Mathematical Spectrum

Volume 6 1973/74 *Number 2*

A Magazine of *Contemporary Mathematics*

Published by the *Applied Probability Trust*

Mathematical Spectrum is a magazine for the instruction and entertainment of student mathematicians in schools, colleges, and universities, as well as the general reader interested in mathematics. It is published by the Applied Probability Trust, a non-profit making organization established in 1963 with the support of the London Mathematical Society. The object of the Trust is the encouragement of study and research in the mathematical sciences.

This number 2 completes Volume 6 of *Mathematical Spectrum* (1972/73); the first issue of the Volume was published in the Autumn of 1973.

Articles published in *Mathematical Spectrum* deal with the entire range of mathematical disciplines (pure mathematics, applied mathematics, statistics, operational research, computing science, numerical analysis, biomathematics). Both expository and historical material may be included, as well as elementary research and information on educational opportunities and careers in mathematics. There is also a section devoted to problems.

Articles are normally commissioned by the Editors; the Editorial Committee also welcomes the submission of suitable material, including correspondence, queries and solutions to problems, for publication in *Mathematical Spectrum*. All correspondence about the contents should be sent to:

The Managing Editor, *Mathematical Spectrum*,
Hicks Building, The University, Sheffield S3 7RH.

Who Was the First Non-Euclidean?

UNDERWOOD DUDLEY
DePauw University

In a recent issue of *Mathematical Spectrum* (Vol. 5 No. 2, 'The Parallel Axiom', pp. 58–66) Rogers stated that non-Euclidean geometry was discovered 'quite independently' by Gauss, Bolyai, and Lobachevsky. There remains, I think, some doubt about this: it is the purpose of this article to suggest that the main credit for the discovery of non-Euclidean geometry should go to Carl Friedrich Gauss.

In mathematics, great discoveries are made by great mathematicians. This should be no surprise, for it is a truism: if you make a great discovery, then you are by definition a great mathematician. Work by lesser mathematicians is seldom of great consequence or use to later mathematicians, and the history of mathematics would not be very different if most of the less gifted mathematicians had never lived. Take a list of some of the great giants of mathematics from 1600 to 1850, among the greatest minds the world has ever seen:

Newton	Leibniz	Bernoulli	Euler
Legendre	Laplace	Cauchy	Gauss

If no other mathematicians had existed during those 250 years, these men could probably have discovered most of the new mathematics developed in that time; most mathematicians would agree that they were capable of it, singlehanded.

There is one seeming exception to the rule that lesser mathematicians do not make big advances in mathematics. That is the discovery of non-Euclidean geometry; the consensus from many sources is that the discoverers were Nicholas Lobachevsky and Janos Bolyai, independently, around 1830. Yet Lobachevsky and Bolyai lacked one of the main characteristics of great mathematicians: versatility. A great mathematician contributes new discoveries to many different areas of mathematics. Newton not only discovered calculus, he gave us the binomial theorem, finite differences, and celestial mechanics. Euler's name is attached to theorems in calculus, mathematical physics, geometry, and number theory; Gauss's name lives on in physics, astronomy, and at least five different branches of mathematics, and so it is for all of the mathematical giants listed previously. But this is not so for Bolyai and Lobachevsky: neither of them did anything of consequence in any other branch of mathematics. Instead, they devoted their mathematical lives to elaborating their geometry. The published work of Bolyai and Lobachevsky made little impression when it first appeared, and one reason for this was the unclear and ill-organized writing, quite uncharacteristic of the work of the great mathematicians. Gauss wrote of Lobachevsky's work that it was

> 'a confused forest through which it is difficult to find a passage and perspective, without having first gotten acquainted with all the trees individually.' (reference 4, p. 182)

Lobachevsky and Bolyai could not be numbered among the very great mathematicians, if judged by the exalted standard necessary when considering their place in human history. Yet they are credited with the discovery of non-Euclidean geometry, which was an enormous leap in mathematics and had profound influence on the future of physics and philosophy. It could be called the most important mathematical discovery of the nineteenth century, as calculus had been for the two hundred years before. Various authorities have attributed this discovery to them, and written:

> 'J. Bolyai... shares with Lobatschewsky the honor of the discovery of non-Euclidean geometry.' (reference 3, p. 96)
>
> 'At this point we should give credit to Johann Bolyai, a Hungarian mathematician, who worked out the notion of a non-Euclidean geometry simultaneously with Lobashevsky, but independently. Here we have a recurrence of the Newton–Leibniz phenomenon, the sort of thing that usually occurs when mathematics is ripe for a new discovery.' (reference 5, p. 259)
>
> 'The actual development of non-Euclidean geometry, however, was not based on Saccheri's work and did not come about until the early nineteenth century, over two thousand years after Euclid. Amazingly, it was then developed independently by three people, Lobachevsky, Bolyai, and Gauss.' (reference 6, p. 209)
>
> 'It was a century after Saccheri that three mathematicians in three different countries (Nicholas Ivanovitch Lobachevski, Russia; Janos Bolyai, Hungary; and Karl Friedrich Gauss, Germany), independently and apparently without knowledge of Saccheri's curious contact with non-Euclidean space, came to the conclusion that Euclid had known exactly what he was doing when he made his statement about parallels a postulate instead of a theorem.' (reference 7, p. 154)

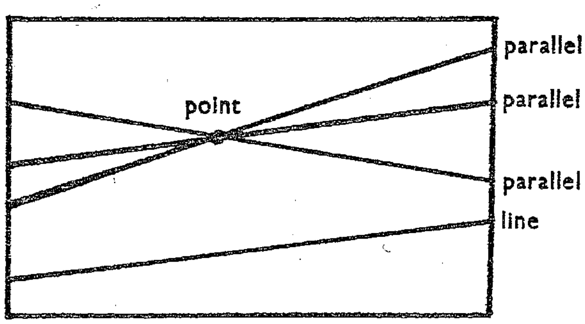
Other examples could be added to these.

I hope to adduce evidence that there was nothing amazing about the development of non-Euclidean geometry, and that it was not like the discovery of calculus, where the great minds of Leibniz and Newton independently worked out the subject. There is much circumstantial evidence that Gauss was the originator of non-Euclidean geometry. Bolyai and Lobachevsky may not have discovered it independently; it seems possible that they may indeed have received the idea indirectly from Gauss. Because Gauss did not publish his discoveries, Bolyai and Lobachevsky deserve the credit for *developing* the subject, but perhaps not for having the original idea which made it possible.

The idea which led to non-Euclidean geometry came from Euclid's fifth postulate. Euclid's five postulates were, in effect:

1. A straight line can be drawn between any two points
2. A straight line can be indefinitely continued
3. A circle can be drawn with any center and radius
4. All right angles are equal to each other
5. Through a point outside of a given line, one and only one line can be drawn parallel to the given line.

The first postulate says that we may use a straightedge, the second says that it may be as long as we like, and the third says that we may use a compass. One could hardly object to the fourth. But the fifth postulate has an entirely different character. For one thing, given an actual point and line, any parallel which we tried to draw would intersect the line, perhaps ten miles away if we drew it very, very carefully, but more likely within ten feet. You could thus argue that in the physical world of pencils and rulers there are no parallels at all. On the other hand, if the universe were finite, then we would be able to draw many parallels to a given line through a point, if by 'parallel' we mean 'a line which does not intersect a given line'. The fifth postulate is not quite as intuitively appealing as the first four, and many mathematicians through the years attempted to prove it from the other four postulates and the axioms of Euclidean geometry. None succeeded, for it is impossible to prove the fifth postulate: it is independent of them, and could be replaced with some other assumption without introducing any logical difficulties.



There are two assumptions about parallel lines which could be made instead of the fifth postulate, namely that either more than one, or no, parallel can be drawn. Both of these assumptions lead to non-Euclidean geometries. What is important is the idea that the fifth postulate could be replaced by some other postulate. The person who first had that idea deserves to be called the founder of non-Euclidean geometry.

One of the people who thought that he had proved the fifth postulate wrote a whole book on it: *Euclid Freed of Every Flaw* by Saccheri, was published in 1733. Saccheri started by assuming that more than one parallel could be drawn to a given line through a point outside the line, and hoped to deduce from this something which was clearly false. He made no mistakes in logic, and towards the end of the book he asserted that one of his deductions was false because it was 'contrary to the nature of the straight line'. That reason was not sufficiently convincing for most mathematicians, and Saccheri's work was forgotten. But Saccheri's deductions in his search for a contradiction later became *theorems* of non-Euclidean geometry. What he was not capable of providing was the idea that Euclid's fifth postulate could be anything but a true theorem, and it is this idea which is the great achievement.

Did Bolyai and Lobachevsky, independently have this idea? We will never know for sure, but there is some evidence that the idea may have come from Gauss. Gauss was thinking about geometry from his early years on: in 1792, when he was 15, he wrote in a letter that he had thought of a geometry

> 'which would have to occur and would occur in a rigorously consistent manner, if Euclidean geometry is not the true one.' (reference 3, p. 175)

Here is the first foreshadowing of non-Euclidean geometry, in 1792, almost forty years before the published works of Bolyai and Lobachevsky. Gauss still seemed to think that Euclidean geometry must be the geometry of the physical world, but he was only 15 and, as we will see, he changed that opinion later. He continued to work out the consequences of his idea: if Saccheri could do that, then with how much greater ease and power could it be done by Gauss! Gauss merely mentioned his results in letters; he had recognized that the sum of the angles in a non-Euclidean triangle would be less than 180° by 1794, he had developed non-Euclidean trigonometry by 1816, and in 1819 he wrote that he had developed non-Euclidean geometry almost completely. In 1817 he wrote a letter which shows that he had the modern idea that Euclidean geometry is not necessarily the geometry of the physical world, and Gauss was the first to have this idea;

> 'I am coming more and more to the conviction that the necessity of our geometry cannot be proved, at least not by human intelligence nor for human intelligence. Perhaps we shall arrive in another existence at other insights into the essence of space, which are now unattainable to us. Until then one would have to rank geometry not with arithmetic, which stands *a priori*, but approximately with mechanics.' (reference 4, p. 180)

Gauss here expressed explicitly the idea which made non-Euclidean geometry so revolutionary. Until Gauss, no one doubted that Euclidean geometry was the true geometry of the physical world; no one thought that any other geometry was *possible*. If you do not have the idea that something is possible, you will certainly not develop it, and sometimes, as with non-Euclidean geometry, once the idea that something is possible becomes known, then several people are able to develop it, as I would suggest that Bolyai and Lobachevsky did.

Gauss's mathematical talent was recognized early. He soon surpassed his schoolmaster, and an assistant, Johann Martin Bartels (1769–1836), was assigned to teach him more mathematics. They studied and learned together:

> 'Between the assistant of seventeen and the pupil of ten there sprang up a warm friendship which lasted out Bartels' life.' (reference 1, p. 222)

Gauss and Bartels were in personal contact until 1796, and again from 1805 to 1807. It is unlikely that Gauss did not tell Bartels about his new geometry, for he wrote about it in letters to friends less close than Bartels. He already knew in 1794 that the sum of the angles in a non-Euclidean triangle was less than 180°, and who could resist telling that picturesque discovery to his mathematical friends? It is to my mind most likely that Bartels had, at second hand from Gauss, the essential ideas of non-Euclidean geometry.

