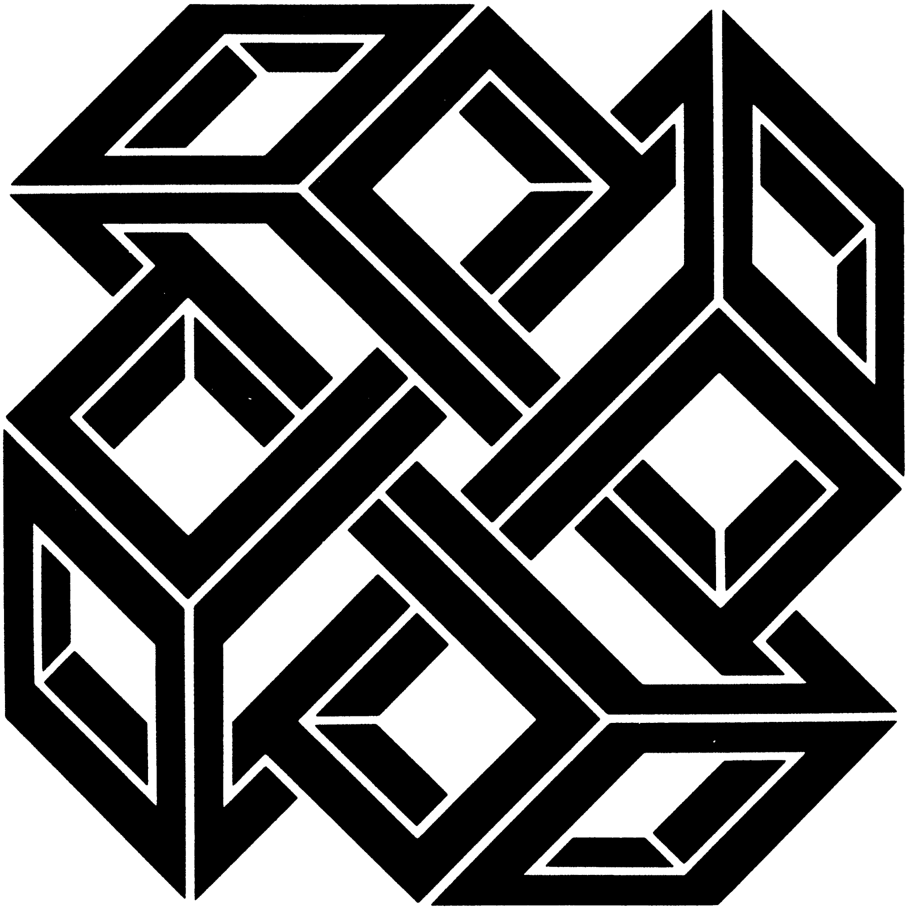
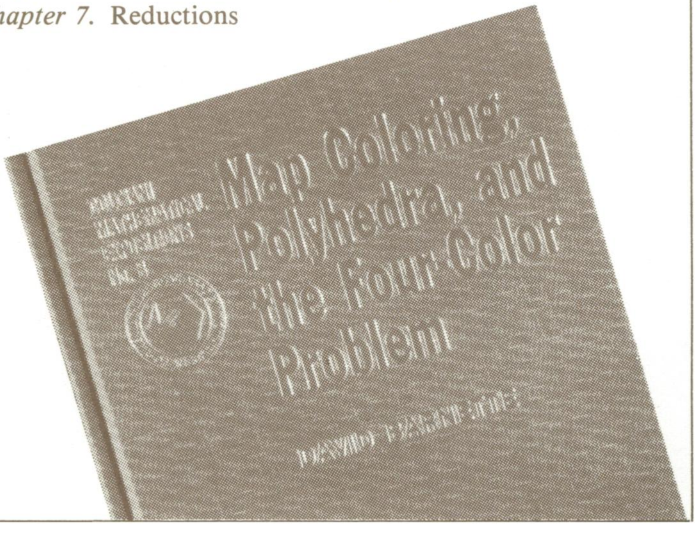
MATHEMATICS MAGAZINE



Vol. 58 No. 4
September 1985

KNOW/DON'T KNOW • DOUBLE ENTRY BOOKKEEPING
PRIME PRODUCERS & P.I.D.'S • SLICING RUBIK'S CUBE


Map Coloring, Polyhedra, and the Four Color Problem
David Barnette

MATHEMATICS MAGAZINE

Volume 58, Number 4, September 1985
(ISSN— 0025- 570X)

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS

COVER: Impossible cubes. Design by Tamás F. Farkas, Budapest, Hungary.

