordinates. It is not quite clear how this should be done, but since we want the series in (4) to converge, we are likely to need some restriction on the coordinates c_n. In Section 4 below, we give some examples of vector spaces which might be used for this purpose.

One example will suffice to show that the simple alternative noted above for finite systems, with a "good" case and a "bad" case, does not apply to infinite systems. Consider the system

$$\left.\begin{array}{r} x_1 + x_2 + x_3 + x_4 + \dots + x_n + x_{n+1} + \dots = y_1 \\ x_2 + x_3 + x_4 + \dots + x_n + x_{n+1} + \dots = y_2 \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ x_n + x_{n+1} + \dots = y_n \end{array}\right\} \quad (5)$$

$$\dots\dots\dots$$

This is an infinite-dimensional analogue of the most straightforward of the finite systems considered above, the reduced system (2) in the "good case" where each ε_i is 1. For even greater simplicity, we have taken each coefficient k_{ij} to be 1.

Neither of the methods, used above to solve (2), can be applied to the system (5); for example, we cannot start with the last equation (and work back from there) because there isn't one. However, an easy procedure is available. Upon subtracting from each equation the one that follows it, we deduce that

$$x_1 = y_1 - y_2,\ x_2 = y_2 - y_3, \dots, x_n = y_n - y_{n+1}, \dots; \quad (6)$$

and (apparently) obtain a unique solution. If you believe this, you will obtain interesting conclusions by looking at particular cases; for example, try $y_1 = y_2 = \dots = 1$, or $y_1 = 1, y_2 = 2, y_3 = 3, \dots$ This should convince you that, for this system, we have one feature of the "good" case (when there is a solution, it is unique), but also one feature of the "bad" case (for certain choices of $y_1, y_2, \dots$, there is no solution).

In fact, it is easy to show that (5) has a solution (then necessarily given by (6)) if and only if $\lim y_n = 0$. This emphasizes something which is already implicit in the occurrence of infinite series in (5); the problem is no longer purely algebraic, since some of the concepts of analysis are now required.

Although it is no longer an area of major research activity, there is an extensive theory of infinite systems of linear equations. I cannot conclude this section without mentioning the book [2] by F.Riesz; it is worth a trip to the library, even if you don't get beyond the historical remarks made in the preface.

3. Linear integral equations

We turn now to a second subject, which can reasonably be viewed as an infinite-dimensional analogue of the topic discussed in Section 1. In the system

$$\sum_{j=1}^{n} a_{ij} x_j = b_i \qquad (i = 1, \dots, n) \qquad (1)$$

we replace the finite sequences $x_1,\ldots,x_n$ and $b_1,\ldots b_n$ by functions x(s) and b(s) of a continuous variable, $0 \le s \le 1$, and the coefficients a_{ij} by a function a(s,t). If we recall also that an integral can be viewed as the limit of certain (finite) approximating sums, it is natural to replace the sums in (1) by integrals. In this way we arrive at the equation

$$\int_0^1 a(s,t)x(t)dt = b(s) \qquad (0 \le s \le 1) \tag{7}$$

It does not change the nature of the system (1) if the diagonal coefficient a_{ii} is replaced by $1 + a_{ii}$, for each $i = 1,\ldots,n$. With this modification to (1), the above process of analogy leads to an equation

$$x(s) + \int_0^1 a(s,t)x(t)dt = b(s) \qquad (0 \le s \le 1) \tag{8}$$

When the same process is applied to the reduced system (2), in the "good" case where each ε_i is 1, we arrive at the equation

$$x(s) + \int_0^1 k(s,t)x(t)dt = y(s) \qquad (0 \le s \le 1) \tag{9}$$

The three equations (7),(8) and (9) are examples of <u>linear integral equations</u>; the functions a(s,t),b(s), k(s,t), y(s) are given, and the function x(s) is to be found. Of course, one requires some restriction, such as continuity, on all these functions, to ensure the existence of the integrals on the left hand sides. Equations such as (7) and (8) are described as <u>Fredholm integral equations</u> (of the <u>first type</u> and <u>second type</u>, respectively); those such as (9), in which the range of integration has the variable limit s, are described as <u>Volterra integral equations</u>.

In order to treat such equations by vector space methods, we shall need to use a space in which a "vector" is a function c(s) $(0 \le s \le 1)$. Our infinite-dimensional problems have moved much further towards analysis, rather than algebra. By now, however, it is time to look at the kind of vector spaces we need.

4. <u>Banach spaces</u>

Suppose that X is a vector space, with complex scalars. By a <u>norm on X</u> we mean a function which assigns, to each vector x in X, a real number denoted by $\|x\|$, in such a way that

(i) $\|x\| \ge 0$, $\|x\| = 0$ if and only if $x = 0$;

(ii) $\|cx\| = |c|\,\|x\|$;

(iii) $\|x+y\| \le \|x\| + \|y\|$ (The triangle inequality);

whenever $x,y \in X$ and $c \in \mathbb{C}$. These three properties are analogous to those of the modulus function for complex numbers. By means of a norm on X, we can introduce various concepts of mathematical analysis, for elements of X. Suppose, for example, that $x, x_1, x_2,\ldots, s, y_1, y_2, \ldots \in X$, and $s_n = y_1 + \ldots + y_n$. We say that the sequence (x_n) converges to x if $\|x_n - x\| \to 0$ as $n \to \infty$; the series Σy_n converges to s if the sequence of partial sums, (s_n), converges to s. A mapping f, from X into X, is continuous if $(f(x_n))$ converges to $f(x)$ whenever (x_n) converges to x. A subset Y of X is closed if it contains all its limit points. Many familiar results from elementary analysis extend at once to this context; for example, "limit of sum = sum of limits", "if a series converges, the terms tend to 0". In order to develop a really worthwhile theory, we need one more assumption:

(iv) If $x_1, x_2, \ldots \in X$ and $\Sigma \|x_n\| < \infty$, the series Σx_n converges to an element x of X.

By a Banach space we mean a complex vector space with a norm which satisfies (iv) (as well as (i), (ii), (iii)). Experts will notice that (iv), although differing from the usual form of the completeness axiom, is nevertheless equivalent to it.

The vector space $\mathbb{C}^n$ becomes a Banach space if the norm of a vector $\underline{c} = (c_1, \ldots, c_n)$ is defined in any one of the following ways:

(a) $\underline{c} = \max\{|c_1|, |c_2|, \ldots, |c_n|\}$;

(b) $\underline{c} = |c_1| + |c_2| + \ldots + |c_n|$;

(c) $\underline{c} = [|c_1|^2 + |c_2|^2 + \ldots + |c_n|^2]^{\frac{1}{2}}$

In each case, the above properties (i) and (ii) of norms are apparent; and (iv) follows from the fact that absolute convergence implies convergence, for a series of complex numbers. The triangle inequality is apparent in cases (a) and (b), and reduces to a classical inequality in case (c).

The following examples are nore interesting, although the verification of the axioms, especially (iv), requires more effort in some cases; I omit the details, and suggest that you do too, unless you are feeling especially enthusiastic.

EXAMPLE 1. Let X_1 be the set of all bounded sequences $\underline{c} = (c_1, c_2, \ldots)$ of complex numbers. Given two such sequences $\underline{c}$ and $\underline{d}$ and a complex number a, define

$$\underline{c} + \underline{d} = (c_1 + d_1, c_2 + d_2, \ldots), \quad a\underline{c} = (ac_1, ac_2, \ldots).$$

Then X_1 is a complex vector space, and is a Banach space when the norm is defined by $\|\underline{c}\| = \sup_n |c_n|$.

EXAMPLE 2. Let X_2 be the set of all complex sequences $\underline{c} = (c_1, c_2, \ldots)$ such that $\Sigma |c_n| < \infty$. Just as in Example 1, this set becomes a complex vector space. It is a Banach space when the norm is defined by $\|\underline{c}\| = \Sigma |c_n|$.

EXAMPLE 3. Let X_3 be the set of all complex sequences $\underline{c} = (c_1, c_2, \ldots)$ such that $\Sigma |c_n|^2 < \infty$. Just as in the previous examples, this set becomes a vector space.

It is a Banach space when the norm is defined by $\|\underline{c}\| = [\Sigma|c_n|^2]^{\frac{1}{2}}$.

These three examples can be viewed as infinite-dimensional analogues of $\mathbb{C}^n$ (with the three norms considered in (a), (b), (c) above). Each of them might be useful in considering suitable problems about infinite systems of linear equations. For linear integral equations, the next two examples are appropriate.

<u>EXAMPLE 4</u>. Let X_4 be the set of all complex-valued functions x(s), defined and continuous on the interval $0 \le s \le 1$. With the obvious definitions for sums and complex multiples of such functions, X_4 is a complex vector space. It is a Banach space when the norm is defined by

$$\|x\| = \sup_{0\le s\le 1} |x(s)| .$$

[In fact, convergence in this space is "uniform convergence on the interval $0 \le s \le 1$", and axiom (iv) is the "Weierstrass M-test" for uniform convergence of infinite series of functions.]

A complete understanding of the next (and last) example requires a knowledge of Lebesgue integration; but if you don't have this, you will come to little harm by ignoring the word you don't understand.

<u>EXAMPLE 5</u>. Let X_5 be the set of all complex-valued functions x(s), defined and measurable on the interval $0 \le s \le 1$. and such that

$$\int_0^1 |x(s)|^2 ds < \infty .$$

Then X_5 is a complex vector space, and becomes a Banach space when the norm is defined by

$$\|x\| = [\int_0^1 |x(s)|^2 \, ds]^{\frac{1}{2}} .$$

Among the examples mentioned above, X_3 and X_5 are probably the most important. Although they look very different at first sight, it can be shown that they are "essentially the same"; by this I mean that there is a one to one linear mapping U, from X_3 onto X_5, such that $\|U(x)\| = \|x\|$ for each x in X_3. Both X_3 and X_5 are examples of the simplest type of infinite-dimensional Banach spaces, namely <u>Hilbert spaces</u>.

In Section 1, we used an iterative process to obtain the solution of (2') in the form (3). The following theorem shows that, under certain conditions, a similar process can be used in Banach spaces.

<u>THEOREM</u> Suppose that T is a continuous linear operator from a Banach space X into itself, and, for each x in X, the series $\Sigma_0^\infty \|T^n(x)\|$ converges. Then, given any y in X, the equation

$$x + T(x) = y \tag{10}$$

has a unique solution x in X, given by

$$x = y - T(y) + T^2(y) - T^3(y) + \dots \qquad (11)$$

Proof Since $\Sigma_0^\infty \|(-1)^n T^n(y)\| = \Sigma_0^\infty \|T^n(y)\| < \infty$, it follows that the series in (11) converges (and hence, $T^n(y) \to 0$ as $n \to \infty$) for each y in X. If we define x_0 in X by

$$x_0 = y - T(y) + T^2(y) - T^3(y) + \dots ,$$

then

$$Tx_0 = T(y) - T^2(y) + T^3(y) - \dots$$
$$= y - x_0 ;$$

so x_0 satisfies (10).

Conversely, if x is any solution of (10), an iterative procedure (similar to the one used to obtain the solution of (2') in the form (3) shows that

$$x = y - T(y) + T^2(y) - \dots + (-1)^{n-1}T^{n-1}(y) + (-1)^n T^n(x),$$

for $n = 1, 2, \dots$; when $n \to \infty$, $T^n(x) \to 0$, and thus

$$x = y - T(y) + T^2(y) - \dots = x_0.$$

Accordingly, (10) has a unique solution x in X, given by (11).

We now use the above theorem to sketch a proof that, when the functions $k(s,t)$ and $y(s)$ are continuous, there is a unique continuous function $x(s)$ which satisfies the Volterra integral equation (9). Of course, this is what one might have hoped for, since (9) was obtained by analogy from the reduced system (2) (in the "good" case), which always has a unique solution.

For X we use the Banach space X_4 described above; so our vectors are continuous complex-valued functions $x(s)$ defined on the interval $0 \le s \le 1$. Given such a function x, we can define another, Tx, by the equation

$$(Tx)(s) = \int_0^1 k(s,t)x(t)dt \quad (0 \le s \le 1). \qquad (12)$$

It is easy to check that T is a linear operator from X into X. We shall write $T^n x$, rather than $T^n(x)$ as hitherto, to avoid excessive proliferation of brackets. The integral equation (9) can now be written in the form $x + Tx = y$; we shall establish the existence of of a unique solution by showing that T satisfies the conditions set out in the above theorem.

Let M be a constant such that $|k(s,t)| \le M$ for all s,t. Upon replacing x by $T^n x$ in (12), we obtain

$$(T^{n+1}x)(s) = \int_s^1 k(s,t)(T^n x)(t)dt, \text{ and therefore}$$

$$(T^{n+1}x)(s) \le \int_s^1 |k(s,t)|\ |T^n x)(t)|dt$$

$$\le M \int_s^1 |(T^n x)(t)|dt.$$

From this, it is easy to check (by induction on n) that

$$|(T^n x)(s)| \leq \frac{M^n (1-s)^n \|x\|}{n!} \quad (0 \leq s \leq 1;\ n = 0,1,2,\ldots) \quad (13)$$

when $n = 0$, the required inequality $|x(s)| \leq \|x\|$ is apparent, since $\|x\|$ is defined as $\sup |x(s)|$.

By taking the suprema of both sides of (13) for $0 \leq s \leq 1$, we obtain

$$\|T^n x\| \leq \frac{M^n \|x\|}{n!} \ ;$$

in particular, $\|Tx\| \leq M \|x\|$. From this, it is apparent that $\Sigma \|T^n x\| < \infty$, whenever $x \in X$. Moreover T is continuous; for if (x_n) converges to x, we have

$$\|Tx_n - Tx\| = \|T(x_n - x)\| \leq M \|x_n - x\| \to 0 \ ,$$

and therefore (Tx_n) converges to Tx. Accordingly, we have verified that T satisfies the conditions required in the above theorem; and as already noted, this suffices to prove the existence of a unique solution to the Volterra integral equation (9).

The Fredholm integral equation (8) (with all functions again continuous) can be considered in much the same way. We again use the Banach space X_4, and define a continuous linear operator by

$$(Tx)(s) = \int_0^1 a(s,t)x(t)dt \quad (0 \leq s \leq 1) . \quad (14)$$

With minor modifications, the method used above for (9) can be applied here also, to prove the existence of a unique solution of (8), provided there is a constant M such that $|a(s,t)| \leq M < 1$, for all s and t. Without the conditon $M < 1$, the method fails (and indeed, (8) may have no solution).

In this section, I have tried to show that Banach spaces and continuous linear operators arise naturally, and are relevant to certain classical problems. The examples I have chosen, involving linear integral equations are among the earliest and the easiest. There are many more recent, and more sophisticated applications, for example in the theory of partial differential equations.

5. <u>Some problems, great and small</u>

Section 1 includes the Alternative Theorem, for a linear operator acting on a finite-dimensional vector space. Does it apply also to every continuous linear operator acting on a Banach space? It does not! In the Banach space X_4 (continuous functions) we can define a continuous linear M by the equation

$$M(x)(s) = sx(s) \quad (0 \leq s \leq 1).$$

It is easy to check that the null space of M consists of the zero function only; but M has no inverse, since $(Mx)(0) = 0$ for each x, and thus the range of M does not contain all continuous functions.

When $a(s,t)$ is a continuous function, equation (14) defines a continuous linear operator T on the Banach space X_4. It can be shown that the Alternative Theorem applies to the operator I + T, but not (in general) to T itself.

This is good for the integral equation (8), but not so good for (7).

A linear operator acting on a finite-dimensional complex vector space always has an eigenvector. However, this result does not extend to continuous linear operators acting on Banach spaces; for example, it is not difficult to show that the operator M, defined above, has no eigenvector.

If T is a linear operator, acting on a complex vector space X of finite dimension $n(>1)$, there is a subspace L of X such that $\{0\} \neq L \neq X$ and $Tx \in L$ whenever $x \in L$ (why?). Does the result apply in an infinite-dimensional Banach space X, if we assume that T is continuous, and require L to be closed? This problem is much harder than the earlier ones. There is an example (as yet unpublished), due to the Scandinavian mathematician P. Enflo, which shows that in general the answer is negative. In the case of Hilbert space operators, the question remains a well known unsolved problem - the "invariant subspace problem".

Every finite-dimensional vector space has a basis. In an infinite-dimensional Banach space X, it seems reasonable to define a basis as a sequence $e_1, e_2, \ldots$ in X, with the following property: each x in X can be represented uniquely as the sum of a convergent series $\sum c_n e_n$, in which $c_1, c_2 \ldots$ are scalars. (Such a basis is called a <u>Schauder basis</u>.) Does every Banach space have a basis?

One has to confine attention to Banach spaces which are separable (which means "not too big", in a certain sense). This "basis problem" was recently solved (in the negative), again by P. Enflo [1], having been first posed over forty years earlier. All the particular Banach spaces described in Section 4 have bases, except X_1 which is not separable, but the proof is none too easy in the case of X_4.

Banach spaces provide the right setting in which to consider a variety of problems. The theory of such spaces includes many deep and powerful theorems, and many hard unsolved problems.

References

[1] P. Enflo, "A counterexample to the approximation problem in Banach spaces", Acta Math. 130 (1973), 309-317.

[2] F. Riesz, Les systémes d'équations linéaires à une infinité d'inconnues, (Paris, Gauthier-Villars, 1913).

The Archimedeans

by Roger Dix, President 1976-1977

The last two years have seen a generally successful attempt by the committees involved to blend serious mathematical discussion with events of a more informal nature.

The 1975/6 programme of evening talks included Professor John Taylor of King's College, London who believed in 'spoon-bending' (Science and ESP); Professor E. C. Zeeman of Warwick University who believed that the Greeks were mathematically ahead of their time (The lost group theory of Eudoxus); and Mr. G. E. Perry of Kettering Grammar School who believed that something Red up there is watching us (The Soviet Space Programme).