Nicholas Ivanovitch Lobachevsky (1793–1856) lived in Kazan, Russia, from the age of seven. He spent forty years of his life at the University of Kazan as student, teacher, and rector; he seems to have been a hard-working and estimable man, much of whose time was taken up with administrative and non-scholarly duties like organizing the University library. The University of Kazan was founded in 1805, and Lobachevsky entered it as a student in 1807. Since the number and quality of scholars in Russia then were not up to the standards of Western Europe, many Germans were offered posts at the new university. The professor of mathematics, the one who taught Lobachevsky, was Johann Martin Bartels. Lobachevsky was far and away the best student Bartels had, so it is natural to suppose that Bartels would give him a good deal of time and attention. Here might be another Gauss! It is not to be doubted that Bartels and Lobachevsky discussed mathematics for many hours, and it is also not to be doubted that Bartels would talk of Gauss, for by this time Gauss had already made a reputation as a powerful mathematician, and acquaintance with him would be a mark of distinction. Indeed, it would be the greatest event of Bartels' mathematical life. It is not unnatural to suppose that one day Bartels might have said, in the jocular manner of professor (who knows the answer) to student,

"Ah, Nicholas, what is the sum of the angles of a triangle?"

"Why, two right angles. Everybody knows that."

"What if I told you that Gauss said that sometimes a triangle has less than two right angles? What would you say to that?"

"But that is impossible! It has been proved."

"But Gauss still says that sometimes it is less. Let me tell you why."

And then Bartels could have proceeded with that main joy of a teacher's life, explaining something new to someone who is interested and will understand. It may of course have been that Lobachevsky came upon the idea of non-Euclidean geometry all by himself and free of outside influence, but it seems unlikely to me. If he had made other contributions to mathematics it would seem less doubtful; if he had never known Bartels, it would be less doubtful still.

It is not hard to find the link between Gauss in Germany and Bolyai in Hungary. Gauss was a student at Göttingen at the same time as the father of Bolyai, Wolfgang. They were not just at the same university at the same time, they were both studying mathematics and they were good friends:

> 'Gauss met Wolfgang Bolyai, the young Hungarian student who became the most intimate friend of his entire life. Needless to say, the theory of parallels was one of their chief mutual interests.' (reference 4, p. 176)

Thus although the author of this quotation maintained that the younger Bolyai discovered non-Euclidean geometry quite independently of Gauss, the truth could well be different. The elder Bolyai did not let his interest in mathematics stop when he left Göttingen. He thought that he had found a proof of the parallel postulate in

1799 (Gauss set him straight on that) and he continued to work on it after he left Göttingen to become a professor of mathematics at a Hungarian university. That is where his son, Janos Bolyai, grew up. It would not be hard to construct another imaginary dialogue between father and son to make it credible that the younger Bolyai learned of Gauss's ideas. I think it not unlikely; but even though he may not have been its originator, Janos Bolyai deserves great credit for being one of the first to work out the consequences of non-Euclidean ideas.

When Gauss was sent the work of the younger Bolyai, he wrote

> 'I will add that I have recently received from Hungary a little paper on non-Euclidean geometry in which I rediscover all my own ideas and results worked out with great elegance.... The writer is a very young Austrian officer, the son of one of my early friends, with whom I often discussed the subject in 1798.' (reference 8, p. 337)

Gauss thought that he was being the soul of tact by including 'worked out with great elegance' but he would not let tact interfere with his habit of being strictly accurate and objective in all things. It is not surprising that when Bolyai read this—'little paper ... all my own ideas and results'—he was disappointed and angry, and dislike of Gauss was with him for the rest of his life.

Gauss has not been given proper credit by historians because he never published anything on non-Euclidean geometry. His motto was *Pauca sed matura* (Few but ripe), and he would not publish anything until it was up to his lofty standards of what is ripe. It is a motto which many other writers of mathematics might well adopt. Gauss feared controversy if he published anything that would go against the prevailing opinion that Euclidean geometry was the only possible one. He liked a quiet, well-regulated life—he never left Göttingen for another position, though he had many chances to do so—shielded from controversies, stresses, and distractions, and thanks to the care of his two wives, he enjoyed such a life. He should have published, for if he did, the mathematical world would have taken notice of non-Euclidean geometry more quickly than it did, and mathematics would have advanced that much more rapidly. Even though he did not publish, if it is true that great mathematical discoveries are made by great mathematicians, then it seems likely that Gauss may have been the first discoverer of non-Euclidean geometry.

References

1. E. T. Bell, *Men of Mathematics* (Simon & Schuster, New York, 1937).
2. E. T. Bell, *The Last Problem* (Simon & Schuster, New York, 1961).
3. Roberto Bonola, *Non-Euclidean Geometry* (Open Court, Chicago, 1912; reprinted by Dover, New York, 1955).
4. G. Waldo Dunnington, *Carl Friedrich Gauss: Titan of Science* (Hafner, New York, 1955).
5. Edna E. Kramer, *The Main Stream of Mathematics* (Oxford, New York, 1951).
6. National Council of Teachers of Mathematics, Historical Topics for the Mathematics Classroom, Washington D.C. (1969).
7. Constance Reid, *A Long Way from Euclid* (Oxford, New York, 1951).
8. David Eugene Smith, *History of Mathematics* (2 volumes) (Ginn, Boston, 1925; reprinted by Dover, New York, 1953).

The Continuum Hypothesis

PAT ROGERS
Goldsmiths' College, London

In an article in the last volume of *Mathematical Spectrum*, on the Parallel Axiom (reference 2), I attempted to show how the careful modern axiomatic approach to mathematics illuminates and effectively disposes of a problem that had taxed mathematicians for over 2000 years. A similar situation arose more recently in the theory of infinite sets and has only within the last ten years been resolved, again in an altogether surprising manner.

Catherine Smallwood, writing in *Mathematical Spectrum* on 'Infinities' (reference 3), described the notion of an infinite set and explained that, in a certain sense, some sets are more infinite than others. This discovery, made by Georg Cantor in the 1870's, was the first non-trivial one in set theory and heralded some of the most exciting and philosophical mathematical activity to this day.

For the sake of completeness I shall review some of the ideas explained in reference 3. The basis of that discussion was the technique for comparing the sizes of two sets. In the case of finite sets one simply counts up the elements in each set and compares the totals. An alternative procedure, which does not involve the ability to count, is to match up the elements in the two sets one by one—if this can be done, then the sets have the 'same size'. The advantage of the latter method is that we may also use it to compare the sizes of infinite sets. More precisely, two sets A and B are said to be *similar* if their elements can be grouped in pairs (a, b), with a in A and b in B, in such a way that each element of A and B appears exactly once. This pairing is called a *one–one correspondence* between A and B.

Thus, the set N of natural numbers and the set S of squares of natural numbers are similar via the one–one correspondence:

$$\begin{array}{cccccccccccc}
N: & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & \dots \\
 & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \\
S: & 1 & 4 & 9 & 16 & 25 & 36 & 49 & 64 & 81 & 100 & \dots
\end{array}$$

Finite sets may now be defined as sets which are similar to $\{1, 2, \dots, k\}$, for some k in N; a set which is not finite is called infinite. S and N are both examples of infinite sets.

Let A be any set similar to N and, using some one–one correspondence between A and N, denote by a_n the element of A which corresponds to n in N. We may now arrange the elements of A in the infinite sequence $a_1, a_2, a_3, \dots$. For this reason we call sets which are similar to N *countably infinite*. Many surprising sets fall into this category; for instance the set Q of fractions does so.

This whole exercise is interesting only if there are sets which are not similar to N. Using his famous diagonal argument (see reference 3, p. 13), Cantor was able to show that the set R of real numbers (which is also called the *continuum*) is, in fact, not similar to N. Since N is similar to a subset of R (for instance, N itself, or Q), we deduce that N is 'smaller' than R. It turns out, actually, that the countably infinite sets are the smallest infinite sets (reference 3, p. 14). Now that we have started the climb up the ladder of 'infinities' it is not too hard to show that we will never reach the top! If A is any set, the *power set* of A, $P(A)$, is defined to be the set of all subsets of A. Cantor proved that $P(A)$ is not similar to A. (This is the last theorem in reference 3. The set $P(A)$ is there denoted by Σ). It follows from this that $P(A)$ is 'larger' than A. In particular, it can be shown that R is similar to $P(N)$, and this yields an alternative proof of the fact that R is larger than N. Now just as we obtained $P(A)$ from A, we can next form $P(P(A))$ and so on, as long as we please. Thus we have an ever increasing chain of infinite sets with the countably infinite sets at the bottom.

A question arises at this point. Is every infinite set similar to one of the sets

$$N,\ P(N)\ (\text{or } R),\ P(P(N)),\ P(P(P(N))), \ldots ?$$

More particularly, are there infinite sets 'between' N and $P(N)$ (or R), i.e., are there infinite subsets of the set R which are neither countably infinite nor similar to R? Cantor tried, without success, to find such a subset and eventually conjectured, in what became known as the *Continuum Hypothesis*, that none existed.

According to Cantor a set was 'any collection into a whole of definite and separate objects of our intuition or our thought'. But this seemingly harmless formulation turned out to conceal many problems and paradoxes, the most famous being Russell's*. Thus Cantor's attempt to rescue the infinite from the vague and mystical ideas of the past was seen to be based on rather shaky foundations. Evidently, to solve Cantor's very difficult problem, one could no longer rely on such an informal notion of a set. At this stage an analogy with geometry is a good one. Euclid had tried to give Greek geometry a firm foundation by means of an axiomatic treatment, but a critical examination of the system revealed many flaws; and it was the discovery of these that alerted mathematicians to the necessity of being absolutely clear about their assumptions. In fact, the analogy goes much further, as we shall see.

The dictionary defines axiom as 'an established principle, a self-evident truth'. We may assume that Euclid had such a definition in mind. The axioms he chose were certain properties satisfied by lines and points; only statements which could be obtained by logical deduction from these axioms were regarded as theorems. (A discussion of this system and its modern replacement appears in reference 2.) The study of Euclid's geometry led, in spite of its faults, to a new understanding of the nature of an axiom and consequently helped to launch an entirely new branch

* A set may or may not be a member of itself. Since most sets are not, we call these *normal*. An example of a set which is not normal is provided by the set consisting of all infinite sets, for clearly there are infinitely many infinite sets. Now consider the set A of all normal sets. Is this set normal or not? If A is normal, then, by the definition of normality, A does not belong to A, i.e., A is not normal, and we have a contradiction. On the other hand, we are also led to a contradiction if we suppose that A is not normal.

of mathematics (mathematical logic). We shall give just the briefest glimpse of that theory.