EDITORIAL POLICY

Mathematics Magazine is a journal which aims to provide inviting, informal mathematical exposition. Manuscripts for the *Magazine* should be written in a clear and lively expository style and stocked with appropriate examples and graphics. Our advice to authors is: say something new in an appealing way or say something old in a refreshing way. The *Magazine* is not a research journal and so the style, quality, and level of articles should realistically permit their use to supplement undergraduate courses. The editor invites manuscripts that provide insight into the history and application of mathematics, that point out interrelationships between several branches of mathematics and that illustrate the fun of doing mathematics.

The full statement of editorial policy appears in this *Magazine,* Vol. 54, pp. 44-45, and is available from the Editor. Manuscripts to be submitted should not be concurrently submitted to, accepted for publication by, nor published by another journal or publisher.

Send new manuscripts to: G. L. Alexanderson, Editor-elect, Mathematics Magazine, University of Santa Clara, Santa Clara, CA 95053. Manuscripts should be typewritten and double spaced and prepared in a style consistent with the format of *Mathematics Magazine.* Authors should submit the original and one copy and keep one copy. Illustrations should be carefully prepared on separate sheets in black ink, the original without lettering and two copies with lettering added.

AUTHORS

John O. Kiltinen ("Goldbach, Lemoine, and a Know/Don't Know Problem") received his Ph.D. from Duke University. He has served on the exam committee for the MAA Michigan Section's high school competition and edits the section newsletter. His interest in know/don't know problems was piqued by Martin Gardner's column in *Scientific American.* After some work on the problem, he organized a research participation class at Northern Michigan University to pursue the subject with four students.

Co-author **Peter B. Young** is a research scientist whose work involves the design and implementation of a signal and image processing system to be used in the analysis of infrared satellite imagery. After completing his B.S. degree, he participated in this research project, validating the refined Lemoine conjecture to 53,000 and identifying the level two and three numbers shown in Tables 1 and 2.

Daniel Fendel ("Prime-producing Polynomials and Principal Ideal Domains") received his Ph.D. from Yale University. His main interest is mathematics education and teacher training, for which he recently completed a textbook, *Understanding the Structure of Elementary School Mathematics* (Allyn and Bacon, fall 1986). The ideas in this article were developed while teaching a graduate abstract algebra course, which included discussion of different types of integral domains.

ILLUSTRATIONS

Thomas Cappuccio captures Sam and Prudence, p. 198.

Dave Logothetti caricatures Borg and McEnroe, p. 224.

All other illustrations were provided by the authors.

The MATHEMATICS MAGAZINE (ISSN 0025-570X) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, D.C. 20036 and Montpelier, VT, five times a year: January, March, May, September, and November.

The annual subscription price for the MATHEMATICS MAGAZINE to an individual member of the Association is $11 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $22. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The non-member/library subscription price is $28 per year. Bulk subscriptions (5 or more copies) are available to colleges and universities for classroom distribution to undergraduate students at a 41% discount ($6.50 per copy—minimum order $32.50).

Subscription correspondence and notice of change of address should be sent to the Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036. Back issues may be purchased, when in print, from P. and H. Bliss Company, Middletown, CT 06457. Microfilmed issues may be obtained from University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

Advertising correspondence should be addressed to Ms. Elaine Pedreira, Advertising Manager, The Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036.

Second class postage paid at Washington, D.C. and additional mailing offices.

Postmaster: Send address changes to Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036.

PRINTED IN THE UNITED STATES OF AMERICA

Goldbach, Lemoine, and a Know/Don't Know Problem

Sum knowledge is productive,
product knowledge is better, but
Goldbach and Lemoine determine the limits.

JOHN O. KILTINEN
Northern Michigan University
Marquette, MI 49855

PETER B. YOUNG
Infrared and Optics Division
Environmental Research Institute of Michigan
Ann Arbor, MI 48107

Know/don't know problems are a popular form of mathematical puzzle. They give an account of two or more persons, each of whom has partial information about a situation into which they have been placed. We are then told something of what they know or do not know, and what effect some sharing of the information by one has on what another knows. We are then asked to decipher all this and deduce some particular piece of information.

As a class, they are problems in logic, and can be successfully analyzed as such (e.g., [**1**]). However, when problems of this type involve the natural numbers, they can be of interest because of the number theoretic issues which they raise, for example, the recent problem in this MAGAZINE [**6**] which involves the sums of squares. Our objective in this note is to study a particular sort of number theoretic know/don't know problem, and explore its relationships with the famous conjecture of Goldbach and with a refinement of a lesser known conjecture of Lemoine.