Four Lunch meetings were held each term and were reasonably well attended. The Invariants were defeated in the problems drive and a tape recording of the President (Hilary Stark) opened the games afternoon on our visit to Oxford. The perennial pre- and post-examination events were enjoyed by all who attended; the ramble following the Roman road to Hildersham on a scorching pre-drought day and then, at closing time, wending its way somewhat stochastically back to Cambridge.

In 1976/7, the evening speakers had a varied reception. A small but enthusiastic crowd gathered for the more serious talks such as the one on the race to find simple groups of large order (The search for simplicity) given by Dr. P. M. Neumann of the Queen's College, Oxford; whereas more general lectures, such as the one on peculiarities of perception (Visual space) given by Professor Richard Gregory of Bristol University, were better attended. The meeting most well remembered, however, must have been the one addressed by Mr. G. Spencer-Brown on "A Proof of the Four Colour Theorem". Unfortunately, the minutes secretary was unable to follow what was said, as were the rest of the audience, so no record of this 'proof' is available. Room A of the Arts School was filled to capacity with latecomers having to stand - a record?

At the first lunch meeting, addressed by Dr. Reid, we were unfortunately unable to serve food but we were back to normal at the second meeting, when Dr. Conway spake on "Formulae for π and other things", and for the six other meetings.

On November 26th 1976, the Society held its Triennial Dinner in the Old Kitchens, Trinity College. We were honoured to have as guests Drs. J. C. Burkill, H. Burkill and K. Moffatt. A most enjoyable time was had by the seventy or so members and their friends who attended.

It was sad to note the demise of both the Bridge Group and the Puzzles and Games Ring during the year: the continual growth of new 'games' societies in Cambridge has taken its toll at last. The Bookshop and Computer Group continued to operate smoothly.

Unfortunately, we lost the Problems Drive, probably for the first time in its history (although the wooden spoon still went to an Oxford pair!), and so we are now more determined than ever to regain the ashes of the pencils and paper burned that sad evening when we visit Oxford next year.

We were pleased to welcome a group of about thirty maths/physics students from Utrecht State University to Cambridge and, together with CUPS, to give them a guided tour of the colleges, followed by a reception and dinner in Trinity College Hall, finishing up in a local pub, or the computer lab. according to taste.

Some of the summer events were marred by bad weather and the new president defied aquatic tradition by not entering the Cam, nay, not even entering a punt, on the trip to Grantchester. He had, however, been immersed several times during the punt jousting and claimed that this was sufficient, though of course not really necessary!

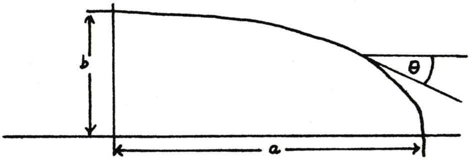
Our thanks are due to all the members of the committee over these two years, to the college representatives, to members who have lent a hand at some time or other and especially to the two secretaries, Dave Peters (who, due to a disagreement with the Part IB examiners, is no longer with us) and Mike Kipling.

The Disappointing Mean Circle

by J. J. Hitchcock

It can easily be shown that an ellipse with semi-major axis a and semi-minor axis b has perimeter

$$P = \int_0^{2\pi}\sqrt{a^2\cos^2\theta + b^2\sin^2\theta}\, d\theta \;;$$



this equals 4aE(m), where E is the complete elliptic integral of the second kind, and the parameter m is the square of the eccentricity, i.e.

$$m = \frac{a^2 - b^2}{a} .$$

According to all the books (e.g. [1]) E(m) cannot be expressed in terms of elementary functions; but it can be expanded as a power series which gives

$$P = 2\pi a \left(1 - \frac{m}{4} - \frac{3}{64}m^2 - \frac{5}{256}m^3 - \frac{175}{16384}m^4 - \ldots\right) .$$

This article is about finding a simple approximation to P by taking the circumference of a circle with radius equal to some mean of a and b.

The obvious answer is $2\pi(a + b)/2$, while Spiegel [2] suggests using a root-mean-square radius, $2\pi\sqrt{(a^2+b^2)/2}$. To find the "best" solution, consider the generalised mean

$$2\pi \left(\frac{a^p + b^p}{2}\right)^{1/p}$$

Expressing this in terms of a and m, and expanding using the binomial theorem gives

$$2\pi a \left(1 - \frac{m}{4} - \frac{(3 - p)}{32} m^2 - \ldots\right) .$$

The first two terms are exact, so let $(3-p)/32 = 63/64$. This makes $p = 1\frac{1}{2}$. The next term in the series is found to be $-\frac{(7-3p)}{128}m^3$, and substituting $1\frac{1}{2}$ for p gives $-\frac{5}{256}m^3$, which surprisingly is also exact.

When I discovered this, I set about calculating the next coefficient, and when twenty minutes later I found that it was $\frac{-176}{16384}m^4$ (and not $\frac{-175}{16384}m^4$) I felt that there must have been a mistake in my arithmetic. Unfortunately there wasn't.

Accepting that it <u>is</u> an approximation, how good is it? As I have only calculated the next two coefficients (they are correct to within 4%), consider two "practical" examples:

Let a = 6378.142 km, b = 6356.757 km, then P (the polar circumference of the earth) is given with an error of about 0.0049 mm.

Let a = 1, b = 0. The perimeter of this degenerate ellipse is 4, while $2\pi(\frac{1}{2})^{2/3} = 3.958$; so the error is only about 1 in 100.

Anyone wishing to consider the matter further should consult Smirnov [3] where there is a similar (but less accurate) approximation.

<u>References</u>

[1] M. Abromowitz and I. A. Stegun, "Handbook of Mathematical Functions", p. 589.

[2] M. R. Spiegel, "Mathematical Handbook", p. 7.

[3] V. I. Smirnov, "A Course of Higher Mathematics; Vol I", pp. 345-348.

Pentaplexity

A class of non-periodic tilings of the plane

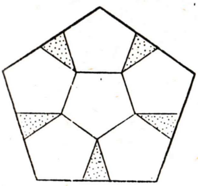
by Professor R. Penrose

Some readers may be acquainted with an article by Martin Gardner in the January 1977 issue of Scientific American. In this he described a pair of plane shapes that I had found, called 'kites' and 'darts', which, when matched according to certain simple rules, could tile the entire plane, but only in a non-periodic way. The tilings have a number of remarkable properties, some of which were described in Gardner's article. I shall give here a brief account explaining how these tiles came about and indicating some of their properties.

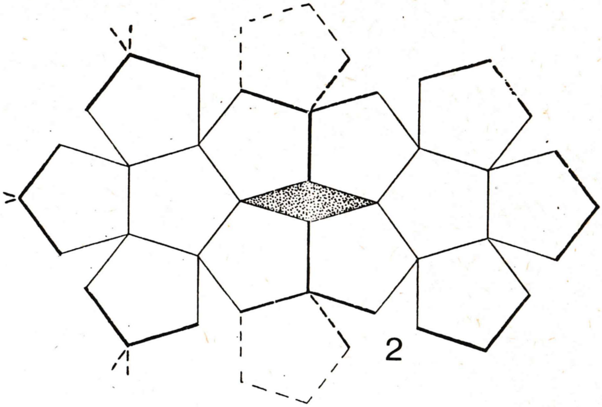
The starting point was the observation that a regular pentagon can be subdivided into six smaller ones, leaving only five slim triangular gaps. (See Fig. 1; this is familiar as part of the usual 'net' which folds into a regular dodecahedron.) Imagine, now, that this process is repeated a large number of times, where at each stage the pentagons of the figure are subdivided according to the scheme of Fig. 1. There will then be gaps appearing of varying shapes and we wish to see how best to fill these. At the second stage of subdivision, diamond-shaped gaps appear between the pentagons (Fig. 2). At the third, these diamonds grow 'spikes', but it is possible to find room, within each such 'spiky diamond', for another pentagon, so that the gap separates into a star (pentagram) and a 'paper boat' (or jester's cap?) (Fig. 3). At the next stage, the star and the boat also grow 'spikes', and, likewise, we can find room for new pentagons within them, the remaining gaps being new stars and boats (as before). These subdivisions are shown in Fig. 4.

Since no new shapes are now introduced at subsequent stages, we can envisage this subdivision process proceeding indefinitely. At each stage, the scale of the shapes can be expanded outwards so that the new pentagons that arise become the same size as those at the previous stage. As things stand, however, this procedure allows an ambiguity that we would like to remove. The subdivision of a 'spiky diamond' can be achieved in two ways, since there are two alternative positions for the pentagon. Let us insist on just <u>one</u> of these, the rule being that given in Fig. 5. (When we examine the pattern of surrounding pentagons we necessarily find that they are arranged in the type of configuration shown in Fig. 5). It may be mentioned that had the opposite rule been adopted for subdividing a 'spiky diamond', then a contradiction would appear at the next stage of subdivision, but this never happens with the rule of Fig. 5.

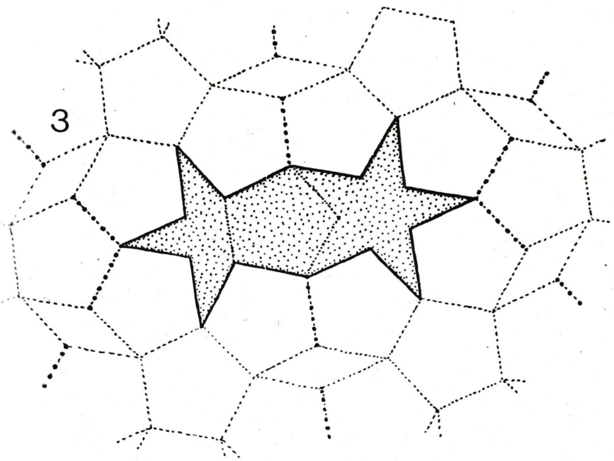
This procedure, when continued to the limit, leads to a tiling of the entire plane with pentagons, diamonds, boats



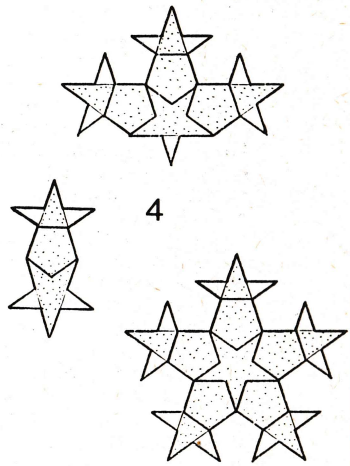
1



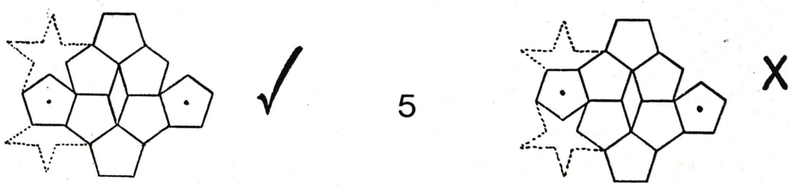
2



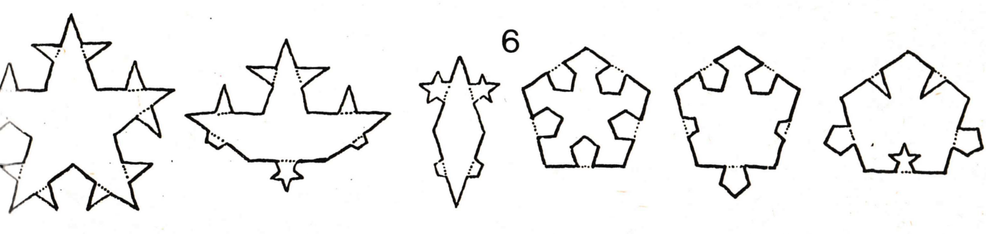
3



4



5



6

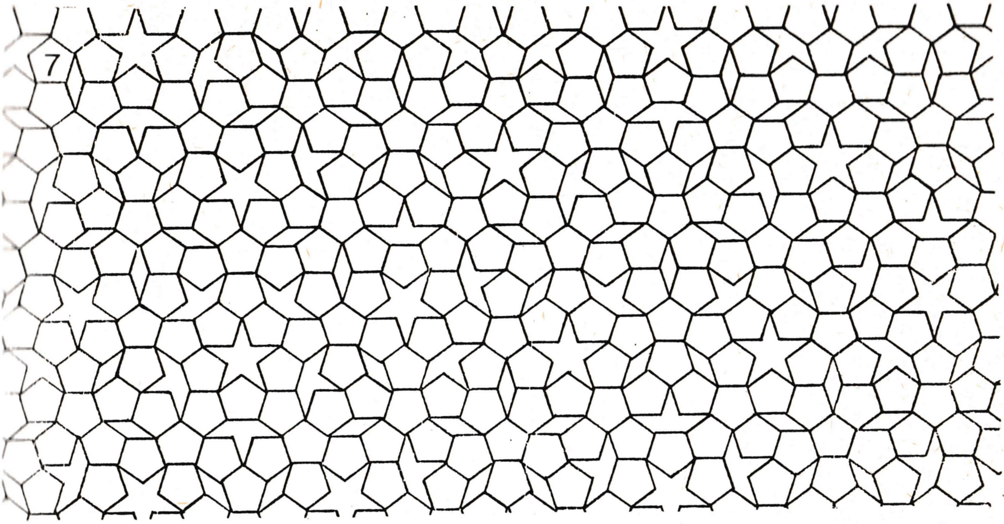
and stars. But there are many 'incorrect' tilings with these same shapes, being not constructed according to the above prescription. In fact, 'correctness' can be *forced* by adopting suitable matching rules. The clearest way to depict these rules is to modify the shapes to make a kind of infinite jigsaw puzzle, where a suggested such modification is given in Fig. 6. It is not too hard to show that any tiling with these six shapes is forced to have a hierarchical structure of the type just described.

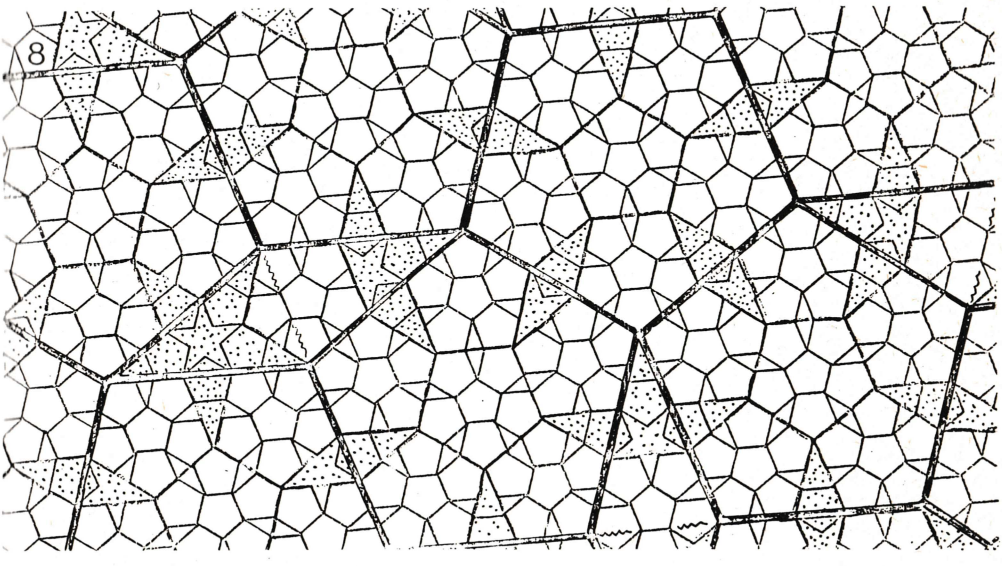
Furthermore, the forced hierarchical nature of this pattern implies that the tiling has a number of very remarkable properties. In the first place, it is necessarily *non-periodic* (i.e. without any period parallelogram). More about this later. Secondly, though the completed pattern is not uniquely determined - for there are $2^{\aleph_0}$ different arrangements - these different arrangements are, in a certain 'finite' sense, all indistinguishable from one another! Thus, no matter how large a finite portion is selected in one such pattern, this finite portion will appear somewhere in *every* other completed pattern (infinitely many times, in fact). Thirdly, there are many unexpected and aesthetically pleasing features that these patterns exhibit (see Fig. 7). For example, there are many regular decagons appearing, which tend to overlap in places. Each decagon is surrounded by a ring of twelve pentagons, and there are larger rings of various kinds also. Note that every straight line segment of the pattern extends outwards indefinitely, to contain an infinite number of other line segments of the figure. The hierarchical arrangement of Fig. 7 is brought out in Fig. 8.

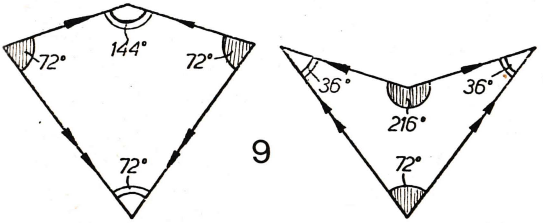
After I had found this set of six tiles that forces non-periodicity, it was pointed out to me (by Simon Kochen) that Raphael Robinson had, a number of years earlier, also found a (quite different) set of six tiles that forces non-periodicity. But it occurred to me that with my tiles one could do better. If, for example, the third 'pentagon' shape is eliminated by being joined at two places to the 'diamond' and at one place to the bottom of the 'boat', then a set of *five* tiles is obtained that forces non-periodicity. It was not hard to reduce this number still further to four. And then, with a little slicing and rejoining, to *two*!

The two tiles so obtained are the 'kites' and 'darts' mentioned at the beginning (*). The precise shapes are illustrated in Fig. 9. The matching rules are also shown, where vertices of the same colour must be placed against one another. There are many alternative ways to colour or shade these tiles to force the correct arrangements. One way which brings out the relation to the pentagon-diamond-boat-star tilings is shown in Fig. 10. A patch of assembled tiles (partly shaded in this way) is shown in Fig. 11 (overleaf).