No attempt is made, in a modern axiom system (or *theory*), to define the primitive terms and relations as Euclid did. For example, in Hilbert's theory of Euclidean geometry (reference 2, p. 61), the words 'point', 'line' and 'between', among others, are undefined. Instead we regard them as marks or meaningless symbols on paper, e.g., $\otimes$, $\triangle$, $\overset{\uparrow}{\wedge}$,.... Certain finite sequences of these symbols are set aside as the *expressions* of the system (just as only certain combinations of words in the English language are recognized as forming sentences). Two subclasses of these expressions are selected: one contains the *axioms* and the other the *logical rules of inference*. A *proof* is a list $E_1, \ldots, E_n$ of expressions such that, for each i, E_i is either an axiom or a direct consequence of some previous expressions in the list by one of the rules of inference. A *theorem* is then the final expression E_n in such a proof. These are the only expressions which we accept as theorems. Of course, in the present example we still wish to study geometry. Thus our choice of axioms will not be arbitrary, but will be suggested by our intuitive notions of how lines and points behave. But it should be clear from the definitions above that this is as far as we may go with intuition! The fact that objects exist which satisfy the axioms is irrelevant to the process of formal juggling by which we prove theorems. On the other hand, it is obviously important that the axioms do not lead to contradictions, i.e., there should be no expressions E for which both E and its negation are theorems. A non-contradictory set of axioms is called *consistent*.

Now, expressions in a theory derive meanings only when the symbols (i.e., $\otimes$, $\triangle$, $\overset{\uparrow}{\wedge}$,...) are given interpretations. If this can be done in such a way that all the theorems turn out to be true, the interpretation is called a *model* for the theory. It then follows that the axioms are consistent. (To prove this, suppose they are not. We should then be able to derive a contradiction; so there exists an expression E such that both E and its negation are theorems. Thus E is true and false in any model of the theory—but this is absurd!)

The axiom which precipitated the above ideas was the Parallel Axiom: 'through a given point, exactly one parallel may be drawn to a given straight line'. The difficulty here was that this axiom did not have the intuitive quality shared by the other axioms (let us agree to call the collection of these other axioms absolute geometry). Mathematicians believed for over two thousand years that the Parallel Axiom was in fact a theorem of absolute geometry and therefore redundant as an axiom of Euclidean geometry. As we saw in reference 2, they were very wrong. Models of absolute geometry were found in which the Parallel Axiom was false! Since the axiom is also consistent with absolute geometry, its negation cannot be proved either. Thus the Parallel Axiom is *independent* of absolute geometry.

It may come as a surprise to learn that the status of the Continuum Hypothesis in set theory is exactly the same as that of the Parallel Axiom in geometry. As a reminder we state Cantor's guess once more.

The Continuum Hypothesis (*CH*). Every infinite subset of R is either countably infinite or similar to R.

Common sense would assert unhesitatingly that the Continuum Hypothesis is either false or true: either there exists an infinite subset of R which is similar neither to R nor to N, or else there is no such set. But then mathematicians before the mid-nineteenth century thought in the same way about geometry—they could not conceive of a situation in which the Parallel Axiom did not hold. And the reason for this was their persistence in interpreting the most elementary notions of geometry, such as point and line, in accordance with pure intuition: not only a vague practice, but also a very restrictive one. In the case of set theory Bertrand Russell showed that a vague, intuitive idea about sets only led to difficulties. So, since the Continuum Hypothesis is about sets, the important thing is to be precise in one's conception of and language about sets. Experience showed that nothing but the full-blooded axiomatic approach to set theory would do.

The first system of axiomatic set theory was Ernst Zermelo's (1908). This was later refined by Abraham Fraenkel and is known in the trade as ZF set theory. There is only one primitive term: 'set'; and there is one primitive relation: 'is a member of'. Examples of a few of the axioms follow.

Axiom of Extensionality. Two sets are equal if and only if they have the same members.

Axiom of the Null Set. There exists a set with no members.

Axiom of the Power Set. For every set X, there exists the set $P(X)$ of all subsets of X.

The stage was set (pardon the pun) for a proof or disproof of CH. However, it was another thirty years before even a partial answer to Cantor's problem was provided, and when the answer was complete, it turned out that CH was *independent* of ZF. In other words, we are free to take our pick: we can add CH as an *axiom* to ZF and so obtain the kind of set theory that Cantor envisaged; or we can leave ZF without the addition of CH, in which case we must be prepared for situations where CH is false.

In 1938 the great mathematician Kurt Gödel showed that CH could not be disproved from ZF. This he did by building a model, called the *Constructible Universe*, in which all the axioms of Cantorian set theory (i.e., ZF+CH) are true. To prove CH independent of ZF it remained to construct a model of ZF in which CH was false, i.e., a model of non-Cantorian set theory. This would show that CH was not a theorem of ZF. Clearly such a model would have to be very different from Gödel's. But for a long time it was very difficult to imagine any other.

The credit for completing the independence proof goes to Paul Cohen. In 1963 he invented a powerful new method, called *forcing*, which enabled him to construct a model of non-Cantorian set theory. Anyone interested in a non-technical description of the Forcing Method should refer to the popular account by Paul Cohen and Reuben Hersh in *Scientific American* (reference 1).

The history of non-Euclidean geometry has taught us some humility; ideas that first met with derision were later proved to be necessary precursors of Einstein's theory of relativity. So far no application has been found for non-Cantorian set theory—but who can say what the future may hold?

References

1. Paul Cohen and Reuben Hersh, Non-Cantorian set theory. *Scientific American* **217** (6) (1967), 104–116.
2. Pat Rogers, The parallel axiom. *Mathematical Spectrum* **5** (1972/73), 58–66.
3. Catherine V. Smallwood, Infinities. *Mathematical Spectrum* **5** (1972/73), 11–15.

A Matrix Method for Solving Cubic Equations Numerically

JOAN M. HOLLAND

1. Recapitulation

In a previous issue of the 1972/3 *Mathematical Spectrum* (Vol. 5 No. 2, pp. 48–53) cubic equations were solved numerically by the use of recurrence equations. This article presents a closely related but more powerful method of obtaining the same solutions. But first, we provide a brief recapitulation as a source of reference for the ensuing sections.

1.1 A study was made of the particular cubic

$$x^3 = x^2+2x-1 \tag{1}$$

whose roots α, β, γ were found to be approximately

$$\alpha \simeq 0{\cdot}4450, \quad \beta \simeq -1{\cdot}247, \quad \gamma \simeq 1{\cdot}802.$$

The computer calculations involved in obtaining these solutions are listed in Table 1.

TABLE 1

n	a_n	b_n	c_n	α_n	β_n	a_{n+1}/a_n
3	1	2	−1			
4	3	1	−1			
5	4	5	−3			
6	9	5	−4	0·4444	−1·000	
7	14	14	−9	0·4449	−1·445	
8	28	19	−14	0·4450	−1·124	
9	47	42	−28	0·4450	−1·339	
10	89	66	−47	0·4450	−1·187	1·782
11	155	131	−89	0·4450	−1·290	1·846
12	286	221	−155	0·4450	−1·220	1·773
13	507	417	−286	0·4450	−1·267	1·823
14	924	728	−507	0·4450	−1·230	1·788
15	1652	1341	−924	0·4450	−1·256	1·812

1.2 The coefficients a_3, b_3, c_3 are those of the right-hand side of (1) while $a_4 = 3$, $b_4 = 1$, $c_4 = -1$ were obtained as follows. Since α is a root of (1), its fourth power is

$$\begin{aligned}\alpha^4 &= \alpha(\alpha^2+2\alpha-1) = \alpha^3+2\alpha^2-\alpha\\ &= (\alpha^2+2\alpha-1)+2\alpha^2-\alpha = 3\alpha^2+\alpha-1.\end{aligned}$$

All higher powers of α may be reduced in this way to a residual quadratic expression in α

$$\alpha^n = a_n\alpha^2+b_n\alpha+c_n. \tag{2}$$

1.3 a_{n+1}/a_n tends to a limit and that limit is the largest root γ. The residual quadratic equation

$$a_n x^2+b_n x+c_n = 0 \tag{3}$$

derived from (2) has for its roots α_n and β_n where the asymptotic approach of $\alpha_n \to \alpha$ and $\beta_n \to \beta$ is clearly visible in Table 1.

1.4 The same method may be applied to the general cubic

$$x^3 = px^2+qx+r. \tag{4}$$

If a_{n+1}/a_n tends to a limit, this is the numerically largest root of the equation; the two smaller roots appear as the limiting values of the roots α_n, β_n of the residual quadratic equation (3). The coefficients of this equation are given by recurrence formulae:

$$a_n = pa_{n-1}+qa_{n-2}+ra_{n-3}, \tag{5}$$

$$b_n = qa_{n-1}+ra_{n-2}, \tag{6}$$

$$c_n = ra_{n-1}. \tag{7}$$

1.5 For Equation (1) these formulae reduce to

$$a_n = a_{n-1}+2a_{n-2}-a_{n-3}, \tag{8}$$

$$b_n = 2a_{n-1}-a_{n-2}, \tag{9}$$

$$c_n = -a_{n-1}. \tag{10}$$

Formula (8) generates the recurrence sequence

$$\{\ldots, -25, -11, -5, -2, -1, 0, 0, 1, 1, 3, 4, 9, \ldots\}$$

which we refer to as $\{a_n\}$.

2. A matrix method for the same results

The successive powers of the matrix

$$F = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \tag{11}$$

develop the terms of the well-known Fibonacci sequence

$$\{\ldots, 2, -1, 1, 0, 1, 1, 2, 3, 5, 8, \ldots\}.$$

Each of the terms is the sum of the previous two, and powers of this matrix have the form

$$F^2 = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}, \quad F^3 = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}, \quad F^4 = \begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix}, \quad F^8 = (F^4)^2 = \begin{pmatrix} 13 & 21 \\ 21 & 34 \end{pmatrix}. \tag{12}$$

Because the second row of F^n is the first row of F^{n+1}, all these can be conveniently combined into the table of overlapping matrices shown below.

$$\begin{array}{rcl} F = & \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} & \\ & \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} & = F^2 \\ F^3 = & \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} & \\ & \begin{bmatrix} 2 & 3 \\ 3 & 5 \end{bmatrix} & = F^4 \\ & \cdots & \end{array} \tag{13}$$

The Fibonacci sequence is also related to the quadratic equation

$$x^2 = x+1 \tag{14}$$

and applying the reduction method of Section 1, we see that

$$x^3 = 2x+1, \quad x^4 = 3x+2, \quad x^5 = 5x+3. \tag{15}$$

Plainly the coefficients of these residual linear expressions follow the same pattern as the rows of the matrices in (13). The matrix F has a characteristic equation and it is no accident that this turns out to be Equation (14).

The outstanding advantage of F is that by repeated squaring to obtain $F^2, F^4, F^8, \ldots, F^{2n}, \ldots$ we can derive numbers far along the sequence without going through the step-by-step process of the recurrence formulae. Thus for F as in (11) suppose that we require F^{27}; then since $F^{27} = F^{16}.F^8.F^2.F$, we can obtain, by rapid calculations relying only on powered matrices of type F^{2n} $(n = 0, 1, 2, \ldots)$ the result

$$\begin{aligned} F^{27} &= F^{16} \,.\, F^8 \,.\, F^2 \,.\, F \\ &= \begin{pmatrix} 610 & 987 \\ 987 & 1597 \end{pmatrix}\begin{pmatrix} 13 & 21 \\ 21 & 34 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 121{,}393 & 196{,}418 \\ 196{,}418 & 317{,}811 \end{pmatrix}. \end{aligned} \tag{16}$$

It is clear that higher terms of the sequence are obtained much more rapidly in this way than by the recurrence formula.

These considerations suggest that it might be profitable to look for a similar 3×3 matrix associated with the particular cubic $x^3 = x^2+2x-1$ to obtain a rapid build-up of the sequence $\{a_n\}$.