The situation we explore involves choosing two integers, r and s, both of which are at least 2. Their sum $m = r + s$ is given to one mathematician, Sam, while their product $n = r \cdot s$ is given to another, Prudence.

Suppose that after some thought, they exchange the following information*:

P: I don't know your sum, Sam.
S: I knew you didn't, Prudence. (1)

What does this tell us? Prudence's first statement tells us that r and s are not both primes, for if $n = r \cdot s$ for primes r and s, there is but one possible value for m. Sam, of course, realizes this too, so he has checked his number to see if it is a sum of two primes. From this statement, we, as well as Prudence, can infer that it is not. Enter C. Goldbach, whose famous 1742 conjecture states the following:

GOLDBACH'S CONJECTURE [GC]. *Every even integer which is at least 4 is the sum of two primes.*

*It will be assumed throughout that the players always reveal something new about the state of their knowledge.

If we grant the truth of [GC]**, we may conclude that Sam's sum m is not even. Now suppose that the conversation concludes like this:

$$\begin{array}{ll} \text{P:} & \text{Now I know your sum.} \\ \text{S:} & \text{And now I know your product.} \end{array} \tag{2}$$

What more can we deduce? Well, since Prudence now knows that the sum cannot be even, she knows, as we do, that r and s must be of opposite parity. Furthermore, she knows that $r+s$ does not equal $p+2$, where p is a prime, for if it did, Sam would not have been so sure of himself in (1). Thus, since she now knows his sum, we may conclude that only one of the several ways to factor her product n involves terms of opposite parity such that their sum is not equal to $p+2$ for some prime p.

Sam has reasoned similarly, and arrived at the same conclusion about the factorizations of n. He reviews all products of pairs of integers which add up to m, and apparently finds that only one of these has factorizations fitting the pattern.

Do we now have sufficient information to deduce values for m and n? Yes, but not uniquely. The smallest solution is $m = 17 = 4 + 13$ and $n = 52 = 4 \cdot 13$. Clearly 17 is not the sum of two primes and 52 has only one factorization into terms of opposite parity. Furthermore, all of the other products of summands of 17, namely, 30, 42, 60, 66, 70 and 72 have other factorizations into terms of opposite parity with sums not two more than a prime. For example, 60 factors as $20 \cdot 3$, and $20 + 3 - 2 = 21$ is not prime. The reader may also check that the pairs $127 = 16 + 111$ with 1776, $757 = 556 + 201$ with 111,756, and $997 = 576 + 421$ with 242,496 are three other possible solutions.

A problem involving the dialog in (1) and (2) appeared in this MAGAZINE [**11**], submitted by D. Sprows. Martin Gardner [**7**] also discussed the problem and conjectured in private correspondence that the number of solutions is infinite, a conjecture which seems plausible to us.

Our goal is to look at variations on this problem involving different dialogs. For example, are there pairs (m, n) for which the exchange in (1) would take place several times before Sam or Prudence knew the other's number? Could the exchange in (1) go on *ad infinitum* for suitable m and n? What about solutions compatible with the following exchange?

$$\begin{array}{ll} \text{P:} & \text{I don't know your sum.} \\ \text{S:} & \text{Thanks for that information.} \\ \text{P:} & \text{I still don't know your sum.} \\ \text{S:} & \text{Now I know your product.} \\ \text{P:} & \text{And now I know your sum.} \end{array} \tag{3}$$

In order to handle questions such as these, we need to develop some terminology. We begin by naming the relation between a sum and product which occurs in these problems. We will say that two natural numbers m and n are **PF-related** and write $\boldsymbol{m}\,\mathbf{PF}\boldsymbol{n}$ if there exists a pair of numbers r and s, both at least 2, such that $m = r + s$ and $n = rs$. The notation "PF" is intended to convey that a two-term *partition* of m yields a *factorization* of n. If $m\,\mathrm{PF}n$, we will also call m a **P-partner** of n and n an **F-partner** of m. FIGURE 1 illustrates the relation PF for small m and n.