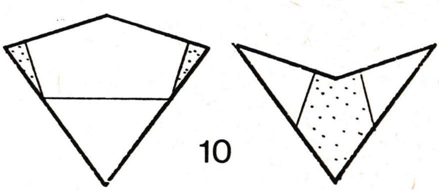
(*) These names were suggested by John Conway.


7


8



9

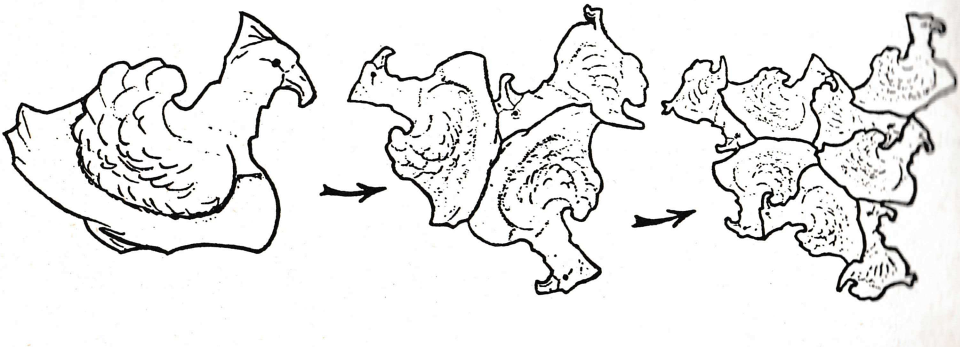

10

The hierarchical nature of the kite-dart tilings can be seen directly, and is illustrated in Fig. 12. Take any such tiling and bisect each dart symmetrically with a straight line segment. The resulting half-darts and kites can then be collected together to make darts and kites on a slightly larger scale: two half-darts and one kite make a large dart; two half-darts and two kites make a large kite. It is not hard to convince oneself that every correctly matched kite-dart tiling is assembled in this way. This 'inflation' property also serves to prove non-periodicity. For suppose there were a period parallelogram. The corresponding inflated kites and darts would also have to have the same period parallelogram. Repeat the inflation process many times, until the size of the resulting inflated kites and darts is greater than that of the supposed period parallelogram. This gives a contradiction.

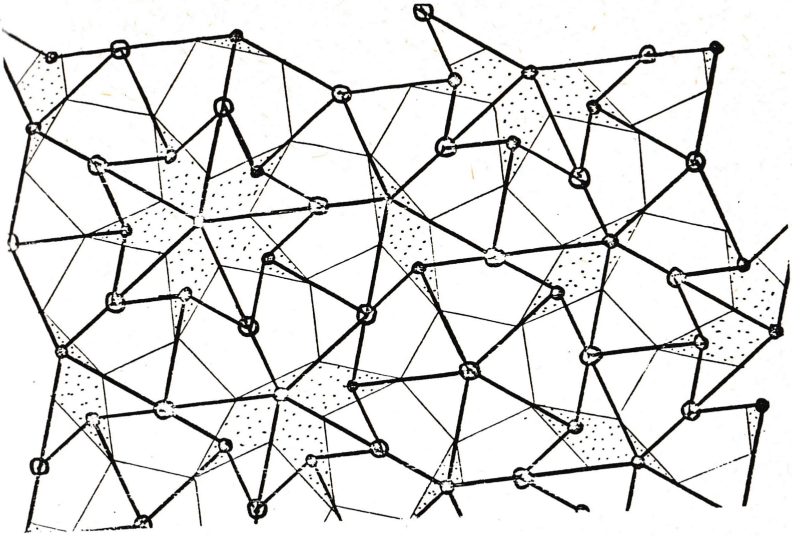
The contradiction with periodicity shows up in another striking way. Consider a very large area containing d darts and k kites, which is obtained referring to the inflation process a large number of times. The larger the area, the closer the ratio $x = k/d$ of kites to darts will be to satisfying the recurrence relation $x = (1+2x)/(1+x)$ (since, on inflation, a dart and two kites make a larger kite, while a dart and a kite make a larger dart). This gives, in the limit of an infinitely large pattern, $x = \frac{1}{2}(1+\sqrt{5}) = \tau$, the golden ratio! Thus we get an <u>irrational</u> relative density (+) of kites to darts - which is impossible for a periodic tiling.

There is another pair of quadrilaterals which, with suitable matching rules, tiles the plane only non-periodically. This is a pair of rhombuses shown in Fig. 13. In Fig. 14 a suitable shading is suggested where similarly shaded edges are to be matched against each other. In Fig. 15, the hierarchical relation to the kites and darts is illustrated. The rhombuses appear mid-way between one kite-dart level and the next inflated kite-dart level.

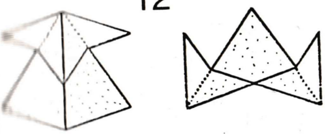
Many different jigsaw puzzle versions of the kite-dart pair or the rhombus pair can evidently be given. One suggestion for modified kites and darts, in the shape of two birds, is illustrated in Fig. 16. The inflation process (in reverse) is illustrated in Fig. 17.



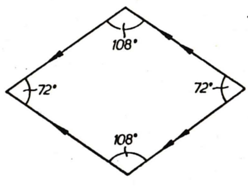
11

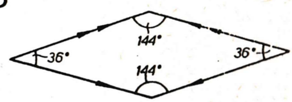


12

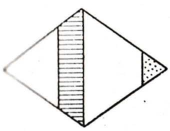


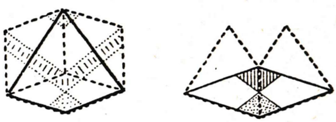
13



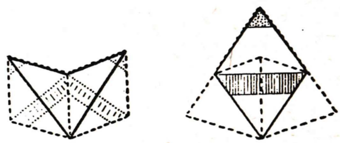


14





15



16

17 →

Other modifications are also possible, such as alternative matching rules, suggested by Robert Ammann (see Fig. 18) which force half the tiles to be inverted.

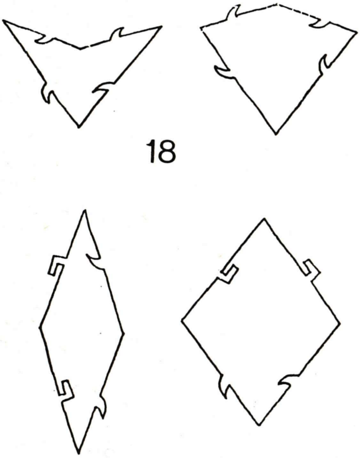
Many intriguing features of the tilings have not been mentioned here, such as the pentagonally-symmetric rings that the stripes of Fig. 14 produce, Conway's classification of 'holes' in kite-dart patterns (i.e. regions surrounded by 'legal' tilings but which cannot themselves be legally filled), Ammann's three-dimensional version of the rhombuses (four solids that apparently fill space only non-periodically), Ammann's and Conway's analysis of 'empires' (the infinite system of partly disconnected tiles whose positions are forced by a given set of tiles).

It is not known whether there is a single shape that can tile the Euclidean plane only non-periodically. For the hyperbolic (Lobachevski) plane a single shape can be provided which, in a certain sense, tiles only non-periodically (see Fig. 19) - but in another sense a periodicity (in one direction only) can occur.

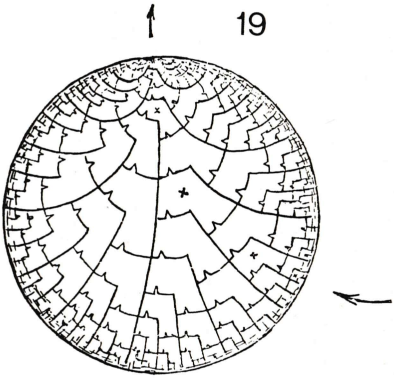
(+) This is the numerical density. The kite has τ times the area of the dart, so the total area covered by kites is τ^2 (= $1+\tau$) times that covered by darts.

References

Gardner, M., Scientific American January 1977, pp. 110-121.
Penrose, R., Bull. Inst. Maths. & its Applns. 10 No. 7/8 (1974) pp. 266-271.



18



19

Little Arrows

by Dr. P. T. Johnstone

One of the salient features of twentieth-century mathematics has been the proliferation of different types of structure which are studied by mathematicians, and the consequent increasing specialization of mathematical research. A notable exception to this trend, in the past thirty years, has been the rise of category theory, which has its origins in the joint work of Samuel Eilenberg and Saunders Mac Lane around 1945 (1). In this article I shall try to describe the aims of category theory, and some of its present areas of application.

Category theory begins with the idea that if you want to study a particular type of mathematical structure, you must study not only the structures of that type, but also the structure-preserving maps between them. Thus if you are interested in topological spaces, it is useless to study the spaces by themselves, without the concept of continuous map which binds them together. Similarly, the study of groups, rings or vector spaces cannot proceed far without the appropriate notion of homomorphism.

A category thus consists of three things: a collection of objects, a collection of arrows or morphisms each of which is associated with a pair of objects called its source and target (we use the diagrammatic notation "f: A → B" for "f is an arrow with source A and target B"), and a composition law which assigns to each "composable pair" of arrows (f: A → B, g: B → C) a third arrow gf: A → C. In addition, we require that composition be associative (i.e. that h(gf)=(hg)f whenever the composites are defined), and that for each object A there should exist an identity arrow 1_A: A → A such that $f1_A = f$ and $1_A g = g$ whenever the composites are defined.

Familiar examples include the category Set of sets and functions, algebraic categories Gp, Rng, Lat, ... of groups, rings, lattices, ... and the appropriate homorphisms, the category Top of topological spaces and continuous maps, the category Ban of Banach spaces and linear contractions, and so on. These are all concrete categories; i.e. their objects are sets with some kind of structure, their arrows are structure-preserving maps, and their composition law is the usual composition of functions. But there are also many categories which are not concrete; for example, we can make any partially ordered set P into a category whose objects are the elements of P, and whose arrows are instances of the order-relation (i.e. there is just one arrow p → q if p ≤ q, and none if p ≰ q). (The composition law in this case is obvious, since there is always at most one arrow with a given source and target.)

The astute reader will have noticed that I used the vague word "collection" rather than the precise word "set" in the definition of a category, and the last paragraph supplies the reason: from Russell's paradox, we know that the collection of all sets cannot itself be a set. The relationship between category theory and set theory is often somewhat strained, precisely because the business of the former is to study such things as the totality of all sets or of all groups, which are inadmissible objects to the latter. There are modifications of standard set theory which provide a "respectable" foundation for category theory (2); but in fact category theory, because of its fundamental nature, can itself be used as a basis for the rest of mathematics, including set theory! We shall return to this point later; for the present, let us merely introduce the term small category (also called a kittygory (3)!) to describe one whose objects and arrows are the elements of some set. A weaker, but equally useful, notion is that of locally small category; we say C is locally small if, for each pair of objects (A,B), the arrows of C with source A and target B are the elements of a set C(A,B). Note that a concrete category is automatically locally small.

We may now take any mathematical concept or theorem which can be expressed solely in terms of objects, arrows and composition, and apply it to any of the categories mentioned earlier. An obvious example is the concept of isomorphism; we say two objects A and B are isomorphic if there exist $f: A \to B$ and $g: B \to A$ such that $gf = 1_A$ and $fg = 1_B$. More interestingly, the notion of cartesian product is a categorical one. Normally we define the cartesian product of two sets by specifying its elements, but we can also define it (at least up to isomorphism) by specifying the arrows into it: a product of two objects A and B is an object A x B together with arrows $p: A \times B \to A$ and $q: A \times B \to B$ such that, for any pair of arrows ($f: C \to A$, $g: C \to B$), there exists a unique $h: C \to A \times B$ with $ph = f$ and $qh = g$.

This definition coincides with the usual one in all the concrete categories mentioned above, but since it doesn't involve elements we can also apply it to non-concrete categories; in a partially ordered set, it coincides with the notion of meet (greatest lower bound). But there is another possibility: by reversing all the arrows in a categorical definition or theorem, we get a new "dual" concept which may be just as useful as the old one. Thus we have the dual notion of coproduct: a coproduct of A and B consists of an object A+B and arrows $A \to A+B$, $B \to A+B$ such that ... As an exercise, try working out what coproducts look like in Set, Top, Gp, Ab (abelian groups) and CRng (commutative rings); the answer in each case is a familiar construction in the appropriate subject.

So far, however, we have not taken to heart the fundamental dictum of category theory: "Always look to the morphisms". The appropriate notion of morphism between categories is called a functor: a functor $C \to D$ consists of two functions sending objects of C to objects of D, arrows of C to arrows of D, and preserving composition and identities.

(Thus (small) categories and functors between them themselves form a category Cat.) This is sometimes called a covariant functor, to distinguish it from the notion of contravariant functor which is similar, but reverses the direction of arrows and the order of composition. Among the simplest examples of (covariant) functors are the forgetful functors Gp $\to$ Set, Top $\to$ Set, etc., which simply forget the additional structure on the objects and arrows of a concrete category. Many familiar constructions in algebra, such as "commutator subgroup": Gp $\to$ Gp, "group of units": Rng $\to$ Gp, and "n×n matrix ring": Rng $\to$ Rng, are in fact functors; although we normally think of them as defined only on objects of the appropriate category, in fact each of them has a natural definition on homomorphisms as well. As an example of a contravariant functor, we give the power-set functor P: Set $\to$ Set, which sends A to the set of all subsets of A and f: A $\to$ B to the map $(B' \mapsto f^{-1}(B'))$.

But there is actually a third level of structure here, since we can also talk about maps between functors. If S and T are two functors from C to D, a natural transformation α: S $\to$ T is a function assigning to each object A of C an arrow α_A: SA $\to$ TA in D, such that for each f: A $\to$ B in C the composites α_B.Sf and Tf.α_A: SA $\to$ TB are equal in D. The best-known example of a natural transformation is the familiar "natural isomorphism" between a finite-dimensional vector space and its double dual; indeed, one could almost say that category theory was invented in order to give a precise meaning to the word "natural" in this context.

To see how it works, let R-Mod be the category of modules over a commutative ring R. For any module A, the set of linear maps A $\to$ R forms a module A*, the dual of A, and the assignment A $\mapsto$ A* is a contravariant functor R-Mod $\to$ R-Mod (since if we are given f: A $\to$ B, composition with f induces a linear map B* $\to$ A*). The composite of this functor with itself, however, is covariant; and there is a natural transformation α from the identity functor $1_{R\text{-}Mod}$ to **, such that $\alpha_A(a)$ is the linear map "evaluate at a": A* $\to$ R. If A is a finitely-generated free module, then α_A is an isomorphism; this is the natural isomorphism we all know and love.

The notion of natural transformation enables us to make the functors from C to D into the objects of a category D^C. An important part of category theory consists in studying these functor categories, particularly those of the form Set^C where C is locally small. The importance of the latter is connected with the existence, for each object A of C, of a functor h^A: C $\to$ Set which sends B to C(A,B) and f: B $\to$ B' to the function "compose with f". (Similarly, we have a contravariant functor h_A: C $\to$ Set which sends B to C(B,A).)

The Yoneda Lemma (4) says that, for any functor T: C $\to$ Set, the arrows $h^A \to$ T in Set^C are in natural 1-1 correspondence with the elements of the set TA; explicitly, a natural transformation α: $h^A \to T$ is completely specified by the element $\alpha_A(1_A)$ of TA. This seemingly innocuous observation has far-reaching consequences. For example, we deduce that arrows $h^A \to h^B$ in $\underline{Set}^{\underline{C}}$ correspond to elements of $h^B(A)$, i.e. to arrows $B \to A$ in $\underline{C}$; thus the assignment $A \mapsto h^A$ is not only a contravariant functor $\underline{C} \to \underline{Set}^{\underline{C}}$, but actually a full embedding (i.e. it is bijective on arrows between a given pair of objects). We say that a functor T: $\underline{C} \to \underline{Set}$ is representable if it is isomorphic to some h^A, and define a representation of T to be a pair (A,a) with A an object of C and $a \in TA$, such that the induced arrow $h^A \to T$ is an isomorphism. Another consequence of the Yoneda lemma is that the representation of T, if it exists, is unique up to canonical isomorphism.

Almost all the important concepts of category theory can be described, at least for locally small categories, as representations of certain functors. For example, a product of A and B is a representation of the (contravariant) functor which sends C to the set of pairs (f: $C \to A$, g: $C \to B$). The contravariant power-set functor is represented by $(2,\{1\})$ where 2 is the two-element set $\{0,1\}$, and the forgetful functor U: $\underline{Gp} \to \underline{Set}$ is represented by (Z,1) where Z is the additive group of integers. More generally, for each set A, the composite $h^A.U$ is represented by (FA, η_A) where FA is the free group generated by A and η_A: $A \to UFA$ is the insertion of the generators.

This last is an example of an adjunction; if F: $\underline{C} \to \underline{D}$ and U: $\underline{D} \to \underline{C}$ are functors, we say F is left adjoint to U if we are given a bijection, natural in A and B, between the sets $\underline{D}(FA,B)$ and $\underline{C}(A,UB)$. A remarkable theorem asserts that every such bijection arises from a pair of natural transformations η: $1_C \to UF$, ε: $FU \to 1_D$ (the unit and counit of the adjunction) such that $\varepsilon_{FA}.F\eta_A = 1_{FA}$ and $U\varepsilon_B.\eta_{UB} = 1_{UB}$ for all A and B; thus the notion or adjunction is an "elementary" one which can be defined without reference to local smallness. A great many key constructions in mathematics turn out to be left or right adjoints to certain familiar functors; we have just mentioned the example of free groups, and the construction of integral group-rings is a functor which is left adjoint to the "group of units" functor mentioned earlier. Again, the categorical notion of product can be described as an adjoint; if each pair of objects of $\underline{C}$ has a product, then the assignment $(A,B) \mapsto A \times B$ defines a functor from the cartesian product $\underline{C} \times \underline{C}$ to $\underline{C}$ which is right adjoint to the "diagonal functor" $A \mapsto (A,A)$.