The first two rows of a suitable matrix M may be written down at once as

$$M = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ . & . & . \end{pmatrix}. \tag{17}$$

This is appropriate because, if used to multiply any other matrix A, the last 2 rows of this will overlap with the first 2 rows of MA, as follows

$$A = \begin{pmatrix} * & * & * \\ l & m & n \\ u & v & w \end{pmatrix}, \quad MA = \begin{pmatrix} l & m & n \\ u & v & w \\ * & * & * \end{pmatrix}. \tag{18}$$

Repeated multiplication by M will produce a series of matrices each overlapping its neighbours by two rows.

It remains to choose the bottom row of M so that it will also produce the sequence $\{a_n\}$ which is generated by Equation (8) whose coefficients are $1, 2, -1$.

It is readily verified that the appropriate form of M is

$$M = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & 2 & 1 \end{pmatrix}, \tag{19}$$

whose first few powers are

$$M^2 = \begin{pmatrix} 0 & 0 & 1 \\ -1 & 2 & 1 \\ -1 & 1 & 3 \end{pmatrix}, \quad M^3 = \begin{pmatrix} -1 & 2 & 1 \\ -1 & 1 & 3 \\ -3 & 5 & 4 \end{pmatrix},$$

$$M^4 = \begin{pmatrix} -1 & 1 & 3 \\ -3 & 5 & 4 \\ -4 & 5 & 9 \end{pmatrix}, \quad M^8 = \begin{pmatrix} -14 & 19 & 28 \\ -28 & 42 & 47 \\ -47 & 66 & 89 \end{pmatrix}. \tag{20}$$

These results are illuminating and contain an agreeable surprise. We have met all these numbers before. The last row of M was chosen to make the third column of its powers follow the sequence $\{a_n\}$ leading to the largest root γ as the limit of a_{n+1}/a_n. In addition we now see that the rows contain exactly the same numbers as those of Table 1 and provide us with the coefficients of the quadratic equations whose roots converge towards α and β. We shall set out the matrices in a combined table (Table 2) under headings a_n, b_n, c_n; we shall also prolong it upwards by using the values for a_n already mentioned in Section 1.5.

TABLE 2

n		c_n	b_n	a_n	
-3		11	3	-5	
-2		5	1	-2	
-1		2	1	-1	
0		1	0	0	M^{-1}
1	I	0	1	0	
2		0	0	1	M
3	M^2	-1	2	1	
4		-1	1	3	
5		-3	5	4	M^4
6		-4	5	9	

We see that the identity matrix is in the centre of the table with overlapping matrices M, M^2,... below it. There seems a strong likelihood that the matrix immediately above I is M^{-1} the inverse matrix; this can be confirmed by verifying that $MM^{-1}=I$. It is perhaps worth noting that if we start with the reciprocal equation

$$x^3 = 2x^2+x-1,$$

then its matrix

$$M' = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & 1 & 2 \end{pmatrix} \tag{21}$$

is the same as M^{-1} with the rows and columns reversed. As might be expected the cubic and its reciprocal are the characteristic equations of M and M' respectively.

3. The matrix for the general cubic

It is now simple to write down a matrix and its inverse for the general cubic

$$x^3 = px^2+qx+r. \tag{22}$$

These are

$$M = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ r & q & p \end{pmatrix}, \quad M^{-1} = \begin{pmatrix} -q/r & -p/r & 1/r \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}. \tag{23}$$

Without going into details, we shall outline the proof that M and its powers lead to the same results as the recurrence formulae of Section 1.

First let us assume that

$$M^n = \begin{pmatrix} c_n & b_n & a_n \\ c_{n+1} & b_{n+1} & a_{n+1} \\ c_{n+2} & b_{n+2} & a_{n+2} \end{pmatrix}. \tag{24}$$

Then

$$M \times M^n = \begin{pmatrix} c_{n+1} & b_{n+1} & a_{n+1} \\ c_{n+2} & b_{n+2} & a_{n+2} \\ C & B & a_{n+3} \end{pmatrix} \tag{25}$$

where

$$C = rc_n + qc_{n+1} + pc_{n+2}. \tag{26}$$

If we now make use of Equations (7) and (5), we see that

$$\begin{aligned} C &= rra_{n-1} + qra_n + pra_{n+1} \\ &= r\{pa_{n+1} + qa_n + ra_{n-1}\} \\ &= ra_{n+2} \\ &= c_{n+3}. \end{aligned} \tag{27}$$

Similarly we can show that $B = b_{n+3}$. Hence if M^n is of the form assumed in Equation (24) M^{n+1} will also be of the same form, with n replaced by $n+1$ throughout the matrix. We have already seen that M itself is of the correct form; thus by induction it follows that this form persists through the lower half of Table 2. A similar treatment for M^{-1} takes care of the upper half of the table.

4. The eigenvectors of M

A further interesting feature of M is the form of its eigenvectors. They are

$$\begin{pmatrix} 1 \\ \alpha \\ \alpha^2 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ \beta \\ \beta^2 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 \\ \gamma \\ \gamma^2 \end{pmatrix}. \tag{28}$$

To verify this statement, consider the effect of multiplying the first vector by M. We obtain

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ r & q & p \end{pmatrix} \begin{pmatrix} 1 \\ \alpha \\ \alpha^2 \end{pmatrix} = \begin{pmatrix} \alpha \\ \alpha^2 \\ p\alpha^2 + q\alpha + r \end{pmatrix} = \begin{pmatrix} \alpha \\ \alpha^2 \\ \alpha^3 \end{pmatrix} = \alpha \begin{pmatrix} 1 \\ \alpha \\ \alpha^2 \end{pmatrix} \tag{29}$$

with similar results for the two others.

5. A further numerical example

To illustrate the power of the method, let us consider a second example. The equation $x^3 - 3x^2 - 5x + 4 = 0$ may be rewritten as $(x^2 + x - 1)(x - 4) = 0$. In this case the largest root γ is 4 while α and β are the roots of $x^2+x-1=0$. Here

$$M=\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -4 & 5 & 3 \end{pmatrix},\quad M^2=\begin{pmatrix} 0 & 0 & 1 \\ -4 & 5 & 3 \\ -12 & 11 & 14 \end{pmatrix},$$

$$M^4=\begin{pmatrix} -12 & 11 & 14 \\ -56 & 58 & 53 \\ -212 & 209 & 217 \end{pmatrix}. \tag{30}$$

Already, after only 2 powerings, it can be seen that $217/53 \simeq 4{\cdot}1$ while $217x^2+209x-212=0$ is close to the equation $x^2+x-1=0$. Convergence for this equation is very rapid but this is not true of all cubic equations. The method is generally much more suitable for a computer than for human calculation.

This approach has been successfully applied to quartic equations, and is evidently capable of extension to some equations of higher degree. A much more thorough investigation would be required to examine problems arising from equal roots and pairs of complex roots.

A New Look at the S.H.M. Equation

DIETER K. ROSS AND EDWIN F. CARTER

La Trobe University, Victoria, Australia

1. Introduction

Most textbooks dealing with elementary applied mathematics contain a discussion of the solution of the equation of simple harmonic motion (S.H.M.). An equation of this type can be used to describe the small amplitude oscillations of a simple pendulum,

$$\ddot{\theta}+g/l\theta=0, \tag{1}$$

where θ is the angle of the swing, l is the length of the pendulum, g is the gravitational constant and a dot denotes differentiation with respect to time; or the motion of a simple elastic spring,

$$\ddot{y}+n^2y=0,$$

where y is its displacement from the static equilibrium position and n is a constant depending on the elastic modulus of the spring and on the weight it supports.

Although these equations are very simple in form and the solution is known most authors solve the S.H.M. equation by using a fallacious argument. It is the purpose of this paper, first of all, to point out some of these errors and then to give a rigorous method of solution which is comprehensible to the present sixth form student.

2. Definitions

To begin with we remind the reader of the following three definitions.

(a) A function is thought of as a rule which assigns to every value of x, a *corresponding* and *unique value* of y. This we conventionally write in the form

$$y = F(x).$$

(b) The function F is said to be differentiable at x if

$$\lim_{h \to 0} \frac{F(x+h)-F(x)}{h} = F'(x), \quad \text{say},$$

exists.

(c) The most general second-order differential equation can be written in the form

$$G(x, y, y', y'') = 0, \tag{2}$$

where y is regarded as a function of x and the dash denotes differentiation. We are now in a position to say that the function $w(x)$ is a solution of (2) if $w(x)$ is twice differentiable and if

$$G(x, w(x), w'(x), w''(x)) = 0$$

is an identity for all x under consideration. This may seem obvious but it is important to be careful. A much more formal set of definitions than those given here can be found in the rather more advanced texts of Dieudonné (reference 1) and Petrovski (reference 2).

3. The S.H.M. equation

In order to illustrate the importance of the definitions given in the above section we shall consider the solution of the differential equation which describes the motion of a particle in S.H.M. subject to the initial conditions

$$y = 0, \quad v = \dot{y} = 1, \quad \text{at time } t = 0. \tag{3}$$

This can be written as the second-order differential equation

$$\ddot{y}+y = 0, \tag{4}$$

where, for simplicity, we have chosen the coefficients to equal one. If we now assume that the chain rule for differentiation applies we obtain

$$\frac{d^2y}{dt^2} = \frac{dv}{dt} = \frac{dv}{dy}\frac{dy}{dt} = v\frac{dv}{dy}. \tag{5}$$

Then, substituting (5) into (4) it is clear that we can integrate with respect to y and so obtain (by using the initial conditions in (3))

$$\dot{y}^2+y^2 = \text{constant} = 1. \tag{6}$$

However, this assumption is incorrect for, although both the displacement y and the velocity v are functions of the time t, it is not true that v is a function (in the sense of (a)), of y. Hence we are not at liberty to use (5) without extra consideration. Nevertheless we shall proceed.

Now for Equation (6) to have a real solution it is necessary that $-1 \leqslant y \leqslant 1$ and in this range one possibility is to write

$$\dot{y} = +(1-y^2)^{\frac{1}{2}} \geqslant 0 \tag{7}$$

which implies that y does not decrease with time.

One solution of this equation which satisfies the initial conditions in (3) is easily seen to be

$$y_1(t) = \sin t$$

which on account of the inequality in (7) is only valid for $-\frac{1}{2}\pi \leqslant t \leqslant \frac{1}{2}\pi$. At the ends of this range we may continue the solution so that

$$y_1(t) = -1 \quad \text{for } t \leqslant -\tfrac{1}{2}\pi$$

and at the other end we may choose

$$y_1(t) = 1 \qquad \text{for } t \geqslant \tfrac{1}{2}\pi.$$

Thus a solution of (7) which is valid for all finite values of t is

$$y_1(t) = \begin{cases} -1, & -\infty < t \leqslant -\frac{1}{2}\pi, \\ \sin t, & -\frac{1}{2}\pi < t < \frac{1}{2}\pi, \\ 1, & \frac{1}{2}\pi \leqslant t < \infty. \end{cases}$$

Notice that Equation (6) cannot be replaced by

$$\dot{y} = -(1-y^2)^{\frac{1}{2}} \leqslant 0,$$

because no solution of this equation can satisfy the initial condition $\dot{y} = +1$ at $t = 0$.