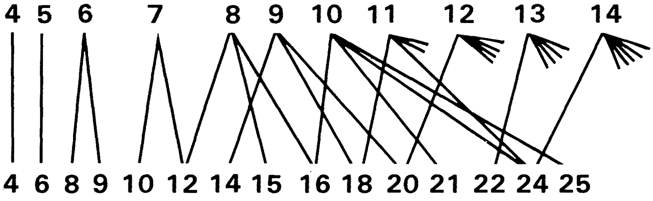
The numbers given to Sam and Prudence are of course PF-related, which means that $m \geqq 4$. For all such m, we will define a concept of **levels of identifiability** which corresponds to the minimal number of exchanges of (1) between Prudence and Sam which must occur before Prudence could possibly know m. To begin, we will say that m is **l-identifiable**, or m is **at level one of identifiability**, and write $L(m) = 1$ provided that m has an F-partner n for which it is the only P-partner. In this case, we will call n a **1-identifier** of m.

We continue, extending this concept recursively. Suppose that the concepts of j-identifiability

**Although [GC] is not a theorem, its truth has been verified by direct computation for all even integers up to at least 10^8, according to the best information available to the authors.

P-partner m:	4	5	6	6	7	7	8	8	8	
r,s: r+s=m r•s=n	2,2	2,3	2,4	3,3	2,5	3,4	2,6	3,5	4,4	• • •
F-partner n:	4	6	8	9	10	12	12	15	16	

(a)



(b)

FIGURE 1. **The relation PF shown by means of a table (a) and a bipartite graph (b). P-partners are shown above and F-partners are shown below.**

and $L(k)=j$ have been defined for $j=1,\ldots,i-1$. Then for $m\geqq 4$, we will define m to be **i-identifiable** ($L(m)=i$, m is **at level i of identifiability**) provided that m is not j-identifiable for any $j<i$ and there exists an integer n such that $m\,\mathrm{PF}n$ and such that all of the P-partners of n other than m have a level of identifiability less than i. That is, $L(m)=i$ if m is at no lower level and if m has an F-partner n such that if $k\mathrm{PF}n$, then $k=m$ or $L(k)<i$. Under these circumstances, we will call n an **i-identifier** for m. See FIGURE 2 for a visual representation of these relationships.

EXAMPLE 1. Let us illustrate these definitions for $m=9$ and 11. First, $L(9)=1$ and 14 is a 1-identifier. This is so since 9 PF 14 ($9=7+2$ and $14=7\cdot 2$), and since 14 has only one proper factorization, k PF 14 if and only if $k=9$. Second, $L(11)=2$, and 18, 24 and 28 are all 2-identifiers for 11. To see this, one must first note that $L(11)\neq 1$. Indeed, the F-partners of 11 are 18, 24, 28 and 30, and each of these has several proper factorizations and therefore several P-partners. Thus, none of these F-partners is a 1-identifier for 11. However, 18 is a 2-identifier since its only P-partners are 9 and 11, and as we saw, $L(9)=1$. One can also show that 24 and 28 are 2-identifiers for 11, but that 30 is not. This last claim is so because 17 PF 30, and $L(17)\neq 1$.

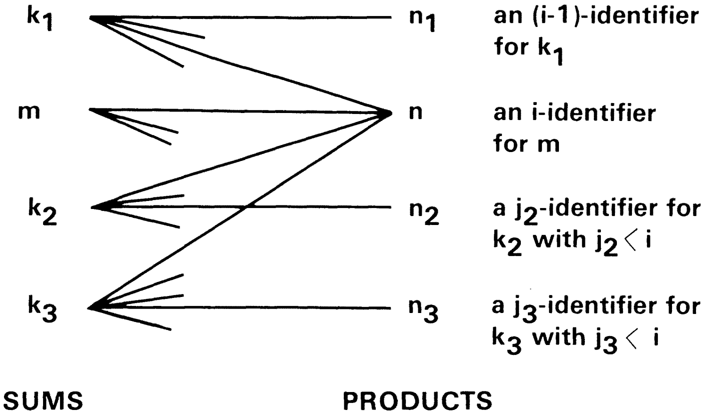

k_1
n_1
an (i-1)-identifier for k_1
m
n
an i-identifier for m
k_2
n_2
a j_2-identifier for k_2 with $j_2 < i$
k_3
n_3
a j_3-identifier for k_3 with $j_3 < i$
SUMS
PRODUCTS

FIGURE 2. **For n to be an i-identifier for one of its P-partners m, the other P-partners of n must all have identifiers at levels less than i, with at least one at level $i-1$.**

This terminology provides a basis for a sharper discussion of our class of know/don't know problems. The reader can verify that the original dialog in (1) and (2) translates into the following succinct form:

P: n is not a 1-identifier for m.
S: m is not 1-identifiable. (1′)

P: n is a 2-identifier for m.
S: m has a unique 2-identifier. (2′)

To illustrate, let us consider Prudence's statement in (2). She knows that Sam's number is not 1-identifiable from his statement in (1). Thus, she can eliminate all 1-identifiable P-partners of her n from further consideration. Since she now knows m, it must be the unique P-partner of n which is not at level one, thus placing it at level two, with n a 2-identifier. Note too that we can now say concisely that to find all solutions for the original problem (1)–(2) means to find all pairs (m,n) such that n is the unique 2-identifier for m.