We say a category $\underline{C}$ is cartesian closed if it has products as above and, for each object A, the functor $(-) \times A$: $\underline{C} \to \underline{C}$ has a right adjoint, the exponential functor $(-)^A$. $\underline{Set}$ is cartesian closed, with $B^A = \underline{Set}(A,B)$, for it is well known that functions $C \to \underline{Set}(A,B)$ correspond to functions $C \times A \to B$. Similarly, $\underline{Cat}$ is cartesian closed, the exponentials being the functor categories defined earlier. The importance of cartesian closedness is that we can regard the exponentials B^A as "internal" equivalents of the hom-sets $\underline{C}(A,B)$, and (by suitably reformulating notions such as representability) we can remove the dependence of what we have done on set-theoretic notions like local smallness - i.e. we can give a truly autonomous development of category theory.

This line of development was pioneered by William Lawvere, who in 1963 (5) gave a number of purely categorical axioms characterizing the categorical properties of Set; more recently, Lawvere and Myles Tierney have introduced the notion of topos, which has been the cause of some of the most exciting categorical developments in recent years. A topos is simply a cartesian closed category which satisfies an "autonomous" version of the representability of the power-set functor (as Lawvere has pointed out, the latter condition for Set is just the categorical version of the set-theorists' Axiom of Comprehension!); to the category-theorist, it serves as a "universe of discourse" within which he is able to carry out categorically all sorts of constructions previously understood only for the category of sets.

The most striking thing about this development is that the concept of topos had been introduced earlier (admittedly with a different definition, which was dependent on set theory) by the French school of algebraic geometers headed by Alexander Grothendieck (6); and indeed examples of toposes arise naturally in algebraic geometry, topology and elsewhere. Thus through topos theory it is possible for the category-theorist to bring ideas of logic (such as the completeness theorem) directly to bear on geometry and algebra, or conversely to apply topological notions such as localization to problems in logic. In this way, it can be seen that category theory is indeed fulfilling its ambition to bring together different areas of mathematical research and expose their fundamental unity.

I close with some suggestions for further reading. (7) is the category-theorist's bible, written by one of the founders of the subject who is also a superb expositor; but it can be rather hard going for the beginning student. (8) is a praiseworthy, and not altogether unsuccessful, attempt to write a category-theory text for those with much less familiarity with abstract mathematics. (9) is a short and readable introduction to the wonders of topos theory.

References

(1) S. Eilenberg and S. Mac Lane, Trans. A.M.S. 58 (1945), 231-294.

(2) e.g. S. Mac Lane, Springer Lecture Notes 106 (1969), 192-200.

(3) Like almost all categorical puns, this is due to Peter Freyd. It comes from his delightful book Abelian Categories (Harper & Row, 1964), now unfortunately out of print.

(4) Attributed to N. Yoneda, J. Fac. Sci. Tokyo 7 (1954), 193-227. In fact the lemma does not appear there explicitly, and Lawvere has pointed out that the idea of the lemma is older than category theory itself - it goes right back to Cayley's representation theorem for groups.

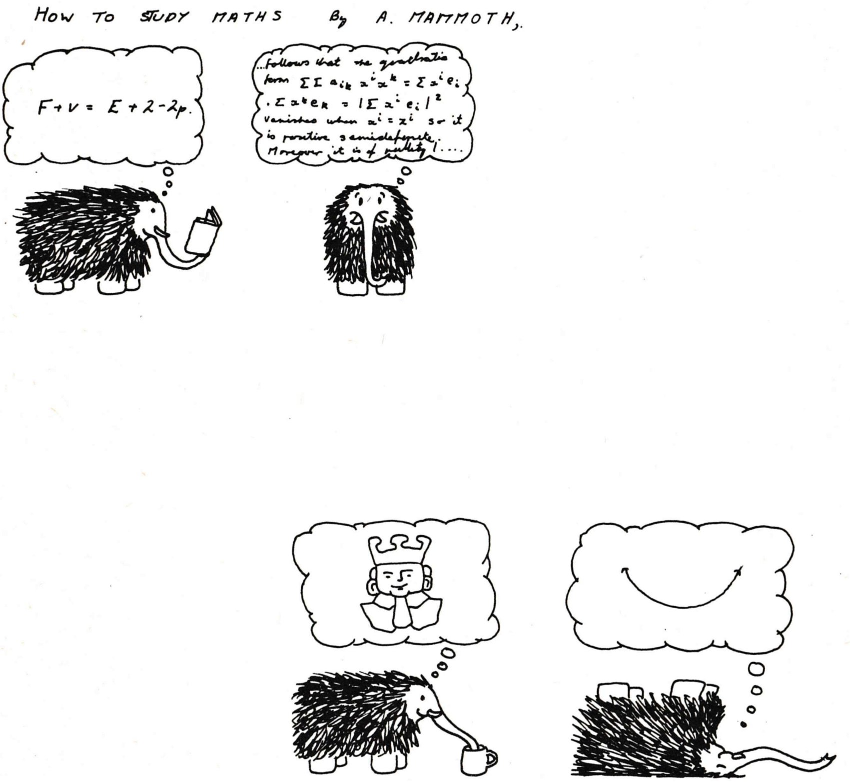
(5) F. W. Lawvere, Proc. Nat. Acad. Sci. 52 (1964), 1506-1511.

(6) J. Giraud, Seminaire Bourbaki 256 (1963).

(7) S. Mac Lane, Categories for the Working Mathematician, Springer Graduate Texts no. 5, 1971.

(8) M. Arbib and E. G. Manes, Arrows, Structures and Functors, Academic Press 1975.

(9) G. C. Wraith, Bull. Inst. Math. and its Applications 12 (1976), 115-119.



Homological Knot Theory

IN THE CONTEXT OF AN EXPANDING ECONOMY

by Professor E. Grimshaw

Section 1: Introduction

In 1973, using the newly developed quasi-topological techniques of Brauer and Suzuki outlined in their ground-breaking paper delivered at the Helsinki colloquium, Kaczinski and Grimshaw proved the long-standing conjecture of Lefschetz on the paranormality of q-reducible semiknots. This had long been the sticking point which had hindered the discovery of any useful theory of Homological Knots. Although it was not realised at the time, the subsequent investigations provided results which were to have applications to a wide range of mathematical disciplines. Perhaps the most surprising use of the theory has been in the solution of economic models for which previously no methods had been available. It is the aim of this note to summarise the advances in this field and to give a new proof of the Atlee-Umagaki Lemma.

Section 2: Preliminaries

The terminology we shall employ is Atlee's [1]. We shall assume the following:

(i) Each pseudo-regular Lindelöf space is paracompact, and this property is distinguished.

(ii) If K is a connected, not necessarily normal topological manifold with the finite cover property, then $H_n(K)$ is trivial for all $n > M(K)$, where $M(K)$ is the Urysohn number of K. [2]

(iii) $H \dagger (K-g)$ paranormal wrt the Schellmann-Zariski topology iff g lies in every q-reducible representation of $K \dagger (K)$ induced by its tensor dual $K \otimes K/(g)$. [3]

(iv) The Slutzky aggregate of any macro-economic model is separable provided the Neumann-Morgenstern equations are satisfied. Since such a model is π-equivalent to the semiknot ideal space $\bar{K}$, which will be assumed paranormal (see [1]), these equations will automatically hold in all our applications. [4]

Section 3: Chief Results

The main people working in the field have been Atlee, F. J. Bingham, Grimshaw, Kaczinski and Forsyth. Following the Kaczinski-Grimshaw paper, Bingham derived an algorithm for q-reducibility of ideal spaces using finite element analysis [5] and in a further series of papers he and Grimshaw extended this to a general semiknot space which for the purpose of this note may be assumed homotopic to its bidual. It was Kaczinski who first noticed that the new methods could be used to tackle a host of other problems (*), notably the solution of the equilibrium state of a macro-economic model

(MEM). Following discussions between him and Grimshaw under the auspices of Heriot-Watt University, an analogous algorithm for MEMs was devised. This algorithm was tested on an IBM-370/145 at Cambridge University, England by Dr. Forsyth and emerged successfully. In fact the model was used subsequently in the predictions of the evolution of the English economy [6]. Basically the results and techniques show that a quasi-static MEM has a finite set of equilibria only if the Slutzky aggregate is double separable. It is the converse implication which forms the content of a brilliant lemma due to Atlee and simultaneously discovered by independent methods by Umagaki [1, 7]. The main interest in the subject now is the extension of their result to systems of MEMs with possibly higher degrees of separability.

Section 4: Proof of the Atlee-Umagaki Lemma

Without loss of generality we may assume the MEM is totally balanced. It may therefore be represented as a semigroup of operators (not necessarily of bounded degree) on a connected topological manifold. As C. Foias has demonstrated in [8], the Slutzky aggregate can be uniquely extended to the compactification of this manifold and hence, provided of course that this compactification is of finite type, that the coefficient of separability remains invariant. But a compact connected manifold is homeomorphic to a pseudo-regular Lindelöf space [9] and hence is paracompact. We now show that the manifold K which we have constructed has the finite cover property. Let $\mathfrak{n}$ be a base of absolutely convex neighbourhoods of K in some localisation of K, say K*. Then, if the maximum filtration of $\mathfrak{n}$ is $\mathfrak{n}^*$, the elements of $\mathfrak{n}^*$ will be a suitably dense cover of K*.

$$K^* = \bigcup_{n \in \mathfrak{n}} (MNn + \varepsilon M(K)K')$$

where M, N are non-convergent integers (in the p-adic topology) and K' and ε are arbitrarily tensor-free. Elementary manipulation of this fundamental equation shows that K* may be expressed as a partially finite disjoint union of reduced elements of $\mathfrak{n}^*$. We may now invoke the Reduced Cover Lemma [8] to show that any partially finite cover on K* is finite; but of course this is exactly what we want. The 2-separability condition may now be substituted in the equation

$$M(K) = \left| S(K) + \frac{3}{8}\sqrt{|S(K)|} \right|$$

where S(K) = -2 using Atlee's convention, so in particular $H_2(K)$ is trivial. Now choose an arbitrary element g contained in an everywhere bounded subset $\mathcal{B}$ of K. What are the absolutely flat neighbourhoods of g in the S-Z topology on (K-g)? Plainly they are a sub-linear set on which a group operation can be defined and using a familiar trick of Umagaki, by considering the duals of these on K, they may be q-realised as distinct (up to homotopy class) generators of $H_1 \dagger (K-g)$, a subgroup (linear hence paranormal) of $H \dagger (K-g)$. But paranormality is a distinguished property so the overgroup is itself paranormal. By (iii), g must lie in every q-reducible representation of $K \dagger K$, and hence of

$H_1 \neq K$. This immediately implies that the absolutely flat neighbourhoods are subdirectly irreducible and as such there can only be finitely many of them. Because the MEM is quasi-static, using Bingham's selection theorem and its associated algorithm, we may set up a bijective map between the approximately dense states of the MEM and the absolutely flat neighbourhoods of the central point g' of K. Since there are only finitely many of them one may infer immediately that they are π-equivalent to equilibria states.

Section 5: Conclusions

The above proof is significantly shorter than that produced by Atlee, who was forced to employ obsolescent transfinite techniques in his proof. It would perhaps be fitting to mention some powerful uses to which this key Lemma can be, and has been, put. Firstly, in Bangladesh the presence of subversive peasant factions in the workforce disrupted the finite set of equilibria states, thus annihilating the 2-separability of the SA. This was manifested by the spectacular drop in world grain prices, which would have occurred but for the intervention of Russian tycoons. Secondly, in Berchampstead, Mrs. J. Spidermouth using a subtle application of the Lemma, succeeded in making the sale of comestibles with less than 31% starch content a viable proposition [10]. Finally, we cite the current instability of the Finnish rural economy as an example of a singly separable Slutzky aggregate situation.

(*) In much the same way as the Golod-Schafarevitch theorem.

References

[1] J. Atlee, Topological Invariants of Manifolds, Duke. Math. J., 1973(2), 571-584.

[2] R. Hiller, Jnr., Topology, Interscience, 1952 (2nd edition).

[3] S. Schanuel, Reducible Representations of the Tree Product, Amer. Math. Monthly, 1970 Jan.

[4] F. M. Slutzky, Moskva Ekonomika Th.

[5] F. J. Bingham, E. Grimshaw, q-Reducibility Criteria in Ideal Spaces, I.E.E.E., 1975, 4-13.

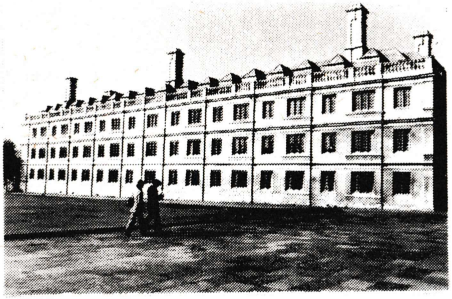
[6] Economic Review for 1976, H. M. Publications.

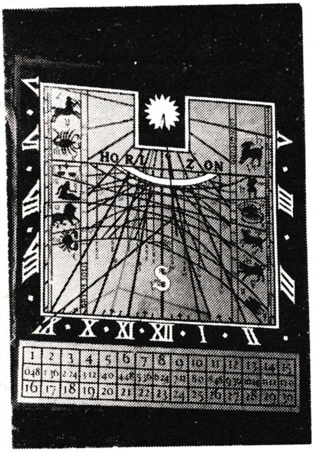
[7] Tsuki Umagaki, A Remark on the Slutzky Aggregate and Equilibria, Tokyo Econ. J., 1974, 841-910.

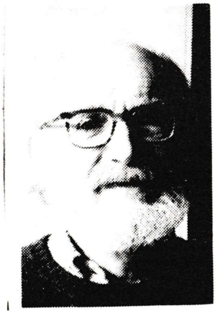
[8] C. Foias, Universal Models and Reduced Covers, Annals of Math., 1954, vol. 60, 345-390.

[9] J. L. Kelley, General Topology, Springer-Verlag, 1955.

[10] Jessica Spidermouth (Mrs.), My Life with Fingal, Penguin, 1977.



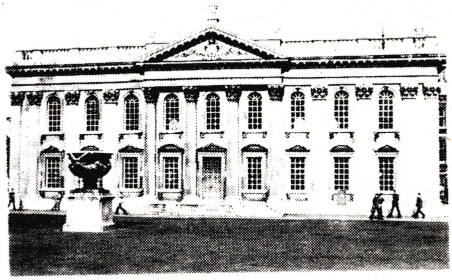


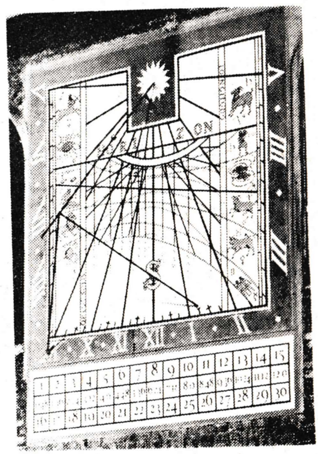


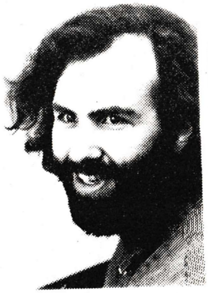
Mathemati[illegible]

From l[illegible] to r[illegible]

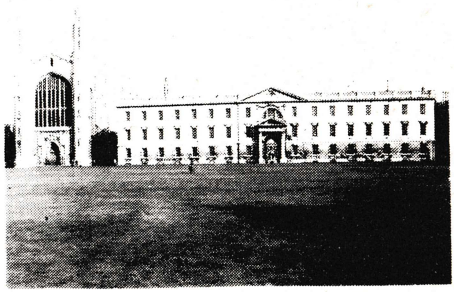
Clare College, Professor [illegible]ill,
Professor Cassels, Dr. [illegible], Se[illegible]
Sundial (Queens'), Profess[illegible]win[illegible]
Hawking, Dr. Reid, Dr. [illegible], D[illegible]
Dr. Friedlander, Dr. Croft, [illegible] Hu[illegible]
Dr. Conway, Dr. Patterson[illegible]
Mathematical Bridge ([illegible]), M[illegible]
P. Verschueren, King's Colle[illegible].







Cambridge



ight:

Professor Polkinghorne,
enate House;
nerton- Dyer, Professor
Dr. Lickorish, Sundial;
uppert, Professor Adams,

M.R. Kipling,

Cambridge Mathematics since 1910

by Professor Sir Harold Jeffreys and Lady Jeffreys

Sir Harold (J) Up to 1909 the Mathematical Tripos was in two parts. The first was normally taken in the third year; by some candidates in the second. The second part was on advanced work and usually had about ten candidates, most of whom got firsts. In Part I the candidates were listed in order of merit; the Senior Wrangler received much newspaper publicity, out of all proportion to his later achievements; in a comparison I made about 1915 the first and second wranglers had exactly the same number of noteworthy performances later in life. In 1909 the top six all did something important later, but the third and the fifth did most. At the other end of the list came the Wooden Spoon, the last man to be classed. His friends presented him with a large spoon, and the last of these treasures, in form something like an oar as the recipient was a rowing man, now hangs in the small Combination Room at St. John's. Teaching for Part I was mainly by coaching. Each coach covered the whole of the Tripos, with special emphasis on the sort of questions likely to be set. There were stories of the last famous coach, R. R. Webb, that he had coached all the Wranglers in a year, and (I think) all but one of the last twenty Senior Wranglers. Webb published little, but his predecessor, E. J. Routh, produced some important papers and several textbooks, better than many later ones.