On the other hand Equation (6) has infinitely many solutions including the obvious

$$y_2(t) = \sin t \quad \text{for all } t.$$

We can also start with the solution $y_1(t)$ and modify it as follows:

$$y_3(t) = \begin{cases} -1, & -\infty < t \leqslant -\frac{1}{2}\pi, \\ \sin t, & -\frac{1}{2}\pi < t < \frac{1}{2}\pi, \end{cases}$$

and, for higher values of t, add more sine waves. Thus

$$y_3(t) = \begin{cases} 1, & \frac{1}{2}\pi \leqslant t \leqslant 19\pi/2, \\ -\sin t, & 19\pi/2 < t < 25\pi/2, \\ -1, & 25\pi/2 \leqslant t < \infty, \end{cases}$$

is a solution of Equation (6).

At this stage the student should check that the function $y_3(t)$ is continuous everywhere, and perhaps construct other solutions of (6). It is a fairly simple matter to show that the functions $y_1(t)$ and $y_3(t)$ are once differentiable in the sense of (b); but they are not twice differentiable at critical points like $t = -\frac{1}{2}\pi$. From the definition of solution given in (c) it is clear that $y_1(t)$ and $y_3(t)$ are not solutions of the S.H.M. equation for all time. The error lies in applying the chain rule (5) in a situation where the velocity $v = \dot{y}$ is not only a function of y but also of the direction of motion. Hence the turning points are critical.

This explanation is not immediately obvious for it would seem as though these functions fail because they are not twice differentiable. That this is not so can be seen by considering the initial value problem $y = 1$, $\dot{y} = 0$ at $t = 0$. In this case $y_4(t) = 1$ is a twice differentiable solution of (6) but it is not a solution of the S.H.M. equation.

4. A simple method of solution

The following method is due to Brown (reference 3). To begin with the S.H.M. equation may be replaced by the pair of equations

$$(\ddot{y}+y)\cos t = 0$$

and

$$(\ddot{y}+y)\sin t = 0,$$

because $\sin t$ and $\cos t$ do not vanish simultaneously. These equations can be integrated to give

$$\dot{y}\cos t + y\sin t = \text{constant} = A, \quad \text{say,}$$

and

$$\dot{y}\sin t - y\cos t = \text{constant} = B, \quad \text{say.}$$

Then on solving these equations for y and $\dot{y}$ we obtain

$$y = A\sin t - B\cos t$$

with a similar expression for $\dot{y}$.

References

1. J. Dieudonné, *Foundations of Modern Analysis* (Academic Press, New York, 1969).
2. I. G. Petrovski, *Ordinary Differential Equations* (Prentice-Hall, Englewood Cliffs, N.J., 1966).
3. A. Brown, The S.H.M. equation: an alternative treatment. *The Australian Mathematics Teacher* **23** (1967), 14–16.

A Non-linear Differential Equation arising in Relativity Theory

A. SACKFIELD AND M. G. FROST
Trent Polytechnic, Nottingham

1. Introduction

Whilst examining a problem in general relativity the differential equation

$$f\left\{\frac{d^2f}{dr^2}+\frac{1}{r}\frac{df}{dr}\right\} = \left(\frac{df}{dr}\right)^2-k^2 \quad (k \text{ constant}) \tag{1.1}$$

arose. Although this differential equation at the time appeared to have no physical significance, it was decided to examine its solution, possibly as a suitable tutorial example. The solution proved very interesting.

A solution in series approach surprisingly yielded an infinite sequence of particular solutions of easily generalisable form. The form of these particular solutions suggested that the general solution could be obtained by transformation of variable techniques. This proved to be so and the general solution of (1.1) was thus obtained and was easily seen to be a generalisation of the previously obtained particular solutions.

Having solved the differential equation (1.1) it was observed that it could, in fact, be obtained from a differential equation of the form

$$\frac{\partial^2\Phi}{\partial x^2}+\frac{\partial^2\Phi}{\partial y^2}+\lambda^2 h\, e^{2h\Phi} = 0, \tag{1.2}$$

where $\Phi(x, y)$ is a function of two independent variables. This equation is known as Liouville's equation (reference 1), and its solution has previously been obtained. It has not, however, been obtained by such elementary techniques nor without reference to physical considerations.

It was further noticed on investigation that (1.1) could be generalised to the form

$$f\left\{\frac{d^2f}{dr^2}+\frac{1}{r}\frac{df}{dr}\right\} = \left(\frac{df}{dr}\right)^2-k^2r^{2\alpha} \quad (\alpha \text{ constant}) \tag{1.3}$$

and a solution easily obtained by a further simple transformation of variable.

Although the main interest in the above work centred on the nature of the solution of the differential equation, it has in fact physical applications which arise through its connection with Liouville's equation. Two such applications are in the theory of nebulae and, more recently, in the theory of thermal explosions.

The work is presented in chronological order to show the way in which it developed.

2. The origin of (1.1)

The following coupled partial differential equations occur in Einstein's theory of general relativity:

$$f\left\{\frac{\partial^2 f}{\partial r^2}+\frac{1}{r}\frac{\partial f}{\partial r}\right\}=\left(\frac{\partial f}{\partial r}\right)^2-\left(\frac{\partial g}{\partial z}\right)^2, \tag{2.1}$$

$$\frac{\partial^2 g}{\partial z^2}=0. \tag{2.2}$$

In these equations f and g are related to the gravitational potential and angular momentum respectively of a rotating body, and r and z have their usual meaning in cylindrical polar co-ordinates (r, ϕ, z).

The equations are obtained from the general system of Einstein's equations by assuming that f is a function of r only and g is a function of z only. In general, of course, both f and g must be assumed to be functions of the time co-ordinate and three spatial co-ordinates.

Now, (2.2) is immediately solvable:

$$g = kz + l,$$

where k and l are arbitrary constants.

Substituting this result into (2.1) gives the differential equation (1.1) for f in terms of r only.

3. A series solution of (1.1)

A first approach, although unusual for a non-linear equation, is to try a series solution. The essence of this technique (the Frobenius method) is to assume a series solution of the form

$$f=\sum_{n=0}^{\infty} a_n r^{n+m}, \quad a_0 \neq 0, \tag{3.1}$$

where a_n $(n = 0, 1, \ldots)$ and m are constants to be determined. Substituting this series representation of f into (1.1) and simplifying, gives

$$\sum_{n=0}^{\infty}\sum_{j=0}^{n} a_n a_j (j+m)(j-n) r^{n+j+2m-1} = -rk^2. \tag{3.2}$$

If (3.1) is to be a solution of (1.1) for all r, then (3.2) needs to be identically true. This can only be so if all the coefficients, except one, on the left-hand side of (3.2) vanish, the exception being the coefficient of r which must equal $-k^2$. Hence equating coefficients of r^{2m} (the lowest power of r which occurs on the left-hand side of (3.2)) gives

$$a_0 a_1 = \begin{cases} 0, & m \neq \frac{1}{2}, \\ -k^2, & m = \frac{1}{2}. \end{cases}$$

If $m=\frac{1}{2}$, equating further coefficients, it is easily shown that $a_n=0$, $n\neq 0,1$. This therefore yields a particular solution of (1.1) for $m=\frac{1}{2}$

$$f=a_0 r^{\frac{1}{2}}+a_1 r^{\frac{3}{2}}=r(a_0 r^{-\frac{1}{2}}-k^2 a_0^{-1} r^{\frac{1}{2}}).$$

This solution contains one arbitrary constant a_0.

Now, if $m\neq\frac{1}{2}$, $a_1=0$ since $a_0\neq 0$, and equating coefficients of r^{2m+1} gives

$$4a_0 a_2=\begin{cases}0, & m\neq 0,\\ -k^2, & m=0.\end{cases}$$

Again, if $m=0$ it is easily shown that $a_n=0$, $n\neq 0,2$. This yields another particular solution

$$f=a_0+a_2 r^2=r(a_0 r^{-1}-k^2(4a_0)^{-1} r).$$

Repeating this procedure for $m\neq\frac{1}{2},0$ we have

$$a_0 a_1=0,\quad \text{i.e., } a_1=0,$$
$$a_0 a_2=0,\quad \text{i.e., } a_2=0,$$

and

$$a_0 a_3=\begin{cases}0, & m\neq -\frac{1}{2},\\ -k^2, & m=-\frac{1}{2}.\end{cases}$$

Again if $m=-\frac{1}{2}$, $a_n=0$, $n\neq 0,3$, and the particular solution is

$$f=a_0 r^{-\frac{1}{2}}+a_3 r^{\frac{5}{2}}=r(a_0 r^{-\frac{3}{2}}-k^2(9a_0)^{-1} r^{\frac{3}{2}}).$$

Continuing with this method, if $m=-1$ a particular solution exists of the form

$$f=r(a_0 r^{-2}-k^2(16a_0)^{-1} r^2).$$

From examination of the sequence of particular solutions it is clear that for integer and half integer values of m of the form $m=1-\frac{1}{2}p$, where p is a positive integer, a particular solution of (1.1) exists in the form

$$f=r(a_0 r^{-\frac{1}{2}p}-k^2(p^2 a_0)^{-1} r^{\frac{1}{2}p}). \tag{3.3}$$

This solution is still referred to as a particular solution since p is not completely arbitrary, but is a positive integer. This restriction is imposed by the form of the series solution.

However (3.3) is quickly seen to be a solution of (1.1) for all values of p. The question now arises as to whether such a solution may be obtained by some other method.

4. A general solution of (1.1) by change of variables

The particular solutions of the previous section are all of the form $rF(r)$. This suggests a transformation of dependent variable of the form $f=rF(r)$. The differential equation (1.1) then becomes

$$F\left\{r^2\frac{d^2F}{dr^2}+r\frac{dF}{dr}\right\} = r^2\left(\frac{dF}{dr}\right)^2-k^2,$$

which, with $r = e^t$, reduces to

$$F\frac{d^2F}{dt^2} = \left(\frac{dF}{dt}\right)^2-k^2.$$

On integrating

$$\frac{dF}{dt} = \pm(k^2+c^2r^2)^{\frac{1}{2}} \quad (c \text{ constant}). \tag{4.1}$$

Integrating again

$$\frac{1}{c}\sinh^{-1}\frac{cF}{k} = \pm t+\log\beta \quad (\beta \text{ constant}).$$

On substituting back this gives

$$f = kr(2c)^{-1}\{\beta^c r^{\pm c}-\beta^{-c}r^{\mp c}\} \tag{4.2}$$

or

$$f = krc^{-1}\sinh(c\log\beta r^{\pm 1}). \tag{4.3}$$

This solution contains two arbitrary constants and using (4.2) is clearly the generalised form of the solution (3.3).

One form of the solution which is not immediately apparent occurs if $c = 0$ in (4.1). Then (4.1) integrates to give

$$F = \pm kt+\alpha \quad (\alpha \text{ constant}).$$

On substituting back this yields the particular solution

$$f_1 = r(\alpha\pm k\log r). \tag{4.4}$$

The solution (4.4) is not independent of the solution (4.3) for

$$\begin{aligned}\lim_{c\to 0} f &= \lim_{c\to 0}\{krc^{-1}\sinh(c\log\beta r^{\pm 1})\}\\ &= kr\lim_{c\to 0}\{\log\beta r^{\pm 1}\cosh(c\log\beta r^{\pm 1})\},\end{aligned}$$

using L'Hôpital's rule; i.e.,

$$\lim_{c\to 0} f = kr\log\beta r^{\pm 1},$$

which is f_1 as written in (4.4) with $k\log\beta = \alpha$.