Lady Jeffreys (J') In about 1965 a pupil told me that she had found a good book on Dynamics in the library by "a man called Routh". Some research into the Girton records shows that in the late 19th century and up to 1914 the teaching was done by a team of lecturers which included N. M. Ferrers, W. M. Hicks, Arthur Berry, A. N. Whitehead, and also W. H. Young, who after his marriage to Grace Chisholm went to live in Göttingen, but returned to teach in Cambridge during term. For a time Constance Herschel, daughter of Sir John Herschel and great-niece of Caroline Herschel, was the resident lecturer in Mathematics and Natural Sciences. Caroline won the Royal Astronomical Society Gold Medal in 1828, and this was later presented to the College by another great-niece, Lady Gordon.

J I came up in 1910, so that the new system had been in operation for four years when I took the Tripos in 1913. In this, Part I was usually taken at the end of the first year and was often used as a preliminary examination to work for other triposes, especially Natural Sciences and Geography. Part I had the peculiarity that anyone failing to be classed could take it again in the second year. Part II was divided into Schedule A and Schedule B, Schedule A more or less corresponding to the old Part I, Schedule B to the old Part II, and both were taken in the same year. St. John's was the only college that had college lecturers covering the whole of Part I and Schedule A; even Trinity sent some people to Baker's Theory of Functions. Otherwise people took lectures, including those for Schedule B, in several different colleges. One consequence of the new arrangement was that more people proceeded to advanced work. It was, however, rather a strain. In fact, those aiming at a star in Schedule B tried to cover nearly all the work for Schedule A in their second year and concentrated on Schedule B in their third, but they had to keep revising all of Schedule A at the same time. The strain was lessened by calling Schedule A Part II in 1934 and Schedule B Part III in 1935, and this system of taking the two parts in different years has persisted. Most of the former coaches became college lecturers in 1909 and these became university lecturers in 1926 in consequence of the Royal Commission of 1919-22. Coaching, so far as it survived, was replaced by college supervision. As it happened, in my year St. John's had four scholars and four exhibitioners, and to a large extent we supervised one another. Baker, as Director of Studies, gave us a standing invitation to go and see him if we were particularly bothered, and we did so about twice a term. The character of both lecturing and supervision has changed. When I lectured for Schedule A lecturers usually set questions and looked over the answers; if questions were found difficult solutions were given in the next lecture. In my time the possible subjects for Schedule B and later Part III were arranged in about 15 groups and spread more or less evenly over six papers. Candidates were not obliged to confine themselves to the subjects they had announced. I remember that I had announced Elliptic Functions, and Dynamics and Hydrodynamics, but I managed to pick up questions in Theory of Functions, Differential Equations, and Celestial Mechanics and Spherical Astronomy. The present arrangement of papers, in closely related subjects, discourages general reading and I do not consider it a good thing.

<u>J'</u> When I came up in 1921 some of the old coaching terminology survived at Girton. We had four coachings a week, in pairs; the coaches included J. C. Burkill, Harold Jeffreys and L. A. Pars. In my first term I went to one course of lectures, by F.J. M. Stratton at Caius, on Optics and Hydrostatics. The subject matter was not enthralling, but I learned from him not to use the word "obviously". The next term was better - S. Pollard at Trinity on Analysis, and in the Easter Term we went to G.P. (later Sir George) Thomson at Corpus, for Electricity and Magnetism. He did experiments in the Corpus lecture room, not very successfully, but at least we were left in no doubt that this was an experimental subject. In later terms I went to more lectures in other colleges, including Arthur Berry's on Elliptic Functions, in King's. He used to stand before the fire with the tail of his gown dangerously near it, and there was

always the prospect that he might catch fire, but I think he never did. I learned a lot from Mary Taylor, who was working with Appleton. Hardy had left Cambridge for a Chair in Oxford, but she lent me her excellent notes of his lectures. Partly through her influence I went to Göttingen for the winter of 1927-8. Until 1933 the Institute for Theoretical Physics shared with Niels Bohr's Institute at Copenhagen the place of highest importance in Quantum Theory. I think my interest in Atomic Physics came in the first place not from Cambridge, but from school physics and from reading a popular book by Sir Oliver Lodge at a friend's house in the vacation before I took Part II.

I spent two terms reading for Physics Part II and proving a very unsuccessful experimenter in the Cavendish Laboratory. In those days if you wanted to do research you went to see the Professor, and Rutherford said "They tell me you aren't much good at experiments, so you had better go and see Fowler". The Ph.D degree was comparatively new and R. H. Fowler supervised the half-dozen or so of us working in Quantum Theory or Statistical Mechanics. There was no regular central meeting place like DAMTP; we worked in college, or the Philosophical Library, or the Cavendish Library, and we went to advanced lectures. There was a weekly colloquium in the Cavendish; I remember one by Dirac, who gave an account of Born's Collision Theory just after the paper came out in 1926. In the Easter Term of 1926 Dirac gave his first course of lectures, attended by about a dozen people, in a lecture room at St. John's. The isolation in which we worked may have been good for us in some ways, by encouraging independence. I thought I missed contacts because I was a woman, but I have since learned that there was very little mixing between the research students from different colleges. There were two evening discussion groups, $\nabla^2 V$ and the Kapitza Club, both now sadly extinct. Women were not eligible for membership of $\nabla^2 V$ until after 1945. At both of these clubs papers were read on widely differing topics, with the hope that workers in different fields might understand one another.

I was away from Cambridge from 1928 until 1938 at Manchester and elsewhere; this was a period when the Tripos was restructured, but its content was not greatly altered. When I returned in 1938 numbers were ~~still~~ compared with those of today, and certainly for Part I lecturers took in written work from their classes. The Archimedeans had been very active in criticising lectures and lecturers, and in 1940 I served on a small liaison committee: there is nothing new under the sun! The truth is that Mathematics is a difficult subject to lecture on-or just a difficult subject.

J The Ph.D degree was started about 1923. Before that, people doing research usually bothered any senior member that they thought likely to be helpful, and in fact most of them were. One consequence of its introduction was that the title "Doctor" carried some prestige away from Cambridge, and this was already held by many D.Sc.'s from other universities. It was said that the introduction of the Ph.D was to attract Americans, for whom it was a necessary condition for an academic post, but in fact this was quite a minor part of the reform. I mostly went to Newall, Eddington, Stratton and occasionally Larmor for advice. Newall was Professor of Astrophysics, but he knew a lot about many things. He lived in Madingley Rise, now the home of the Department of Geodesy and Geophysics. He was very stately (so was Mrs. Newall), especially when driving in his carriage drawn by a beautiful pair of glossy black horses. He did not hold with motor transport and perhaps my allegiance to the bicycle carries on the tradition, in a less elegant way.

At a high table sherry party n couples were present, and much shaking of hands occurred. No one shook his own or his spouse's hand, and no one shook hands with the same person twice. Mr. A asked the other 2n-1 people how many hands they had shaken, and he received a different answer each time. How many hands did Mrs. A shake?

Problems Drive 1977

by J. Mestel

1 Find a positive integer that is multipled by 7 when its last digit is shifted to the front.
Indicate how to construct such an integer larger than any given integer.

2 Dr. Spock was making patterns with his dominoes ($[i,j]$, $0 \leq i \leq 6$, $0 \leq j \leq 6$) when Captain Kirk sat on his ray gun and fused them into the following array.
Draw in the erased boundaries.

5	4	2	4	1	1	5
4	6	3	2	3	6	3
0	0	1	2	0	6	4
3	4	3	0	1	0	4
1	5	5	5	6	4	2
3	3	4	6	1	0	5
2	6	2	1	5	3	6
5	0	2	1	2	0	6

3 The combination to the secret safe behind the washbasin in the outside toilet consists of seven different integers in descending order, whose sum is 36. A burglar knew the first two numbers, and tried to bribe a porter into telling him the value and position of another.
"You'll have to pay me for two, sir", said the porter, "since no matter which one I gave you, you wouldn't be able to deduce the exact code".
Technically he was correct, but he did not get any money.
(The porter is aware of the burglar's knowledge).

4 In the following sums each letter represents a unique digit. Furthermore no two different letters stand for the same digit.
ONE + ONE + ONE + ONE ⇒ FOUR, FOUR + ONE = FIVE,
TWO - ONE - ONE = O
What is NOW FURTIVE?

5 The audience at a Natural Sciences lecture on rag day consists of various numbers of spiders, starfish, tapeworms and relations of Cyclops, a solitary peacock and, of course, Macbeth's corpse. The first lecturer of the day observes as many arms as legs in the audience, and so assumes all is as usual. The second lecturer compares arms and eyes, with a similar result. The third lecturer notes twice as many eyes as heads, but is surprised by how many of them are open.
Describe the constitution of the least possible audience.

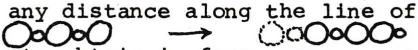
Biological Data

	heads	eyes	arms	legs
spider	1	2	0	8
starfish	0	0	5	0
tapeworm	1	0	0	0
Cyclops	1	1	2	2
Macbeth	0	0	2	2
peacock	1	1002	0	2

6 Three 2p coins and two 1p coins are arranged thus:

A move consists of translating a 2p coin and a touching 1p coin any distance along the line of their centres.

e.g.:

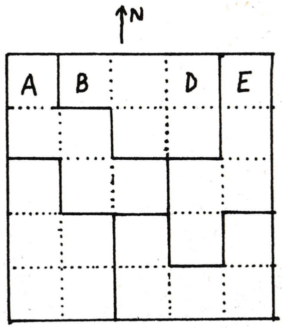


Show how to obtain in four moves a position of the form:

7 AND HERE IT IS ... the inevitable what's the next number in the series problem. Explanations of solutions are advisable.

(a) 0, 1, -1, 2, -1, 5, -4, 29, ...
(b) 6, 11, 37, 135, 2059, ...
(c) 1, 2, 9, 12, 70, 89, 97, 102, 112, 182, ...

8 Following Cambridge University's UDI the 25 colleges (unit squares in the diagram) were each occupied by one of five peace-keeping forces, the Apaches, Bootboys, Cowgirls, Deviants and Extras. This was done in such a way that each N-S strip, each E-W strip and every local parish (in black outlines) contained a college occupied by each force. Given the information in the diagram, colour Cambridge using five colours without reference to any controversial theorems.

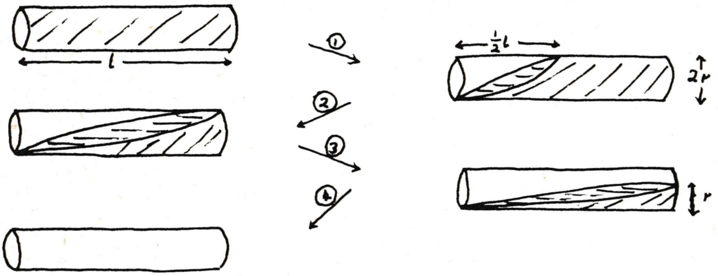


9 The hands on my alarm clock are indistinguishable, and there are no numbers around the outside. Accidentally woken up by it one morning, I observed with a snarl that the hands were both pointing at minute divisions, and that they were 9 minutes apart.
Had it not been for my hangover, what could I have deduced?

10 The C.I.A. have developed a new technique to destroy their secret book-keeping. Unfortunately for the free world, it merely obscures the digits, and does not remove decimal points from a calculation. As an aspiring Commy, can you fill in the blurred digits in the exact long division below?

```
          @@@@.@@@@
     @@@) @@@@@@
          @@@
           @@@
           @@@
            @@@
            @@@
             @@@
             @@@
              @@@@
              @@@@
              ----
```

11 Four impoverished students have a solitary, full, cylindrical glass of beer. They share it out according to the following measurements:



What proportion does each get of the total volume?

MAGDALENE

Ramsey Problems in Euclidean Geometry

by Dr. B. Bollobás

Most of you are probably familiar with Ramsey's theorem, proved by the Cambridge logician F. P. Ramsey [2] in 1930: any colouring of the pairs of the natural numbers with two colours contains a monochromatic infinite set, i.e. an infinite set of natural numbers, all of whose pairs have the same colour. For over twenty years this result seemed to be no more than a curiosity. In the fifties, however, it became increasingly clear (due mostly to the efforts of Professor Paul Erdös) that there are many similar results and problems under rather different conditions.

In the early sixties a branch of set theory was born, called the <u>partition calculus</u>, which aims at answering the question: for which cardinals n, m, r and c is the following statement true?

"Let X be a set with $|X|=m$ and $X(r)$ the set of subsets of X with cardinality r. If we colour $X(r)$ with c colours, then we can find $Y \subsetneq X$ with $|Y|=m$ and all elements of $Y(r)$ having the same colour".

In a variant of this problem we restrict our attention to colourings which are 'regular' in some sense. Thus if we colour $2^{\mathbb{N}} = P(\mathbb{N})$, the set of all subsets of $\mathbb{N}$, with two colours, then there need not be an infinite set $M \subseteq \mathbb{N}$, all of whose infinite subsets have the same colour; but if one of the colour classes is open (in the product topology on $2^{\mathbb{N}}$, the product of countably many 2-point discrete spaces), then there has to be such an M. (This result and some extensions of it are very useful in analysis.)

The nature of the problem changes again if we colour an algebraic or geometric object and look for a monochromatic set with a given structure. A result of this kind was proved by van der Waerden [3] three years before Ramsey's theorem: given n there exists an N such that if $\{1,2,\ldots,N\}$ is coloured with two colours then at least one colour class contains an arithmetic progression of length n. Recently several deep results were proved in this vein, including extensions of van der Waerden's theorem to commutative semigroups.

The aim of this article is to draw attention to some Ramsey type problems discussed in [1], which, though not very deep, are rather amusing, have not been investigated too much and can be tackled by first year undergraduates with a fair chance of success. (See [1] for many more results and problems.)

Let L be a <u>finite</u> set of points in $\mathbb{R}^m$, m-dimensional

Euclidean space. We shall be interested in subsets X, of another Euclidean space $\mathbb{R}^n$, for which in every colouring of X with k colours, there is a monochromatic set L', congruent (in the usual geometrical sense) to L. For brevity, we shall denote this property by $L < (X)_k$. (Some authors prefer to write $X \to (L)_k$.) To gain familiarity with this definition, the reader is advised to prove the following easy lemma.

<u>Lemma 0</u> If $L_2 \subseteq L_1$, $X_1 \subseteq X_2$ and $L_1 < (X_1)_k$ then $L_2 < (X_2)_k$

We say L is <u>Ramsey</u> if for every (finite) k there is an $n = n(k)$ such that $L < (\mathbb{R}^n)_k$. In deciding whether a set L satisfies $L < (\mathbb{R}^n)_k$, and hence whether or not L is Ramsey, it may seem necessary at first glance to consider colourings of the whole of $\mathbb{R}^n$ and not just some finite subset of it. A standard compactness argument, however, shows that this is not the case.

<u>Theorem 1</u> $L < (\mathbb{R}^n)_k$ iff there is a finite set $X \subseteq \mathbb{R}^n$ such that $L < (X)_k$.

<u>Proof</u> By Lemma 0, $L < (X)_k$ immediately implies $L < (\mathbb{R}^n)_k$. To prove the other implication we need some preparation. For simplicity we put $M = \mathbb{R}^n$ and $[1,k] = \{1,2,\ldots,k\}$. We want to find a 'natural' set whose elements correspond exactly to the distinct colourings of M with k colours: $\{1,2,\ldots,k\}$. We take the set S with $|M|$ coordinates, labelled by the points <u>x</u> of M, where the <u>x</u>th coordinate takes values lying in $[1,k]$. The set we have constructed is nothing but $[1,k]^M$, the product of $|M|$ copies of $[1,k]$. Clearly, given any colouring of M, we can uniquely define an element of S as the one which has the colour of <u>x</u> as <u>x</u>th coordinate; and conversely any element of S determines a colouring of M in the obvious way. So we have obtained a natural identification between $[1,k]^M$ and the set of colourings of M with k colours.

We can put a topology on $[1,k]$, namely the discrete topology, and we can then put the product toplogy on $[1,k]^M$. Tychonov's theorem from general toplogy tells us that $[1,k]^M$ is compact since it is the product of compact spaces.

Given a subset X of M, we define a <u>bad</u> colouring for X to be a colouring of M in which X does not contain a monochromatic L' congruent to L, and we let Bad(X) be the set of bad colourings for X. Thus Bad(X) is a subset of $[1,k]^M$ and if X is <u>finite</u>, it is clear that Bad(X) is both closed and open: for being a bad colouring only places restrictions on the (finitely many) coordinates corresponding to X.

We are now ready to prove the second implication. Suppose $L \not< (X)_k$ for every finite set $X \subseteq M$, that is $\mathrm{Bad}(X) \neq \emptyset$ for every finite set $X \subseteq M$. A moment's thought shows that $\mathrm{Bad}(X) \cap \mathrm{Bad}(Y) \supseteq \mathrm{Bad}(X \cup Y)$ for any subsets X,Y of M. Hence the system $\{\mathrm{Bad}(X)\colon X \text{ finite}, X \subseteq M\}$ of closed sets has the finite intersection property. The compactness of $[1,k]^M$ implies that $\bigcap \mathrm{Bad}(X) \neq \emptyset$, that is we can find a colouring $c \in \bigcap \mathrm{Bad}(X)$ which is bad for <u>all</u> finite subsets of

M and hence clearly bad for M.

Let us see what we can say about the simplest non-trivial geometrical configuration.

Theorem 2 In any colouring of $\mathbb{R}^2$ with 3 colours we can find a pair of points distance 1 apart, with the same colour. However, if $\mathbb{R}^2$ is coloured with 7 colours we cannot necessarily find such a pair of points.

Proof Let P be a pair of points distance 1 apart. In our notation, the theorem becomes

$$P < (\mathbb{R}^2)_3, \text{ but } P \not< (\mathbb{R}^2)_7.$$

Figure 1 shows the first assertion. For suppose that in a red-blue-yellow colouring of the seven points there is no monochromatic adjacent pair. We may assume that x is red. Then y_1, z_1 are blue and yellow (in some order) and so x_1 is red. Similarly x_2 is red, but x_1 and x_2 are adjacent.