5. (1.1) as a form of Liouville's equation

Transforming from Cartesian co-ordinates (x, y) to polar co-ordinates (r, θ) the Liouville's equation (1.2) becomes (reference 2)

$$\frac{\partial^2\Phi}{\partial r^2}+\frac{1}{r}\frac{\partial\Phi}{\partial r}+\frac{1}{r^2}\frac{\partial^2\Phi}{\partial\theta^2}+\lambda^2 h\, e^{2h\Phi} = 0.$$

Assuming that the solution is axially symmetric (i.e., Φ is a function of r only), this becomes

$$\frac{\partial^2\Phi}{\partial r^2}+\frac{1}{r}\frac{\partial\Phi}{\partial r}=-\lambda^2 h\,e^{2h\Phi}. \tag{5.1}$$

But, letting $f=e^{\Phi}$ in (1.1) yields precisely this differential equation with $h=-1$ and $\lambda=\pm ik$, where $i=\sqrt{-1}$; i.e., the differential equation (1.1) is a form of Liouville's equation. The solution of Liouville's equation assuming axial symmetry has previously been given (reference 3). The solution of (5.1) is thus

$$f=ikrc^{-1}\cosh(c\log b^{-1}r), \tag{5.2}$$

where b and c are arbitrary constants.

This solution is obtained in the literature by making assumptions about the form of solution and using a knowledge of possible physical interpretation of the solution. The solution (5.2) is easily seen to be another form of the solution (4.3) by putting $\beta=b^{\mp1}e^{i\pi/2c}$. Then from (4.3)

$$\begin{aligned} f&=kc^{-1}r\sinh(c\log b^{\mp1}r^{\pm1}e^{i\pi/2c})\\ &=kc^{-1}r\sinh(c\log b^{\mp1}r^{\pm1}+\tfrac{1}{2}i\pi)\\ &=ikc^{-1}r\cosh(c\log b^{-1}r). \end{aligned}$$

6. The solution of (1.3)

Having solved the differential equation (1.1) it is an easy matter to solve the differential equation (1.3).

Putting $f=r^{\alpha}F$, (1.3) reduces to

$$F\frac{d^2F}{dr^2}+\frac{1}{r}\frac{dF}{dr}=\left(\frac{dF}{dr}\right)^2-k^2,$$

which is precisely the differential equation (1.1).

Therefore, the solution of (1.3), which is a generalised form of (1.1), is simply

$$f=kr^{\alpha+1}c^{-1}\sinh(c\log\beta r^{\pm1}). \tag{6.1}$$

The differential equation (1.3) may easily be seen to be a form of a generalised Liouville's equation

$$\frac{\partial^2\phi}{\partial x^2}+\frac{\partial^2\phi}{\partial y^2}+k^2r^{2\alpha}e^{-2\phi}=0. \tag{6.2}$$

7. Physical applications

Although no physical applications of the differential equation (1.1) were immediately apparent, its connection with the Liouville's equation (1.2) shows that in fact it possesses some physical applications. Liouville's equation has been studied in connection with the theory of nebulae (reference 4), but its most recent application has been in the theory of thermal explosions (reference 5). In this theory the mathematical model for the explosion of a gas contained in a vessel is constructed as follows.

Let T be the temperature at a point with cylindrical polar co-ordinates (r, ϕ, z) in the gas, which has thermal conductivity λ. The critical condition for inflammability of the gas is that the heat generated by a reaction QW (where Q is the heat of reaction and W is the reaction velocity) be equal to the heat lost to the surroundings $-\lambda\nabla^2 T$, where (reference 2)

$$\nabla^2 T = \frac{\partial^2 T}{\partial r^2} + \frac{1}{r}\frac{\partial T}{\partial r} + \frac{1}{r^2}\frac{\partial^2 T}{\partial \phi^2} + \frac{\partial^2 T}{\partial z^2}.$$

For a monomolecular reaction W is taken to be proportional to $e^{-E/RT}$ where E is the energy of activation of the reaction and R is the gas constant.

We have therefore

$$\nabla^2 T = -Ke^{-E/RT} \quad (K \text{ constant}). \tag{7.1}$$

If now, the temperature T is almost equal to T_0, the temperature of the walls of the containing vessel, we may expand E/RT as a Taylor series about T_0 to obtain

$$\frac{E}{RT} = \frac{E}{RT_0} - \frac{E}{RT_0^2}(T - T_0), \tag{7.2}$$

neglecting terms of higher than first order in $(T-T_0)$. Putting $\theta = (E/RT_0^2)(T-T_0)$ in (7.2), (7.1) becomes

$$\nabla^2 \theta = -\left[\left(\frac{KE}{RT_0^2}\right) e^{-E/RT_0}\right] e^{\theta}.$$

If the containing vessel is very long and cylindrical with axis the z-axis, the solution is axially symmetric and this equation reduces to one of the form (5.1).

In (7.1) the constant K depends on several physical quantities which have usually been assumed to be constant. The solution of the generalised equation (1.3) and (6.2) however allows K to be replaced by a term of the form $K_0 r^{2\alpha}$ and an exact solution for the temperature distribution obtained. If, further, K is replaced by a term of the form which may be expanded in powers of r or $1/r$ then the behaviour of the solution either close to or at great distances from the axis respectively may be examined. It is conceivable that variations of this form could occur. Confirmation of this would require further study of the chemical physics of the situation.

References

1. W. F. Ames, *Non-linear Partial Differential Equations in Engineering* (Academic Press, New York, 1965), p. 66.
2. D. E. Rutherford, *Vector Methods* (Oliver & Boyd, Edinburgh, 1962).
3. W. F. Ames, *Non-linear Partial Differential Equations in Engineering* (Academic Press, New York, 1965), p. 180.
4. G. W. Walker, Some problems illustrating the forms of nebulae. *Proc. Roy. Soc.* A **91**. (1915), 410–420.
5. P. L. Chambré, On the solution of the Poisson–Boltzmann equation with application to the theory of thermal explosions. *J. Chem. Phys.* **20** (1952), 1795–1797.

Letter to the Editor

Dear Editor,

Fermat's Theorem

Your readers may be interested in a proof of the famous theorem of the French mathematician Fermat. This method of obtaining the proof is different from Fermat's own proof.

Theorem. If a is prime to p and p is a prime number, then

$$a^p - a \equiv 0 \pmod{p}.$$

Proof.

$$\begin{aligned} a^p &= (a-1+1)^p \\ &= (a-1)^p + {}^pC_1(a-1)^{p-1} + \ldots + {}^pC_{p-1}(a-1) + {}^pC_p \\ &\equiv (a-1)^p + 1 \pmod{p} \quad (\text{since } p \text{ is prime and for} \\ &\qquad\qquad\qquad\qquad r < p,\ {}^pC_r \text{ is divisible by } p) \\ &\equiv (a-2)^p + 2 \pmod{p} \\ &\equiv \ldots \\ &\equiv (a-a)^p + a \pmod{p}. \end{aligned}$$

Hence

$$a^p \equiv a \pmod{p}.$$

Yours sincerely,

A. B. PATEL
(V. S. Patel College of Arts and Science,
Bilimora, India)

Reference

S. Barnard and J. M. Child, *Higher Algebra* (Macmillan, London, 1960).

Problems and Solutions

Sixth formers and students are invited to submit solutions to some or all of the problems below: the most attractive solutions will be published in subsequent issues. When writing to the Editorial Office, please state your full name and the postal address of your school, college or university.

Problems

6.5. Show that, given a finite number of points in the plane which do not all lie on the same straight line, there exists a straight line passing through exactly two of the points.

6.6. Let n be a positive integer which is not divisible by 2 or 5. Show that $1/n$ has decimal expansion of the form $0{\cdot}\dot{a}_1 a_2 \ldots \dot{a}_m$.

6.7. (Submitted by R. J. Webster, University of Sheffield.) Show that

$$1+\frac{1}{2}+\frac{1}{3}+\ldots+\frac{1}{n}$$

is not an integer when $n > 1$.

6.8. (Submitted by B. G. Eke, University of Sheffield.) From the series

$$1+\tfrac{1}{2}+\tfrac{1}{3}+\ldots$$

delete every term whose denominator has a '3' in its decimal representation. Show that the modified series is convergent.

Solutions to Problems in Volume 6, Number 1

6.1. The odd integer n can be expressed as a sum of squares in two different ways, say $n = x^2+y^2 = u^2+v^2$, where $x < u$ and x, u are even. Prove that there are positive integers a, r, b, s such that

(i) $u-x = 2ar$, $y-v = 2as$,

(ii) r and s are relatively prime,

(iii) $(x+ar)\,r = (y-as)\,s = brs$.

Deduce that n cannot be prime.

Solution

We have

$$u^2-x^2 = y^2-v^2,$$

i.e.,

$$(u+x)(u-x) = (y+v)(y-v).$$

Now x, u are even and y, v are odd, so $u-x$ and $y-v$ are both even; let their highest common factor be $2a$. Then we can write

$$u-x = 2ar, \quad y-v = 2as,$$

and r, s are relatively prime. This gives (i), (ii). Thus

$$(2x+2ar)\,2ar = 2as(2y-2as)$$

so

$$(x+ar)\,r = (y-as)\,s.$$

Thus $s \mid (x+ar)\, r$, whence $s \mid (x+ar)$, because r, s are relatively prime. If we write $x+ar = bs$, we have

$$(x+ar)\, r = (y-as)\, s = brs,$$

which is (iii). Note that $y-as = br$. Now

$$\begin{aligned} n &= x^2+y^2 \\ &= (bs-ar)^2+(br+as)^2 \\ &= (a^2+b^2)\,(r^2+s^2), \end{aligned}$$

so that n is not prime.

Also solved by David Seal (Winchester College), Michael Beasley (Kingston Grammar School).

6.2. In a gathering of people, some shake hands with others. Show that there are two people who shake hands the same number of times.

Solution by Howard Hiller (Cornell University)

Suppose that there are n people at the gathering. Then each one can shake at most $n-1$ hands. Hence to each of the n people we assign an integer k, $0 \leqslant k \leqslant n-1$. By the 'pigeon-hole principle', the only possible way to do this to avoid the conclusion is to assign the j^{th} person the integer $j-1$ $(1 \leqslant j \leqslant n)$, where we have allowed a renaming of the people. But then the n^{th} person shakes hands with everyone and the first person with no one, an impossible situation.

Also solved by P. R. Knight (University of Exeter), David Seal, N. Richards (Melton Mowbray College of Further Education).

6.3. Show that, for every positive integer n, there is a finite set of points in the plane with the property that every point of the set is distant one unit from exactly n points of the set.

Solution by David Seal

Suppose there is a finite set S of points in the plane with this property for $n = k$. Then one can translate this set of points one unit of distance in any direction to get another set of points T with the same property. It is possible to choose the direction of this translation in such a way that each point of T is at a distance of one unit from only one point of S. Indeed, there are infinitely many possible directions for the translation, and only a finite number are excluded by this condition. But now the union of S and T will be a finite set of points and each point of it will be at a distance of one unit from exactly $k+1$ points of the set. When $n = 0$, we may take as our set a single point. It follows by induction that such a set is possible for all non-negative integers n.