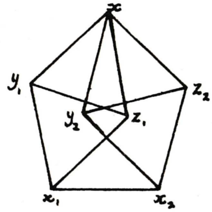


Figure 1. Adjacent points are at distance 1.

To show $P \not< (\mathbb{R}^2)_7$, we merely have to exhibit a colouring of $\mathbb{R}^2$ with 7 colours in which in each monochromatic component the distance between any two points can never be 1. For this purpose we tesselate the plane with regular hexagons with side a, where $\frac{1}{2} < a < \frac{4\sqrt{5}-5}{10}$ and colour them as in Figure 2.

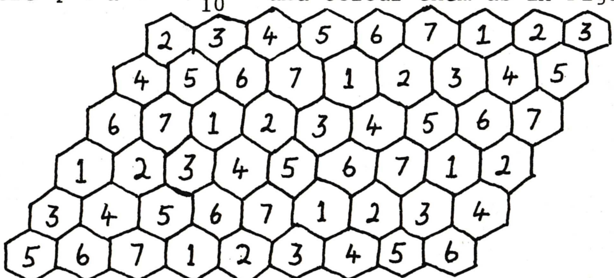


Figure 2, which shows that $P < (\mathbb{R}^2)_7$ is false.

Problem 1 Determine the maximal value of k for which

$$P < (\mathbb{R}^2)_k .$$

Theorem 3 Let Q_2 be the set of vertices of a unit square. Then

$$Q_2 < (\mathbb{R}^6)_2 .$$

Proof We must find a finite set $X \subseteq \mathbb{R}^6$ for which $Q_2 < (X)_2$. In fact for X we take the 15 points:

$x_{12} = (1/\sqrt{2},1/\sqrt{2},0,0,0,0)$, $x_{13} = (1/\sqrt{2},0,1/\sqrt{2},0,0,0)$, ...
$x_{56} = (0,0,0,0,1/\sqrt{2},1/\sqrt{2})$, and in general

$$x_{ij} = (x^1_{ij},\ldots,x^6_{ij})$$

$(1 \leqslant i < j \leqslant 6)$ is defined by $x^k_{ij} = 1/\sqrt{2}$ $(k = i \text{ or } j)$; and $x^k_{ij} = 0$ $(k \neq i,j)$. It is easily verified that the points $x_{ab}, x_{bc}, x_{cd}, x_{da}$ form a unit square (a, b, c, d distinct). Now suppose we have a red-blue colouring of $\mathbb{R}^6$. We now reduce our problem to a combinatorial one: let K^6 be the complete graph of order 6, that is a graph with 6 vertices $\{v_1,\ldots,v_6\}$ say, where every pair of vertices $\{v_i,v_j\}$ is joined by an edge v_iv_j. We next define a red-blue colouring of K^6 by colouring the edge v_iv_j $(i < j)$ with the same colour as x_{ij}. It is left as an exercise to the reader to check that every colouring of the edges of a K^6 with two colours contains a monochromatic quadrilateral. (This can be done by a case by case discussion.) Suppose $v_av_b, v_bv_c, v_cv_d, v_dv_a$ is such a quadrilateral. Then $x_{ab}, x_{bc}, x_{cd}, x_{da}$ form a monochromatic square.

Given $L_1 \subseteq \mathbb{R}^m$, $L_2 \subseteq \mathbb{R}^n$, their <u>cartesian product</u> $L_1 \times L_2$ lies in $\mathbb{R}^m \times \mathbb{R}^n \cong \mathbb{R}^{m+n}$.

<u>Theorem 4</u> If L_1 and L_2 are Ramsey then so is $L_1 \times L_2$.

<u>Proof</u> Since L_1 is Ramsey, given k there is an m for which $L_1 < (\mathbb{R}^m)_k$. Hence by Theorem 1 there is a <u>finite</u> set $X \subseteq \mathbb{R}^m$ such that $L_1 < (X)_k$, where the k colours are $1,2,\ldots,k$, say. We now define a set of new colours: we let every distinct k-colouring of X be a new colour, so that we obtain $c = k^{|X|}$ new colours. Since L_2 is Ramsey, there is an n such that $L_2 < (\mathbb{R}^n)_c$.

We claim $L_1 \times L_2 < (\mathbb{R}^{m+n})_k$. By Lemma 0, it suffices to show $L_1 \times L_2 < (X \times \mathbb{R}^n)_k$. Suppose then that we are given a k-colouring of $X \times \mathbb{R}^n$. This automatically determines a c-colouring of $\mathbb{R}^n$ as follows: any point $\underline{y}$ of $\mathbb{R}^n$ is simply given the new colour corresponding to the k-colouring of X induced by that of $X \times \{\underline{y}\}$. Since $L_2 < (\mathbb{R}^n)_c$, we can find a monochromatic $L_2' \subseteq \mathbb{R}^n$ congruent to L_2; that is, whatever point $\underline{y}$ of L_2' we choose, $X \times \{\underline{y}\}$ induces the same colouring on X. For this colouring of X, we can find a monochromatic $L_1' \subseteq X$ congruent to L_1, since $L_1 < (X)_k$. Clearly it follows that $L_1' \times L_2'$ is a monochromatic subset of $\mathbb{R}^{m+n}$ congruent to $L_1 \times L_2$. Hence $L_1 \times L_2$ is Ramsey.

We now generalise the two examples we have looked at, that is P and Q_2, and define the notion of a <u>brick</u>. A <u>brick</u> in $\mathbb{R}^n$ is a set congruent to $B = (\varepsilon_1 a_1,\varepsilon_2 a_2,\ldots,\varepsilon_n a_n)$: $\varepsilon_i = 0$ or 1 where every $a_i > 0$. Thus a brick is just the cartesian product of a finite number of different sized point-pairs (for example $P \times P = Q$). A unit simplex in $\mathbb{R}^k$ (that is, in $\mathbb{R}^2$ a triangle with unit sides, in $\mathbb{R}^3$ a tetrahedron with unit sides, etc.) shows that $\{0,1\} < (\mathbb{R}^k)_k$ so every point-pair is Ramsey. So by Theorem 4 and Lemma 0 we immediately have

<u>Theorem 5</u> Any subset of a brick is Ramsey.
Call a set $L \subseteq \mathbb{R}^m$ <u>spherical</u> if it is embeddable in some sphere (of arbitrary dimension and radius). It is somewhat more complicated to show that every Ramsey set is spherical. It is intriguing that Theorems 4 and 5 say all that is known about Ramsey sets.

<u>Problem 2</u> Is there a spherical set which is not Ramsey?

<u>Problem 3</u> Is there a Ramsey set which is not a subset of a brick?

Since an obtuse-angled triangle is spherical but not a subset of a brick, in particular we have the following very simple looking problem.

<u>Problem 4</u> Is every triangle Ramsey?

References

[1] P. Erdös, R. L. Graham, P. Montgomery, B. L. Rothschild, J. Spencer and E. G. Straus, Euclidean Ramsey Theorems, I., J. Combinatorial Theory (A) 14 (1973) 341-363; II and III, Infinite and Finite Sets, North Holland 1975, pp. 529-557 and 559-583.

[2] F. P. Ramsey, On a problem of formal logic, Proc. London Math. Soc. 30(2) (1930) 264-286.

[3] van der Waerden, Ein Satz über Klasseneineilung von endlichen Mengen, Abh. Math. Sem. Hamburg Univ. 5 (1927) 185-188.

Algernon and Basil had been poring over some problems in an old copy of Eureka. "Golly", said Basil, "Look what I've found here, Algie. One of those dreadful problems which ask you to find the next term in the series." "Oh, I can never be bothered with those", replied Algernon, "I have my own foolproof way of handling that sort of question. I simply take a polynomial, possibly complex, of smallest degree which has the given terms of the series as its values at the first however many positive integers and then I use it to calculate the rest of the series." Basil appeared puzzled. "I don't know if I'm being an utter oaf, Algie, or whether you're being dashed clever, but I can't for the life of me see firstly why all the subsequent terms worked out according to your rule should be integers, and secondly why the series you get should be unique". Algernon smiled. "Bad luck, old chap".
Who was right - Algernon or Basil?

The Thirteen Spheres Problem

by A. J. Wassermann

In an unpublished notebook at Christ Church, Oxford there is an account of a conversation held in 1694 between David Gregory and Isaac Newton about the distribution of stars of various magnitudes. The question arose as to whether a solid sphere could simultaneously be brought into contact with thirteen other spheres of the same size: Gregory believed that it was possible, but Newton disagreed. It took 180 years before R. Hoppe proved that Newton was in fact right.

The problem we wish to solve is the following: how many spheres of unit radius can simultaneously touch a fixed sphere of unit radius? Clearly this is equivalent to determining the maximum number of points, N say, which can be placed on the surface of a unit sphere which are separated from each other by a distance of at least 1 (that is, great circle distance $\pi/3$). (See Figure 1). It is an easy matter to show that N=12 or 13. For on the one hand if we place 12 points on the surface of the sphere so that they form the vertices of a regular icosahedron, it can be checked that the minimum distance constraint is not violated; and on the other hand, given the N points lying on the sphere, if we draw a circle round each of great circle radius $\pi/6$, our condition implies that they do not intersect. But the area of each such circle is $(2-\sqrt{3})\pi$, so we have $N(2-\sqrt{3})\pi \leq 4\pi$ and hence $N \leq 8+4\sqrt{3}$, which forces $N \leq 13$.

The solution of the problem we give is due to Leech [1] and is a variation on a method used by van der Waerden and Schütte [3], which we will briefly discuss later. Basically the idea of the proof is to consider areas in some detail and hence to reduce the problem to an elementary question in graph theory. Before we embark on the proof, however, we recall some results about spherical triangles. If ΔABC is a spherical triangle on a unit sphere, with all edges great circles, then

(i) $\dfrac{\sin a}{\sin \alpha} = \dfrac{\sin b}{\sin \beta} = \dfrac{\sin c}{\sin \gamma}$ (Sine Rule)

(ii) $\cos c = \cos a.\cos b + \sin a.\sin b.\ \sin \gamma$ (Cosine Rule)

(iii) Spherical area of $\Delta ABC = \alpha + \beta + \gamma - \pi$.

(See Figure 2).

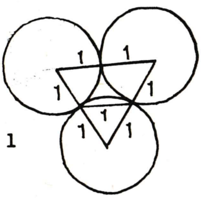


Figure 1

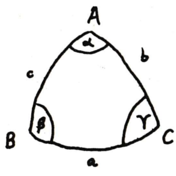


Figure 2

(Note that as ΔABC grows smaller, it approximates more closely to a planar triangle and (i) and (ii) become the familiar (planar) sine and cosine rules.)

Our proof proceeds by way of contradiction, that is we assume Gregory was right and N=13. In the following all the spherical distances are great circle distances and a (spherical) line joining two points on our sphere is a shortest great circle arc joining them (ambiguity can occur only when the points are antipodal). We start by constructing a <u>graph</u> on the surface of the sphere by taking the thirteen given points as vertices and joining any two vertices whose distance apart is strictly less than $\cos^{-1}(1/7)$. We observe that no edges of the graph cross, that is have a common interior point; to show this it suffices to show that a quadrilateral on the sphere with sides of length $\pi/3$ must have a diagonal longer than $\pi/2$ ($>\cos^{-1}(1/7)$). Since the distance between any two points inside the quadrilateral is always less than the length of the longer diagonal and since we can always construct an equilateral quadrilateral of side $\pi/3$ inside the quadrilateral, it is clear that we may assume that it is actually equilateral of side $\pi/3$. But then it is clear that the length of the larger diagonal is least when such a quadrilateral is regular; in which case, by the cosine rule, the diagonal distance is $\pi/2$.

The next step is to simplify the graph so that it divides the sphere into polygons. We employ 2 strategies:

1. Each vertex may be moved so that it is joined to at least 2 other vertices.
2. We may ensure that the graph is connected, since if the graph has more than one component we can slide one of them across the surface of the sphere until one of its points is within $\cos^{-1}(1/7)$ of some point of another component, thus reducing the number of components.

Each angle of the polygons so formed is strictly greater than $\pi/3$: one merely needs to consider spherical triangles ΔABC with sides a,b,c where $a,b,c \geq \pi/3$, $b,c \leq \cos^{-1}(1/7)$ and minimise $\angle BAC$. The minimum (which is unattained in the graph) occurs when $a = \pi/3$, $b = \pi/3$, $c = \cos^{-1}(1/7)$ and $\alpha = \pi/3$, $\beta = \pi/3$, $\gamma = \cos^{-1}(-1/7)$. Hence at most <u>five</u> edges meet in any one vertex, that is each vertex has <u>valency</u> ≤ 5.

We now show, by considering areas, that all the polygons, except for possibly one quadrilateral, must be triangles. First of all we determine when the areas of the polygons are minimised:

<u>Triangle</u> equilateral of side $\pi/3$, with angles of $\cos^{-1}(1/3)$ and area $=\Delta = 3\cos^{-1}(1/3) - \pi \geq 0.5513$ by (iii).

<u>Quadrilateral</u> equilateral of side $\pi/3$, one of diagonals of length $\cos^{-1}(1/7)$ and area $= 2(\cos^{-1}(1/7) + 2/3\pi-\pi)$ ≥ 1.334 by (iii).

<u>Pentagon</u> equilateral of side $\pi/3$, with two diagonals emanating from the same vertex of length $\cos^{-1}(1/7)$ and area ≥ 2.226 by (iii).

Thus if we define w_n = (least possible area of an n-gon)-(n-2)Δ we have $w_4 \geq 0.231$ and $w_5 \geq 0.572$. It is also clear that w_n increases with n, for given an (n+1)-gon of minimal area we may remove a triangle formed by three consecutive vertices to obtain an n-gon: shortening the new side of the n-gon if necessary so that it is less than $\cos^{-1}(1/7)$ can only decrease the area and so the result follows by an easy induction argument. Thus $w_n \geq 0.572$ for $n \geq 5$.

Now let V, E, F be the number of vertices, edges and faces (=polygons) in the graph. By Euler's formula we have $V + F = E - 2$. Let F_n = number of n-gons in the graph. Then $F = F_3 + F_4 + \ldots$ and $2E = 3F_3 + 4F_4 + \ldots$ Euler's formula now gives $2V-4 = F_3 + 2F_4 + 3F_5 + \ldots$

Furthermore the area of the sphere is equal to the sum of the areas of the polygons in the graph, so

$$4\pi \geq \sum_{n\geq 3} (w_n + (n-2)\Delta)F_n = (2V-4)\Delta + (w_3F_3 + w_4F_4 + \ldots)$$

$$\geq 22 \times 0.5513 + 0.231F_4 + 0.572(w_5 + \ldots)$$

So $0.231F_4 + 0.572(F_5 + \ldots) \leq 0.438$, and hence $F_4 = 0$ or 1 and $F_5 = F_6 = \ldots = 0$. So all the polygons are triangles except for possibly one quadrilateral.

Case 1: All the polygons are triangles We have $F_3 = 2/3\,E$ and by Euler's formula $13 + 2/3E = E + 2$. Hence $E = 33$ and the average number of edges at each vertex is $66/13 > 5$, a contradiction.

Case 2: One polygon is a quadrilateral By Euler's formula we obtain $F_3 = 20$, $E = 32$ and all vertices have valency five except for one which has valency four. This configuration, however, can never arise on the surface of a sphere for topological reasons: to verify this one should sit down with pencil and paper for an hour or two and just try to construct such a system (the proof is elementary, but rather long and tedious). This contradiction shows that Newton was right and completes the proof.

The proof of van der Waerden and Schütte ([2] , [3]) also uses graph theory. They consider the smallest sphere (of radius r say) on which 13 points can lie separated by an (actual) distance of at least 1. Their graph is formed by joining all points which are separated by a distance of exactly 1 - as before edges do not cross. However, instead of simplifying the graph, they split it up. To be more precise they define a point to be removable if it can be displaced a small distance so that its minimum distance from all other vertices is strictly greater than 1. The graph can thus be transformed so that it is the disjoint union of isolated points and irreducible (connected) components: plainly not all vertices can be isolated by the minimality of the sphere. Assuming that $r \leq 1$, they show that the graph consists only of one irreducible component, containing only triangles, quadrilaterals and pentagons all with interior angles less than 180°. Furthermore the graph can only contain one pentagon or one regular quadrilateral and not both together. A contradiction is then produced by considering excess areas as in our proof.

The thirteen spheres problem has been much generalised - it leads naturally into such subjects as the packing of spheres. More specifically, however, one can ask how many spheres of the same size can simultaneously touch a unit sphere, if one allows the size of the smaller spheres to vary; equivalently, let $N(\phi)$ be the largest number of points on the unit sphere which are separated by a spherical distance of at least ϕ - can $N(\phi)$ be determined? In fact, $N(\phi)$ has been determined almost completely for $\phi \geq 25^{\circ}$ and some useful upper and lower bounds due to Fejes Toth and van der Waerden are known (for a full discussion see [4]). One would also like to know in what configurations the points must lie for those critical values of ϕ for which $N(\phi)$ 'jumps' between integers. In connection with these jumps, it is interesting to note that if 5(or 11) spheres of equal size can simultaneously touch a unit sphere, then so can 6 (or 12) spheres of the same radius.

References

[1] J. Leech, The problem of the thirteen spheres, Math. Gazette 40:22-23 (1956).

[2] K. Schütte and B. L. van der Waerden, Auf welcher Kugel haben 5, 6, 7, 8 oder 9 Punkte mit Mindesabstand Eins Platz? Math. Ann. 123:96-124 (1951).

[3] K. Schütte and B. L. van der Waerden, Das Problem der dreizehn Kugeln, Math. Ann. 125:325-334 (1953)

[4] H. S. M. Coxeter, The problem of packing a number of nonoverlapping circles on a sphere, N.Y. Acad. Sci. 1961: 320-331.