Editorial comment. If we begin with the three vertices of an equilateral triangle of side 1 for the case $n = 2$ and use the method described by David Seal, we see that for $n \geqslant 2$ we can always find a set of 3.2^{n-2} points with the desired property. Professor Paul Erdös of the Hungarian Academy of Sciences has raised the question whether 3.2^{n-2} can be replaced by a smaller number. He suspects that this is the case; but the problem appears to be difficult.

6.4. Show that every positive integer has a multiple which has decimal form $99\ldots900\ldots0$.

Solution

Consider the positive integer n. If we divide each term in the sequence

$$9, \quad 99, \quad 999, \quad 9999, \ldots$$

by n, we obtain a sequence of remainders, say $r_1, r_2, r_3, \ldots$, where $0 \leqslant r_i < n$ for each i. It follows that two of these remainders must be the same, say $r_i = r_j$ for some i, j, $i < j$. Then we have

$$99 \ldots 9 = k_i n + r_i,$$

$$99 \ldots \ldots \ldots 9 = k_j n + r_j,$$

so that

$$99 \ldots 900 \ldots 0 = (k_j - k_i)\, n,$$

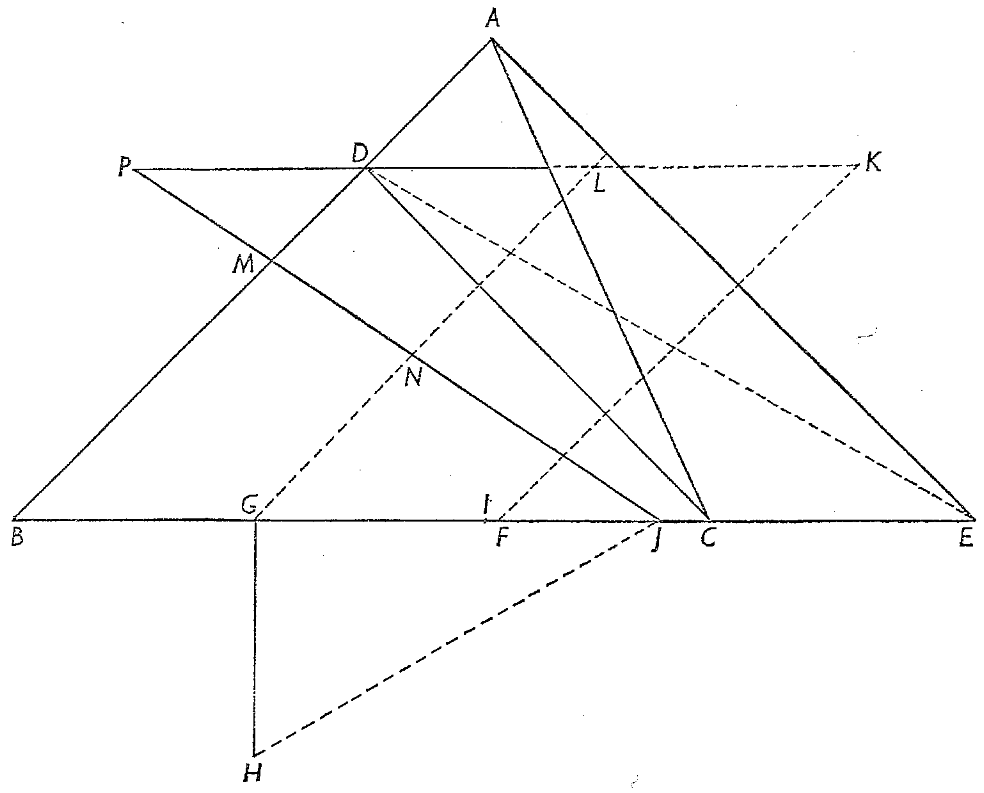
which gives the result.

Also solved by David Seal, Michael Beasley.

Solution to a problem submitted by Major W. H. Carter

Major Carter submitted the following problem published in Volume 6 Number 1: given any triangle and any point, construct a straight line through the point which bisects the area of the triangle. Three solutions were supplied by A. R. Pargeter of Blundell's School, Tiverton, Devon. One involved a discussion of the tangents to a certain hyperbola, another used homographies. We publish the third solution below.

Let ABC be the given triangle and P the given point, shown outside the triangle. Draw PD parallel to BC to meet AB at D. Join DC. Draw AE parallel to DC to meet BC (produced) at E. Bisect BE at F. Draw the perpendicular bisector GH of BF (G being the midpoint of BF). Cut off GH and GI (GI lying along GC) equal to PD. With centre H and radius BI, draw an arc to cut BC at J. Join PJ, which is the required line.



Proof.

$$\text{Area}\,(DBFK) = \text{Area}\,(\Delta DBE) = \text{Area}\,(\Delta ABC).$$

Hence

$$\text{Area}\,(DBGL) = \tfrac{1}{2}\,\text{Area}\,(DBFK) = \tfrac{1}{2}\,\text{Area}\,(\Delta ABC).$$

Now

$$\Delta PLN : \Delta PDM : \Delta JGN = PL^2 : PD^2 : JG^2$$

and

$$JG^2 = JH^2 - GH^2 = BI^2 - PD^2 = PL^2 - PD^2,$$

so that

$$\begin{aligned}\text{Area}\,(\Delta JGN) &= \text{Area}\,(\Delta PLN) - \text{Area}\,(\Delta PDM)\\ &= \text{Area}\,(LDMN).\end{aligned}$$

If we add the quadrilateral *MBGN*, we obtain

$$\text{Area}\,(\Delta JBM) = \text{Area}\,(LDBG) = \tfrac{1}{2}\,\text{Area}\,(\Delta ABC),$$

as required.

Book Reviews

Arithmetical Excursions. By H. Bowers and J. E. Bowers. Constable and Company Limited, London, 1973. Pp. xi+320. £1·50 (paperback).

This book was first published in America in 1961 and was intended for use with intelligent high school pupils from the age of 11 onwards. It has 27 chapters, some only 4 pages long; although much of the work will be found familiar it contains many facts about numbers that the modern pupil will not know. There are many items of interest that do not form part of an examination syllabus, such as calculating prodigies; the rule for division by 11; the Russian peasant's method of multiplication; Fermat numbers; Goldbach's conjecture; figurate, perfect and amicable numbers; and so on.

Most chapters have a set of exercises for which there are answers, but a bright pupil will find them rather trivial. I do not envisage this book being used for any O-level course, and it is too elementary for post O-level, unless it is used for a mathematics course in a sixth form general studies scheme. I would recommend anyone, mainly in the age range 11–16, to read it if they are interested in numbers. Provided the large amount of familiar work between the more unusual items does not bore readers, this book should give much enjoyment.

St Paul's School, London S.W.13 — W. G. Nunn

New Mathematics III. By K. S. Snell and J. B. Morgan. Cambridge University Press, 1972. Pp. vi+266. £1·50.

New Mathematics IV. By K. S. Snell and J. B. Morgan. Cambridge University Press, 1973. Pp. viii+280. £1·60.

These two volumes are second editions of texts first published in 1961. The first three volumes of the set are intended to cover the essential ground for O-level syllabus. As one would expect from the authors, the texts are clear and thorough, well set out and suited to their object.

At the present time, mathematical teaching in this country is in something of a turmoil. No longer is it possible to know that because a pupil has reached O-level he has met and acquired, with very small variations, a well-defined body of mathematical knowledge. Teachers are uncertain whether they ought to be teaching matrix transformations or solutions of quadratic equations. Universities are conflicting in their requirements and contradictory on the value they place on some topics.

In the face of all this, these volumes try to steer a middle course. Here are sets and matrices; here also are proofs of geometrical theorems. Inevitably no topic is dealt with in great depth, and obviously no pupil will be expected to cover all the ground. The authors have provided a text which, used with wisdom, should be of value to pupils of varied ability, allowing the brightest to take a broad view of the subject and others to concentrate on their own syllabus.

Brentwood School, Essex O. B. TYMMS

Calculus. By J. HUNTER. Blackie, Glasgow, with W. & R. Chambers Ltd, Edinburgh, 1972. Pp. v+245. £2·00.

Aimed at sixth form pupils in schools and first year students at universities and colleges, this tightly packed book covers in detail all of the basic theory of differentiation and integration together with many mathematical applications. Except for a short section on partial differentiation, it is only concerned with functions of one real variable. The applications given include curve sketching, length, area, volume, circle of curvature, Taylor series, Simpson's rule and differential equations of the first and second orders.

It is a sensible book which develops the subject logically wherever this is practicable but points out and states clearly those results which go too deep for an adequate proof at this level. This enables the author to keep in close touch with intuition and to avoid unnecessary abstractions without encouraging slipshod habits of thought. In addition to many worked examples in the text, each chapter closes with a generous supply of problems having answers in the back of the book. These provide the practical experience which is so important for an understanding of the rules of differentiation and integration. The layout of the book is attractive with well printed diagrams. In general, the definitions of technical terms are clear and easily found. But the explanation of some recently coined terms such as injection, composition, etc. seems rather brief and may cause unnecessary indigestion. However this is only a trivial blemish. On the whole it is a clear, comprehensive and carefully written book which should be particularly useful to sixth form pupils hoping to go further.

University of Durham R. A. SMITH

Sets and Numbers. By S. ŚWIERCZKOWSKI. Routledge and Kegan Paul, London, 1972. Pp. iii+74. £0·50.

The book is intended to introduce the fundamental concepts of set theory. Although it does deal with the basic principles of the subject, some familiarity with these ideas would be useful, as the first two chapters deal rather quickly with all the basic concepts. Thus although the elementary operations can be understood, the applications to the later sections of the book might prove difficult.

On the whole the book is very readable: most points are well explained in their context, and not by too many references to preceding pages.

Some examples are given within each chapter to elucidate the points made; however, more of these would be helpful. The problems at the end of each chapter are sometimes used to introduce new material, and in cases where the new concepts are easily grasped this is a good idea. There are answers given at the end of the book. This is very helpful, as in many books there is no way of knowing whether or not an answer is correct. Although some of the problems are quite difficult, they are well explained at the end of the book. A few more examples might be useful in the later exercises, to illustrate the more difficult work.

The book would be of most use to first year undergraduates since, although the material is possibly familiar, it is often the new approach which worries the student. This book overcomes this problem by gradually introducing the new way of thinking.

St Aidan's College, Durham G. CHILVER

Advanced Level Mathematics (Pure and Applied). By C. J. TRANTER and C. G. LAMBE. English Universities Press Ltd, London, 1973. (Third Edition.) Pp. 582. £2·25.

This is the third edition of the combined red, white and blue book: *Advanced Level Applied Mathematics* (Lambe) and *Advanced Level Pure Mathematics* (Tranter) first appeared in 1953, the year of Queen Elizabeth's coronation. The fashion for colours like red, white and blue has changed slightly since then and so has the fashion for the sort of mathematics contained in these books. They exhibit excellent qualities in their careful selection of material and clarity of exposition, but they do not reflect the great changes in mathematical education of the last 20 years.

The series of red, white and blue books claims to 'ensure the passing of examinations'. I cannot confirm that the present volume will guarantee certain success; but the reader who is taking one of the 'traditional' A-level course syllabuses will be able to prepare himself well, especially if he manages the majority of its 2000 examples and exercises.

A certain modern flavour has been injected into the third edition through the inclusion of a chapter on vectors and one on matrices; for publishing reasons, the latter appears as Chapter 22 rather than in its 'natural' position as Chapter 6. For me, however, this misses the point in introducing matrices: while I need to learn some matrix methods, I should also like to understand the algebraic structure of matrices and the many applications that they have. Unfortunately the book succeeds in presenting only some matrix methods; but this it achieves well and some writers of 'modern' texts could learn much from this presentation.

Sub-headings on Efficiency, Power and Banking led me to wonder whether this book might say something of the world of business and the national economy, but applications outside mechanics are few with the exception of some linear differential equations. This is a traditional book; while I cannot recommend it for inspiration in modern mathematics, many readers will find that it possesses admirable qualities as a reference work. I shall certainly keep my volume at hand in case I forget what 'efficiency' is.

Schools Council Project G. R. H. BOYS
Critical Review of the Mathematics Project
University of Nottingham

Practical Programming. By P. N. CORLETT and J. D. TINSLEY. Cambridge University Press, 1972. (Second Edition.) Pp. x+264. £3·90, hardback; £1·50, paperback.

This is the second edition of the book first published in 1968, when it was something of a landmark. Since then several more books for schools have been produced, and conversational languages such as Basic and Telcomp have proved popular in class because of their simplicity. Nevertheless this remains a useful volume both for teachers and committed mathematical sixth formers. Although this is an Algol book, most examples also have flow diagrams with which to analyse the problem.

After chapters on flow diagrams and Algol, there is a set of straightforward worked examples giving the student something concrete to tackle. Arrays are introduced next to enable him to attack a wider group of numerical questions. The statistics includes methods for ranking numbers; these open up a fascinating field of problems.

The theory of Algol procedures leads into iterative techniques, polynomials, and matrices as dealt with in modern A-level courses. Two new chapters by R. A. Court have been added on Fortran, its details and its uses in data processing, including alphanumeric manipulation.

As stated in the book, this is for those whose mathematical education is beyond elementary level, and the text includes most of the numerical methods in the A-level course with the exception of the calculus aspects; there is much else besides. Teachers will find plenty of opportunity for discussion and many chances to set open-ended questions. This is another nicely produced S.M.P. handbook which will be of benefit in the library and on the teacher's desk.

Oundle School M. A. BLOXHAM

Exploring Mathematics on Your Own. By WILLIAM H. GLENN and DONOVAN A. JOHNSON. Dover Publications, Inc, New York, 1972. Pp. 303. £1·25.

This is not a new book but a republication of one first published in 1960. Because it has been designed for the layman with an interest in mathematics, it is very elementary but also interesting. Although the average student will find much of the material very familiar, he will also find much which is fascinating and enjoyable.

The book is divided into six parts, each self contained. The first part looks at numeration systems and shows how to compute in bases other than the usual base ten. Part Two investigates number patterns leading on to infinite series and their summations. Part Three discusses Pythagoras' Theorem; it begins with a mention of his life and teachings and proceeds to demonstrate the usefulness of his theorem today. The elementary ideas of set theory are introduced in Part Four and Part Five takes a look at the intriguing topic of topology. The book concludes with a section of mathematical 'tricks'; the more interesting of these include calendar problems and the trisection of an angle using a watch.

Each section contains many problems, all with full solutions. Active participation by the reader in the practical aspects of the subject is encouraged. Instructions are included for making an abacus to assist with counting in different bases, and a slide rule for determining the hypotenuse of a right-angled triangle with two given sides. Few will be able to resist making a Moebius Strip, that strange one-sided piece of paper, and demonstrating its unexpected properties. At the conclusion of each section are suggestions for further reading; it is regrettable that these have not been updated as all are references to pre-1960 publications.

This is a book which can provide much enjoyment. Easily understood, it should make good reading for the younger student.

12 Barnfield Avenue, Sheffield S10 5TA C. M. NIXON

Recreational Problems in Geometric Dissections and How to Solve Them. By HARRY LINDGREN, revised and enlarged by GREG FREDERICKSON. Constable and Company Limited, London, 1973. Pp. viii+184. £1·00 (in UK).

Many of us are apt to think of dissections as isolated puzzles, ingenious and perhaps amusing: but here is a systematic account of methods for finding them which gives new interest to the subject. If a figure can be arranged so that a continuous repeating pattern can be formed from it, either as a frieze or as a tessellation, then it can be easily dissected into another such figure. For the frieze type, the method is to cross the two strips and use the common area. At this point (if not before) the reader begins to feel the urge to do it himself; it becomes a true recreation, that is to say, something to do, with freedom to choose the material and to do it one's own way. The requirements are geometrical instruments, plenty of tracing paper and lots of time—just the thing for anyone making a long stay in a hospital or prison! Hundreds of dissections are described and illustrated and there is plenty of scope for inventions and improvements. The real addict wants to do a dissection with the minimum number of pieces and, as it is rarely possible to prove that the minimum has been reached, it is always intriguing to try to reduce the number.

Tessellations provide some pleasing examples and we go on to 'rational' dissections (so-called, apparently, for the not very adequate reason that the word begins with R, the letters P, Q, S and T having already been reserved for other types). We then proceed to special treatment of the n-gon ($n = 7, 8, 9, 10$), of rectilinear letters, of stars, of many into one, and finally to some curvilinear dissections and some of solid figures. There are several useful appendices: problems, with solutions, dimensions of regular polygons, an ingeniously arranged index of dissections, a list of sources and a number of clever dissections invented by the reviser.

The book is well produced and good value at the price.

18 West Hill, Charminster, Dorchester, Dorset E. H. LOCKWOOD

Test Your Logic. By GEORGE J. SUMMERS. Constable and Company Limited, London, 1973. Pp. vi + 100. £0·75 (paperback).

The puzzles in *Test Your Logic* do, as the author claims, resemble short 'whodunit' mysteries. The 'clues' are explicitly stated, as is the final requirement, and each problem is a matter of deciding how to induce the best interaction between the clues. Except for the layout the puzzles resemble in type the *Sunday Times* brain teasers.

Forty of the fifty problems do not need any knowledge of algebra, and the knowledge required for the other ten is minimal. Each of the problems is original, and composing fifty dissimilar logic problems which are interesting is certainly an achievement.

The level of difficulty of the problems in the book increases as we read on. The first problems are highly suitable as an introduction to the working of the logic problem, and by the end, the standard is such as to challenge an expert. The problems are interesting enough to maintain a novice's efforts even when he is faltering over the first problem or two; here the idea of providing hints at the bottom of the page (uʍop ǝpᴉsdn) which partially (though only partially) open the doors to the solution of the problem is very helpful.

On the debit side, one or two of the problems are decidedly ambiguous, though on reading the solution one realises that the ambiguity stems, not from a vague wording, but an unstated assumption in the problem. A more serious defect is that one or two, but no more, of the problems are hard slog, trial-and-error types. You can easily spend an hour, or maybe more, meticulously checking possibilities and still be using the best method available.

On the whole, *Test Your Logic* is a well-written, original book, interesting to the specialist and non-specialist alike. It is a valuable addition to any mathematical book collection.

St Paul's School, London S.W.13 P. W. HEWITT

An Introduction to Mathematical Reasoning. By B. IGLEWICZ and J. STOYLE. Collier–Macmillan Publishers, London, 1973. Pp. vii + 231. £2·50.

This book is what the title says. Various areas of arithmetic, algebra and geometry are studied, with particular attention to the methods of proof involved. There are chapters on symbolic logic, proof by induction, direct proof, indirect proof and disproof by counter-example. Almost all of the examples should be accessible to a capable sixth form student, and many of them might be covered in an A-level course. There are many not too difficult exercises, and the book would be suitable for independent reading at school or university; although intending buyers should note that the price is determined by current American rather than British levels.

Here is a very suitable first introduction, but potential readers who require rather more may be disappointed. Chapter 2 introduces symbolic logic including truth tables and the standard notations as far as implication; but this notation is referred to only briefly in the subsequent chapters, where the methods of proof are very much those that would be found in traditional English school texts making no mention of symbolic logic at all. It would be fascinating to find a book at this level which showed symbolic logic, in some detail, as a tool of value in giving deeper understanding of traditional methods of proof, so preparing the student for a more thorough-going approach to formalised mathematics later on. Such a book has still to be written, but these authors point the way.

H.M.I., Department of Education and Science T. J. FLETCHER

Notes on Contributors

Underwood Dudley is an Associate Professor of Mathematics at DePauw University, Greencastle, Indiana. He is a graduate of Carnegie Institute of Technology and the University of Michigan, and taught for a time at Ohio State University. He is the author of a book *Elementary Number Theory* as well as papers on number theory, the history of mathematics and recreational mathematics.

Pat Rogers, who also contributed to the last volume of *Spectrum*, has since then returned from York University, Toronto, and is now a Lecturer at Goldsmith's College in the University of London. After graduating from Oxford she spent a year at Toronto University and subsequently she taught at Ealing Technical College and at the Polytechnic of North London. Her main interest is in mathematical logic.

Joan M. Holland has taught mathematics in schools in St Andrews and Birmingham. She has also held administrative posts with education authorities in Hertfordshire and in Singapore and Kenya. For the final five years before retiring in 1969 she was Senior Lecturer in Mathematics at Bishop Otter College, Chichester. Since retiring she has had a book, *Studies in Structure*, published. Her retirement hobbies are bridge and gardening.

Dieter K. Ross is Senior Lecturer in Applied Mathematics and Acting Head of the Department at La Trobe University, Victoria, Australia. After graduating from Melbourne University, he was appointed to a lectureship at Manchester University. He is very interested in mathematical education in which capacity he is an Academic Adviser to the Victoria Institute of Colleges. He is also an Assistant Editor of the *Journal of the Australian Mathematical Society.*

Edwin F. Carter is a Lecturer in Physical Chemistry at the Swinburne College of Technology, Victoria, Australia. After completing a science degree in chemistry he has commenced studies in mathematics leading to a Master of Science degree at La Trobe University. He is particularly interested in mathematics and its applications to the physical sciences.

Anthony Sackfield is a Lecturer in Mathematics at Trent Polytechnic. His research interest lies in the general theory of relativity.

Mike Frost is a Lecturer in Mathematics at Trent Polytechnic, with research interest in mathematical methods.

INDEX

Volume 4 (1971/72) Volume 5 (1972/73) Volume 6 (1973/74)

Contents

PRICES (*postage included*)

Prices for Volume 6 (Issues Nos. 1 and 2):

Subscribers in Britain and Europe: £0·50
Subscribers overseas: £1·00 (US\$2.50; \$A.2.00)
(These prices apply even if the order is placed by an agent in Britain.)
A discount of 10% is allowed on all orders for five or more copies.

Back issues:

Volume 1 is out of print. Volumes 2, 3, 4 and 5 are available at the back issue price per volume of £1·00 (US\$2.50; \$A.2.00).

Enquiries about rates, subscriptions, and advertisements should be directed to:

Editor—*Mathematical Spectrum*,
Hicks Building,
The University,
Sheffield S3 7RH, England.

Printed in England by John Wright & Sons Ltd. Bristol