"Our new postman is not very efficient. On each of Monday and Tuesday there was one letter addressed to each house in the street, and indeed the postman delivered one of them to each house, but on Monday Alf was ... Bert, while on Tuesday Charlie was ... Dave."

In this extract from a letter from my friend Ed, I have replaced by dots two passages, each consisting of "the person who got the letter addressed to" written several times, $1\frac{1}{2}$ times as many in one passage as in the other. I know the number of houses in Ed's street, so I could deduce that the postman delivered all the letters incorrectly on just one of these days.

How many houses are there in Ed's street?

(It may be assumed that the people mentioned by name live in different houses in the street.)

Magnetic Monopoles

by Dr. P. Goddard

From time to time reports that magnetic monopoles have been discovered reach the front page of the newspapers. In 1975 it was claimed that the track of a magnetic monopole had been seen in observations of cosmic rays using balloon borne detectors over Sioux City, Iowa. However it is very difficult to distinguish a magnetic monopole from a heavy nucleus using such apparatus and this claim is now generally discounted. Although experimental evidence for their existence is lacking, magnetic monopoles are objects of considerable theoretical interest and this has been further enhanced in the last four or five years.

Maxwell's equations contain a tantalising asymmetry between electricity and magnetism. Whereas the electric charge and electric current appear in the equations ($\epsilon_0 = c = 1$),

$$\underline{\nabla}.\underline{E} = \rho, \quad \underline{\nabla} \wedge \underline{B} - \dot{\underline{E}} = \underline{j}, \tag{1}$$

magnetic charge and magnetic current are explicitly excluded by the equations

$$\underline{\nabla}.\underline{B} = 0, \quad \underline{\nabla} \wedge \underline{E} + \dot{\underline{B}} = 0 . \tag{2}$$

In vacuo, the equations are symmetric under the transformation $\underline{E} \to \underline{B}$, $\underline{B} \to -\underline{E}$. This symmetry could be restored when charges are present if we introduce a hypothetical magnetic charge density, σ, and magnetic current density, $\underline{k}$, replacing equations (2) by

$$\underline{\nabla}.\underline{B} = \sigma, \quad \underline{\nabla} \wedge \underline{E} + \dot{\underline{B}} = -\underline{k} , \tag{3}$$

and adding $(\rho,\underline{j}) \to (\sigma,\underline{k})$, $(\sigma,\underline{k}) \to (-\rho,-\underline{j})$ to the transformation law.

One cannot object to this modification (i.e. the replacement of equations (2) by equations (3)) of classical electrodynamics on the grounds of lack of consistency. In some ways it would be less easy to handle since equations (2) are essential to justify the introduction of the electromagnetic potentials ϕ and $\underline{A}$, and these are convenient in solving problems. While at the classical level one might wonder why nature should fail to exploit this potential symmetry, it is necessary to progress to the quantum mechanical level to see why the existence of magnetic monopoles would be so significant theoretically.

The modern theory of magnetic monopoles begins with a paper of Dirac published in 1931. Dirac was trying to generalise quantum mechanics and he found that his efforts led naturally to the existence of magnetically charged particles whose magnetic charge, g, must satisfy

$$qg = 2\pi\hbar n \tag{4}$$

where q is the electric charge of any other given particle, $\hbar$ is Planck's constant and n must be an integer. This is

Dirac's celebrated quantisation condition for magnetic charge. Elementary number theory immediately gives than any electric and magnetic charges occurring would have to be multiples of some basic units, q_o and g_o, respectively, satisfying equation (4): $q_o g_o = 2\pi\hbar N$, for some integer N.

Thus Dirac's argument shows that the existence of a single magnetic monopole would imply the quantization of both electric and magnetic charge (that is that they should both occur in multiples of some basic units). Electric charge quantization is one of the striking facts observed in nature for which a compelling theoretical reason is, as yet, lacking: the charge of each observed particle is an integral multiple of the charge on the electron. The most attractive aspect of the observation of a magnetic monopole is that it would make this observed fact a theoretical necessity.

Arguments leading to the Dirac condition (4) may be given at various levels of sophistication and rigour. In the present article I have only space for the most naïve. Consider a particle of mass m and electric charge q moving in the field of a fixed magnetic monopole of strength g situated at the origin. The equation of motion is

$$m\ddot{\underline{r}} = q\dot{\underline{r}} \wedge \underline{B} = \frac{qg}{4\pi r^3}\,\dot{\underline{r}} \wedge \underline{r} \qquad (5)$$

Although the magnetic field is spherically symmetric, the conservation of angular momentum does not quite take the usual form. In fact the rate of change of the particle's angular momentum is

$$\frac{d}{dt}(\underline{r} \wedge m\dot{\underline{r}}) = \frac{qg}{4\pi r^3}\,\underline{r} \wedge (\dot{\underline{r}} \wedge \underline{r}) = \frac{d}{dt}\left(\frac{qg}{4\pi}\hat{\underline{r}}\right) \qquad (6)$$

where $\hat{\underline{r}} = \underline{r}/r$, the unit vector in the $\underline{r}$ direction. So what is conserved is

$$\underline{J} = \underline{r} \wedge m\dot{\underline{r}} - \frac{qg}{4\pi}\,\hat{\underline{r}} \qquad (7)$$

It may be shown quite easily that the second term in equation (7) is the angular momentum of the complete electromagnetic field and so $\underline{J}$ is the total angular momentum of the system. Now, in quantum mechanics, all components of angular momentum have values which are integral multiples of $\hbar$, or at worst $\frac{1}{2}\hbar$. If this applies to the component of $\underline{J}$ in the direction of $\underline{r}$,

$$\hat{\underline{r}}.\underline{J} = -\frac{qg}{4\pi} \, ,$$

we obtain Dirac's condition (4) immediately.

Dirac drew attention to the fact that his quantization condition meant that the ratio of the force between two magnetic poles to that between two similarly placed electric charges would be large, being

$$g_o^2\, q_o^{-2} = \frac{1}{4}N^2 \left(\frac{q_o^2}{4\pi\hbar}\right)^{-2}$$

if the charges are the smallest possible. If q_o equals the charge on the electron, $q_o^2/4\pi\hbar$ is the fine structure constant, α, which is approximately 1/137, giving the ratio as about 500 N^2. Thus although there is a theoretical symmetry between electricity and magnetism, using the observed value of the fine structure constant, we see that there will be quantitative asymmetries in practice. The larger force should mean that it is more difficult to separate particles with opposite magnetic charges than their electric counterparts. Further, since the strengths of a particle's interactions are thought to be reflected in its mass, magnetic monopoles should be much heavier than electrons. These may be the basic reasons why they have not been observed.

During the last twelve years various theories have been proposed which unify the "weak" nuclear forces, which are responsible for phenomena such as nuclear β-decay, with Maxwell's theory of electromagnetism. In 1974 a Dutch physicist, G.'t Hooft, and, independendently, a Russian, A. M. Polyakov, pointed out that in certain of these unified theories, solutions corresponding to magnetic monopoles exist. These magnetic monopoles are solutions to the classical version of these theories, before quantum mechanics is taken into account. They are not point particles but objects with a certain (very small) spatial extent. At the classical level they have a definite calculable mass.

Our knowledge of the quantum mechanics of such theories is incomplete but certain comments can be made with reasonable confidence. Firstly, each of the fields in the theory can be thought of as made up of particles quantum mechanically, just as the electromagnetic field is made up of photons. Electrically charged elementary particles will enter the theory in this way; in particular, one will obtain particles conventionally called W bosons, which, although electrically charged and massive, mediate the "weak" interactions in the same way that the photons mediate electromagnetism. Secondly, we shall expect to obtain quantum particles corresponding to the classical magnetic monopole solutions. The Dirac condition (4) seems to be built in from the beginning, ensuring consistency. The mass of the magnetic monopole can be shown to be at least 30 times that of the W boson. The W boson itself is still hypothetical, though its discovery is confidently anticipated by many theorists. Its mass must be considerably more than 20 proton masses. This puts the monopoles of 't Hooft and Polyakov way beyond the range of most current observations.

These new theoretical ideas open up a whole range of speculations. It is possible that these new theories of magnetic monopoles are equivalent, at the quantum mechanical level, to "dual" theories in which the roles of electricity and magnetism are reversed. In the dual theories the electrical charges would appear first as classical extended objects. This introduces the possibility of a dual symmetry (indeed a new kind of wave-particle duality) between forces of disparate strengths. Such a symmetry might even explain the relationship between the "strong" and "weak" nuclear forces. But for the moment this is merely piling speculation upon speculation.

et in Cantabrigiense ego...

by Ambrose Bonwicke (1715)

It is observable that among the university men that almost half of them are hypt (as they call it), that is, disordered in their brains, sometimes mopish, sometimes wild, the two different effects of their laziness and debauchery. If there be a sober and diligent tutour, he is affronted, abus'd, injur'd: and when he is so he can find no redress, but brings on himself a greater odium, as in the case of Clare Hall. It may be added that there is no restraint or check on these disorders, but impunity reigns every where, and the most extravagant behaviour is not reform'd. A fellow of St. John's, a rector of a parish not far off of Cambridge, a nephew of an archbishop, runs up and down the country, is at all hors-matches and cockfightings, appears in grey clothes and a crevat. Yet he is not check'd either by the diocesan or the college, though this behaviour is both against canon and statute.

With the immorality of these academics is joynd prophaness and impiety. I have heard them with these ears swear and curse and damn like hectors: and nothing is more usual with them in their common conversation. And this prophane swearing prepares them for that breach of oaths of another nature, which they are guilty of. They solemnly swear to keep the statutes of the university, and of their particular colleges, and yet live in a most visible violation of them, them I mean which respect not only their manners, but their exercises: but at the end of the year they meet in the Regent house, and are absolvd by a priest without shewing any signs of repentance. They shew little regard and reverence for the Lord's day: on all Sundays in the afternoon they go immediately from the church to the coffee-houses, as if they thought it were but passing from one place of diversion to another. Though there was her majesties proclamation against prophaning this day, in which persons were particularly forbid to go to coffee-houses, yet the vice-chancellor and clergy take no notice of it, but act contrary to it. And whether the undergraduates and scholars repair to church on this day, or stay at home, is little minded by their tutours: but when they go, every body knows of it, for they talk aloud in the church, they laugh, they most irreverently behave themselves even in the time of divine service. If they meet not with the desirable spectacle, they run out of the church as if they were frighted: and their practice is to ramble up and down from church to church throughout the town, to gaze on the young women, and (as some of them are wont to confess) to tell how many patches they wear.

Pell's Equation

by R. C. Mason

"'Twas graduation day, and all the scholars about to graduate were assembled in the First Court. As he was inspecting them, the Senior Tutor, a man of no mean mathematical ability, suddenly noticed that those scholars present could be formed into squares of equal size, and numbering forty one. Delighted with this observation, he formed the groups accordingly.

A sorry figure stumbled into the court. The Senior Tutor was furious, and berated him: "Late again, Shufflesworth, you miserable specimen." The Senior Wrangler however, was able to prove his reputation, and spoke thus: "Surely, Professor, if we re-form into one large group we shall make but one perfect square." "

Who was the Senior Wrangler?

We are required to solve $m^2 - 41\,n^2 = 1$ in integers.

Plainly $6n < m < 7n$, so if $m = 6n + a$, then $0 < a < n$ and

$$-5n^2 + 12an + a^2 = 1$$

It follows that $2a < n < 3a$, so if $n = 2a + b$, then $0 < b < a$

and $$5a^2 - 8ab - 5b^2 = 1$$

Then, if $a = 2b + c$, $0 < c < b$ and $-b^2 + 12bc + 5c^2 = 1$.

Then, if $b = 12c + d$, $0 < d < c$ and $5c^2 - 12cd - d^2 = 1$.

Then, if $c = 2d + e$, $0 < e < d$ and $-5d^2 + 8de + 5e^2 = 1$.

Then, if $d = 2e + f$, $0 \le f < e$ and $e^2 - 12ef - 5f^2 = 1$.

which is just the Euclidean algorithm. Now we write $M = e - 6f$ and $N = f$, and then $M^2 - 41\,N^2 = 1$.

Also we have, on substitution, $$\left.\begin{aligned} m &= 2049\,M + 13120\,N \\ n &= 320\,M + 2049\,N \end{aligned}\right\} \quad (*)$$

and it may be verified directly that $M^2-41N^2=1 \Rightarrow m^2-41n^2=1$. Thus every solution is obtained by successive applications of the transformation (*) to the trivial solution $m = 1$, $n = 0$. In other words, if we write

$$m_{t+1} = 2049\,m_t + 13120\,n_t$$
$$n_{t+1} = 320\,m_t + 2049\,n_t$$

$m_o = 1$, $n_o = 0$, then the set $\{(m_t, n_t) : t \in \mathbb{N}\}$ is just the set of solutions. It is immediate that

$$m_t = \tfrac{1}{2}\left[(2049 + 320\sqrt{41})^t + (2049 - 320\sqrt{41})^t\right]$$

and $$n_t = \frac{1}{2\sqrt{41}}\left[(2049 + 320\sqrt{41})^t - (2049 - 320\sqrt{41})^t\right]$$

that is, $m_t \pm n_t\sqrt{41} = (m_1 \pm n_1\sqrt{41})^t$.

It is clear that the above gives an algorithm for producing all the solutions to equations of the form $m^2 - dn^2 = 1$, for d a fixed integer.

There is a short cut to this solution, which is to observe that substituting c = 0, b = 1 gives m = 32, n = 5 a solution of $m^2 - 41n^2 = -1$, squaring which gives $m^4 - 82n^2m^2 + 41^2n^4 = 1$, that is $(m^2 + 41n^2)^2 - 41(2mn)^2 = 1$, and $m_1 = 2049 = m^2 + 41n^2$, $n_1 = 320 = 2mn$ follows.

We gain a little insight into this apparent coincidence if we consider $\mathbb{Z}[\sqrt{41}] = \{a + b\sqrt{41}: a,b \in \mathbb{Z}\}$ as a subring of $\mathbb{C}$. We can define $N: \mathbb{Z}[\sqrt{41}] \to Z$ by $N(a + b\sqrt{41}) = a^2 - 41b^2$, and N is clearly multiplicative, that is $N(xy) = N(x)N(y)$. We are looking for solutions of $m^2 - 41n^2 = \pm1$, i.e. $N(x) = \pm1$, where $x = m + n\sqrt{41}$, which is to say x is a unit of $\mathbb{Z}[\sqrt{41}]$. Then if $N(m + n\sqrt{41}) = -1$, $N((m + n\sqrt{41})^2) = 1$, whence the short cut.

Reasonable proficiency can be acquired in the use of the algorithm, and the reader may care to try his hand at $m^2 - 53n^2 = 1$ and $m^2 - 19n^2 = 1$. The ambitious and numerically perfect may care to attempt $m^2 - 94n^2 = 1$, for which the smallest known solution is m = 2143295, n = 221064 (that is, for m and n strictly positive).

The algorithm may be somewhat simplified by taking a (very) slightly different approach. Define the continued fraction of a positive number x as a sequence of positive integers $r_1, r_2, r_3, \ldots$ and write $x = [r_1, r_2, r_3 \ldots]$ where r_i is defined inductively as follows: Let $x_1 = x$ and $r_1 = [x]^i$, and for i > 0 define $x_{i+1} = 1/(x_i - r_i)$ if $x_i > r_i$ otherwise define $x_{i+1} = 0$, and then define $r_{i+1} = [x_{i+1}]$.

Then $x = r_1 + \cfrac{1}{r_2 + \cfrac{1}{r_3 + \ldots}}$ and so by an abuse of notation we may write $[a_1, a_2, \ldots, a_n, x]$ for $[a_1, a_2, \ldots, a_n, r_1, r_2, \ldots]$.

For a rational m/n the continued fraction is necessarily finite (that is r_i is eventually zero) and the calculation is the same as the Euclidean algorithm for determining the HCF of m and n. For example, 32/5 = [6,2,2] and 2049/320 = [6,2,2,12,2,2], which agrees with our original calculation.

We now make a small observation, namely that if $m^2 - 41n^2 = 1$ then $m^2/n^2 - 41 = 1/n^2$ and so m/n is a good approximation to $\sqrt{41}$. We may now reasonably guess that the continued fraction for $\sqrt{41}$ is intimately connected with that for m/n. Indeed, it is not hard to see that the calculations are essentially the same. It so happens that by a stroke of 'luck' the continued fractions for square roots are quite easy to calculate.

Put $x = \sqrt{41}$, and note that $6 < \sqrt{41} < 7$. Then $x_1 = \sqrt{41}$,

$r_1 = 6$, $x_2 = \frac{\sqrt{41} + 6}{5}$, $r_2 = 2$, $x_3 = \frac{\sqrt{41} + 4}{5}$, $r_3 = 2$, $x_4 = \sqrt{41}+6$, $r_4 = 12$ and $x_5 = x_2$. Thus $\sqrt{41} = [6,2,2,12,2,2,12,2,2,\ldots]$. It is to be noted that only the approximation $6 < \sqrt{41} < 7$ was needed to calculate the continued fraction, which is to say no $x_i = \frac{a\sqrt{41} + b}{c}$ for $|a| > 1$ occurs. We will see that there is no mere coincidence.

If $\sqrt{41} = [\underbrace{6,2,2,12,2,\ldots,2,2}_{3n \text{ terms}},x+6]$ then $x = \sqrt{41}$ and we also see that $\sqrt{41} = \frac{ax + b}{cx + d}$, where $\frac{a}{c} = [\overbrace{6,2,2,12,\ldots,2,2}^{3n \text{ terms}}]$ (obtained by formally putting $x = \infty$). Then it is easy to show that $a = d$, $b = 41c$.

We define, for $z(y) = \frac{py + q}{ry + s}$, $\Delta(z(y)) = ps - qr$. (p,q,r,s integers with no factor common to all). Plainly $\Delta(z(y) - 1) = \Delta(z(y)) = - \Delta(1/z(y))$. (@)
Hence, if $z(y) = \frac{ay + b}{cy + d} = \frac{ay + 41c}{cy + a}$, we obtain by successive applications of (@) that $\Delta(z(y)) = (-1)^{3n} = (-1)^n$. But for n even this is exactly what we want, for then $\Delta(z(y)) = a^2 - 41c^2 = 1$. If $n = 1$ we recover $a = 32$, $c = 5$, a solution of $a^2 - 41c^2 = -1$.

We now see why we had that stroke of 'luck'. In calculating the continued fraction for $\sqrt{41}$, we had occasion to consider the integer part of x_i. By taking n large enough, we can use the above to write $x_i = \frac{px + q}{rx + s}$, for some integers p,q,r,s given by the continued fraction. Upon rationalising the denominator we have $x_i = \frac{p\sqrt{41} + q}{r\sqrt{41} + s} = \frac{(ps-qr)\sqrt{41} + (qs-41pr)}{s^2 - 41 r^2}$ and since $ps-qr = (-1)^{i-1}\Delta(z) = (-1)^{n+i-1}$ we have the stroke.

Thus in general the most efficient way to solve Pell's equation, $m^2 - dn^2 = 1$, in integers, is to take m/n as a suitable truncation of the continued fraction for $\sqrt{d}$.

We end with a glance at a generalisation. It has been remarked that the solutions of $a^2-41b^2=+1$, are, when written as $a + b\sqrt{41}$, the (group of) units of $\mathbb{Z}[\sqrt{41}]$, and the algorithm shows directly that this group must be isomorphic to $C_2 \times C_\infty$, where C_2 is the group generated by -1 and C_∞ is an infinte cyclic group, generated by $(32 + 5\sqrt{41})$, given by the smallest solution of $a^2-41b^2=+1$ in positive integers. Indeed, the algorithm also shows that the group of units of $\mathbb{Z}[\sqrt{d}]$ is isomorphic either to $C_2 \times C_\infty$ or to C_2 (although it does not show that the latter occurs only when d is a perfect square). This result is a special case of Dirichlet's Unit Theorem in the case that d is square-free and $d \equiv 2$ or 3 (4), for in this case $\mathbb{Z}[\sqrt{d}]$ is a ring of algebraic integers. Dirichlet's Unit Theorem states that the group of units of a ring of algebraic integers is isomorphic to $W \times F$, where W is a group of roots of 1 and F is a free Abelian group of finite rank, which is 1 in the case of $\mathbb{Z}[\sqrt{d}]$.

Lecture-Dynamics

by P. Verschueren

This paper attempts a brief introduction to one of the newest and most rapidly growing fields of mathematics, that of Lecture Dynamics. A full account of the foundations and development of this subject is given in the survey article by N. Paulver in the British Association of Lecture Dynamics Quarterly Vol. I pp. 2-3.

Basic Definitions Lecture-Dynamics comprises the study of the way in which a lecturer proceeds along a lecture course, of the mode of transfer of information, and of the induced flow in the students' notes. Collation and analysis of data concerning lecture courses as integral units, without reference to their internal dynamical systems, forms the objective of the complementary science of Lecture-Statistics, usually called Lecture Statics.

Terminology We say that a theorem A is (logically) implied by a theorem B, if the result of B can be used as a (n irredundant) step in the proof of A, written $B \leq A$. The set of the theorems with the ordering relation of logical implication defines the logic space. The equivalence class of a theorem is the set $\{B : B \leq A \text{ and } A \leq B\}$; A and B are then said to be equivalent.

A path is defined as any sequence of theorems in the space. The path is continuous or connected if each theorem of the path is logically implied by its predecessor.

Basic to all the work in Lecture-Dynamics and related fields is the need to isolate quantitative variables which are measurable in the physical environment. This is equivalent to defining a metric on the logic space.

Fundamental Lemma The logic space is metrisable.
Proof We define the following system of measurables and units.

The lecture motive force, E, of a lecturer, is the force which causes information to flow from him to each individual student. A lecture motive force of one Jolt is the force required to overcome the (possibly very complex) impedance to communication generated by holding a standard nine o'clock lecture.

The inductance, L, is the ability of the audience to produce motivation in the lecturer. An inductance of one Noose is that which produces a motivation equivalent to that provided by a single coil (of rope) hanging from the ceiling of the lecture room. This unit is far too large for normal purposes, so we usually use a micro-Noose.

The ability of the lecturer to remain indifferent, when a charge of incompetence or boredom is laid against him, is termed capacitance.

The magnetic attraction of the lecturer's vice (normally negligible) is denoted by B.

Then recent work leads us to believe that, statistically, $E = 2\pi L/C + B$.

The direction and rate of flow of lecture delivery along its course is denoted by $\underline{j}_o(t)$. The unit of magnitude is defined as that rate of flow along two infinitely long, uniformly progressing, parallel lecture courses, which produces a tension (or rivalry) between the lecturers of one insult per lecture.

We have now constructed a system of measurement using Système Internationale or S.I. units almost everywhere. Elementary semantic analysis gives immediately that we therefore have a metric space. Q.E.D.

The metric thus defined is called the Paulver-Lebesgue metric.

We now return to consider the transfer of information from lecturer to student. Using $\underline{j}_o$ as defined above, the lecture course is defined by $\underline{s}_o(t) = \int_0^t \underline{j}_o(x)\, dx$, called the flow-path. We call the paths followed by the notes of the students the note-paths, denoted $\underline{s}_1, \ldots, \underline{s}_n$ (for n students). Obviously, under ideal conditions, these will all follow the lecture course exactly, or $\underline{s}_o = \underline{s}_1 = \ldots = \underline{s}_n$. However, information is of course quantized, so that, by the Uncertainty Principle, miscommunication can occur, causing a discrete jump in the student's notes to a neighbouring connected path, so that the notepath is now disconnected. The complete set of corresponding note-flows $\underline{j}_o, \ldots, \underline{j}_n$ are denoted by $\underline{J}$, the flow tube. We now simplify the situation by assuming the number of students, n, to be large. The flow tube at time t then approximates a multivariate normal distribution, under the P-L metric, with mean $\underline{j}_o(t)$. The variance of the distribution is clearly an increasing function of time, so that the flow tube behaves like a Gaussian wave packet. In fact the variance is here the value of a rather more general measure (which applies for any n), namely the divergence of the flow tube, $\underline{\nabla}.\underline{J}$. Our "ideal conditions" can now be expressed as $\underline{\nabla}.\underline{J} = 0$, known as the conditon for maximised, well-defined flow, or Maxwell condition. The main aim of lecture-dynamics is to minimise $\underline{\nabla}.\underline{J}$.

We are now in a position to state and sketch the proof of the central theorem of Lecture-Dynamics.

Heineken-Borel Theorem In any lecture course, all note-paths will eventually diverge from the flow-path, and tend in the limit to Markov processes. The rate of divergence is directly proportional to the proof density of the course. (Thus, in general, the attempt to absorb courses of high proof content will rapidly induce random walks.)

Sketch Proof: Apply the flow equation. To show Markov, use Kolmogorov's 0-1 Law, and the lack-of-memory property of the students.

We conclude this paper with an example of some recent work in the field.

We define a Russell subject as one in which all lemmata are finitely generated and of essentially bounded proof-length. Henceforth we shall restrict our attention to such subjects.

Theorem (Hopkins-Morris BALD Quar. Vol. VII pp 63-68). The eventual power spectrum of any simple well-formed course in a Russell subject is invariant under finite timetable transformation.

Proof Since the subject is Russell, the set of essentially different theorems is finite, N_0 say. But each final note path is a random walk (Heineken-Borel) and so these theorems will generate the note-path space which thus has dimension N_0. Now by a corollary of the Kürtz-Siegfried lemma (BALD Quar. Vol. II pp 37-39), this means that any finite transformation can be expressed as a finite composition of alternating lecture transpositions without affecting the net final divergence (since the course is simple). But the course is also well-formed so that it must be motivation and communication invariant under such transpositions. Thus the final divergence and associated power spectrum are invariant.

Q.E.D.

Corollary Since all positive-valued courses can be expressed as a direct Cantor sum of simple, well-formed courses, it follows that only degenerate courses are affected by finite transformations.

Hence we see that the uniformly most powerful syllabus contains only Russell subjects lectured in a non-degenerate manner.

A convex polyhedron P_1 with just nine vertices $A_1,\ldots,A_9$ is given. P_i is a polyhedron obtained by translating P_1 through the vector A_1A_i, for i = 1 to 9. Prove that at least two of the polyhedra $P_1,P_2,\ldots,P_9$ have some interior point in common.

A soldier has to test an area of a region shaped like an equilateral triangle, including its boundary, for mines. His detector has a range of half the altitude of the triangle. Starting from one vertex, what is the shortest path for him to take such that he scans the whole region.

Book Reviews

Proofs and Refutations: The Logic of Mathematical Discovery
by Imre Lakatos
£7.50 (Hardback), £1.95 (Paperback) 1976 C.U.P.

This is a serious study of the methodology of mathematics written in the form of a discussion between teacher and students on the Euler conjecture that for a polyhedron vertices, edges and faces are connected by $V - E + F = 2$. The discussion largely mirrors historical development; throughout Lakatos presses the view of mathematics as a growing, maturing study, rather than a series of infallible deductions from basic axioms, which has always been considered the essence of mathematics.

The first problem we face is: given a proof of a conjecture, what do we do in the face of a counterexample which contradicts it? We can reject the conjecture; we can reject the counterexample as a "monster"; we can compile a list of counterexamples and thus construct a "region of validity" of the conjecture; or we can analyse our "proof", discover which lemma (explicit or implicit) in it is contradicted by the counterexample, and modify the theorem (not the proof) accordingly. This is the method of proofs and refutations; by it a succession of counterexamples leads to a creative refinement of our knowledge. That this last course is the best seems obvious; but in an appendix we are shown that it was lack of appreciation of this which was responsible, among other things, for the inordinate difficulty nineteenth-century mathematicians found in correcting Cauchy's "theorem" that "the limit of a convergent series of continuous functions is continuous", when contradicted by series due to Fourier, by inventing the concept of uniform convergence.

This is an immensely readable and very comprehensible book; a second-year undergraduate would cope with it easily, although it will make its deepest appeal to those whose special interest is foundations of mathematics and metamathematics generally. It is a book which has made me think.

O. L. C. Toller

A First Course in Abstract Algebra
by P. J. Higgins
£4.50 (Hardback), £2.25 (Paperback) 1975 Van Nostrand

Professor Higgins has set out to write a book in which "the student's first encounter with any new mathematical concept ... (is) closely followed by a study of the applications which justify its introduction", and in this aim, at least, he has been largely successful (he admits, and I am sure that the pun is unintentional, that "in the case of rings ... there are some difficulties in achieving the ideal"). The structures which he introduces are groups, rings and fields, and the applications, which occupy a considerable proportion of this fairly short text, include separate chapters on factorisation and linear congruences in the integers, polynomials over the familiar number rings, and a study of $\mathbb{Z}_n$ and $\mathbb{Q}$ as canonical ring and field examples. Even the more theoretical chapters are written in a very concrete style, with copious, often unusual, examples offered at every stage, but this does mean that the book is rather short on solid material, so that, for instance, it would not be adequate for any of the Cambridge algebra courses.

In an elementary book of this sort, one might have hoped for rather more exercises, and in particular more theoretical exercises, since a course of abstract algebra should surely leave the student able to handle confidently the concepts involved, as well as giving him facility in calculation. Finally, though no book is perfectly error free, it cannot encourage the beginner to be presented early on with the ambiguous statement "the composite function gof exists only when the domain of g is the same as the range of f".

M. C. Davies

Dimension Theory of General Spaces
by A. R. Pears
£16.50 1975 C.U.P.

Everyone knows that the dimension of $\mathbb{R}^n$ is n. The object of dimension theory is to define and investigate integer or infinite valued dimension functions on more general classes of topological spaces. These functions should assign the same dimension to homeomorphic spaces and give the usual result for the spaces $\mathbb{R}^n$. There are many plausible ways of defining dimension functions, and this book gives a clear account of their properties and interrelationships. In particular, it contains the most lucid presentation that I have ever come across of Prabir Roy's example of a metrisable space with small inductive dimension equal to zero, but with large inductive dimension and covering dimension both equal to one. The printing is well up to the usual high standard set by the CUP, and, at £16.50, this book is excellent value for money. On page 141, one finds the following proposition:

Let X be a normal regular space and let M be a subspace of X which is weakly paracompact, normal and locally strongly paracompact. Then dim M$\leqslant$dim X.

If you are at least a Part III student and if results like this really turn you on, you should rush out and snap up Heffer's last remaining copy of this book.

Dr. A. M. Tonge

Topos Theory
by P. T. Johnstone
£17.40 1977 Academic Press

Pure Mathematicians should be feeling their foundations shaking beneath their feet. In recent years (as the blurb of this book so quaintly puts it, "since the pioneering work ... in 1969 and 1970") it has been realized that one gains a much better understanding of many mathematical matters by replacing the static concept of 'set' by a dynamic concept of 'variable set'. Benefits of this attitude include a fertile interplay between geometry and logic (each enriching the other) and a vast clarification of universal algebra.

This book provides a systematic introduction to the theoretical details of this change of outlook. After a summary of essential background (merely category theory, sheaf theory, and concepts originally motivated by algebraic geometry) it gives the definition and properties of the basic objects, elementary toposes, each of which can be thought of as a suitable replacement, as universe of discourse, for the category of sets. There follow chapters on constructions within and without toposes; logical matters, showing that the logic of toposes is not always 2-valued (true, false), but rather close to that of intuitionism; and concepts, such as cohomology, imported from the geometric side. Especially important are the accounts of such basic objects as natural numbers, real numbers, and algebraic theories in a topos (where the attitude reveals, for instance, the essence of the difference between the Dedekind-section and Cauchy-sequence definition of real numbers); the study of points of a topos, putting the logical Löwenheim-Skolem and Gödel-Henkin theorems into their (geometric) perspective; and the account of forcing using toposes, leading to a topos-theoretic proof of the independence of the Continuum Hypothesis and (as an exercise!) of the Axiom of Choice.

Of course, the difficulty is that all this is incomprehensible: the fate of most prophets in their own time. This is not to say that the book is badly written: on the contrary, in mathematical terms it is excellently presented, and for erudition the phrase tour-de-force is brought to mind. But as the author himself writes: "The average mathematician, who regards category theory as 'generalised abstract nonsense', tends to regard topos theory as generalised abstract category theory". Contrariwise, an understanding of even the basic attitude of the theory requires a formidable background in several difficult areas of mathematics. But given this, the book provides a splendid introduction to one of the most exciting recent developments in pure mathematics.

Dr. B. R. Tennison

Newer Uses of Mathematics
edited by Professor Sir James Lighthill
£2.25 1978 Penguin

The pertinent word in the title of this very readable book is 'uses'. Six eminent mathematicians describe developments in that branch of the subject generally known as applicable mathematics.

Professor Lighthill opens with his own chapter on the Physical Environment, the section on numerical weather prediction being verbatim the text of his talk to the Trinity Mathematical Society in 1976. The chapter on finance is written by Professor R. E. Beard, a prominent actuarial theorist, in which the mathematical complexity of modern accountancy, banking and investment is described. There are four other chapters - on biomathematics, operational research, networks and planning. Catastrophe Theory, however, is not discussed in the book because it is "too new and too technical", but an elementary account was given recently by Professor Zeeman on television. A noticeable theme running through the book is that of the parallel development of applicable techniques and computer technology.

An excellent feature of 'Newer Uses' is the bibliography after each chapter, referring the interested reader to introductory works on all topics covered. The book should be easily within the reach of any sixth-former and I would recommend it to every mathematician who does not see his work applied to problems in the real world.

M. R. Kipling

Nonlinear Ordinary Differential Equations
by D. W. Jordan and P. Smith
£12 (Hardback), £6.50 (Paperback) 1977 Clarendon Press

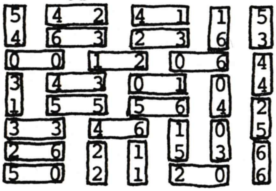
The theory of non-linear systems is one of wide applicability to the physical, biological and social sciences. This book is an elementary account of the theory with a continual emphasis on problems of practical interest, and a qualitative slant to the novel phenomena which obtain in the presence of non-linear effects. The chapters form fairly self-contained groups covering two dimensional systems and second order equations, small parameter singular perturbations and forced oscillations, formal stability, and the existence of periodic solutions in certain representative equations.

The layout is tidy, comprehensive and readable, and there are many examples both worked and as exercises. The result is an excellent text which would be a useful accompaniment to a course on Non-linear equations at the 2nd or 3rd year undergraduate level.

C. M. Noble

Solutions to Problems Drive

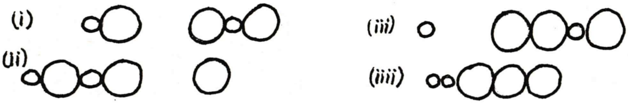
1 $a\dfrac{10^{68n}-1}{69}$, a = 7, 8 or 9; n any positive integer.

2 

3 9, 8, 7, 5, 4, 2, 1

4 439 1806725

5 166 spiders, 266 starfish, 2 tapeworms, 2553 cyclops 1 peacock and Macbeth's corpse.

6 

7 (a) $u_n = u^2_{n-2} - u_{n-1}$, so next term is -13.

(b) if p_n is the nth prime, $u_n = 2^{p_n} + p_n$, so next term is 8205.

(c) $u_n - 1$ is the nth natural number with rotational symmetry, so next term is 610.

8

A	B	C	D	E
D	C	A	E	B
C	E	B	A	D
B	D	E	C	A
E	A	D	B	C

9 The time was 7.48 a.m. (4.12 can never be in the morning).

10 625)631938

11 $\frac{1}{4}, \frac{1}{4}, \frac{1}{2} - \frac{2}{3\pi}, \frac{2}{3\pi}$.