What about repetitions of (1)? We see that the jth instance of this exchange of "I don't know" and "I knew you didn't" would translate into:

P: n is not a j-identifier for m.
S: m is not j-identifiable. (1_j)

If the dialog involved $i-1$ repetitions of (1) followed by (2), then the final exchange becomes:

P: n is an i-identifier for m.
S: m has a unique i-identifier. (2_i)

We also note that we can now rephrase and answer our question about the possibility of an infinite repetition of dialog (1). This would be a sequence $(1_1),(1_2),(1_3),\ldots,$ which clearly means that m is an integer for which no level of identifiability is defined. The following theorem shows that this is impossible.

THEOREM 1. *For every integer m with $m \geqq 4$, there is an integer i such that $L(m)=i$.*

Proof. Assuming the existence of integers k for which $L(k)$ is not defined, let us take m to be the least such integer. Then $L(k)$ is defined for all $k<m$, so we may define i to be $1+\max\{L(k)|k<m\}$. We will obtain a contradiction to our assumption that m has no level of identifiability by showing that $n=2(m-2)$ is a j-identifier for m for some $j \leqq i$. This follows from the fact, which we will prove in a moment, that m is the largest P-partner of n. Thus, if $k\,\mathrm{PF}\,n$ and $k \neq m$, then $k<m$, so $L(k)<i$. Then setting $j=1+\max\{L(k)|k\,\mathrm{PF}\,n$ and $k\neq m\}$, we have $j \leqslant i$ and $L(m)=j$.

To prove the claim that m is the largest P-partner of $n = 2(m-2)$, take any k such that $k\,\mathrm{PF}\,n$ and $k \neq m$. Then one has integers x and y, both being at least 2, such that $x + y = k$ and $xy = n = 2(m-2)$. Since n is even, we may assume that x is even, say, $x = 2z$. Then $2zy = n = 2(m-2)$, so $m = zy + 2$. If $z = 1$ or $y = 2$, we would have $k = m$. Thus, $z > 1$ and $y > 2$. From this, it follows that $2(z-1) < y(z-1)$, so $k = 2z + y < zy + 2 = m$. This completes the proof.

Now that we know that $L(m)$ is defined for every integer $m \geqq 4$, can we readily evaluate this function? The following two theorems address the matter of recognizing integers m such that $L(m)$ is either 1 or 2.

THEOREM 2. *Let m be an integer with $m \geqq 4$. Then $L(m) = 1$ if and only if m is the sum of two primes, or $m = p + p^2$ where p is prime.*

Proof. Suppose that $m = p + q$ with p and q prime. Then clearly $n = pq$ is a 1-identifier for m, so $L(m) = 1$. Similarly, if $m = p + p^2$, then $n = p^3$ is a 1-identifier for m since the only proper factorization of p^3 is $p \cdot p^2$.

Now suppose that m is 1-identifiable. Let n be a 1-identifier for m and let x and y be the integers such that $x + y = m$ and $xy = n$. If x and y are both prime, we are done, so assume that y is not prime. Let $y = pz$, where p is prime and $z \geqslant 2$. Then one has $k_1\,\mathrm{PF}\,n$ where $k_1 = xz + p$, so since m is the unique P-partner of n, $xz + p = k_1 = m = x + pz$. Thus, $x(z-1) = p(z-1)$, so $x = p$. Furthermore, $k_2\,\mathrm{PF}\,n$ where $k_2 = px + z = p^2 + z$, so $k_2 = m$. This means that $p^2 + z = p + pz$, whence, $p(p-1) = z(p-1)$ and $z = p$. Thus, $m = p + p^2$.

Goldbach's Conjecture permits this result to take on the following more descriptive form. (In stating this corollary and other results which follow, we will indicate the conjectures upon which they depend with their abbreviations.)

COROLLARY 2.1 [GC]. *An integer m with $m \geqq 4$ is 1-identifiable if and only if m is even or $m = p + 2$ where p is an odd prime.*

Note that since $m = p + p^2$ is always even, [GC] renders redundant the 1-identifiers of the form p^3 for such m's. However it is interesting to note how our concept of levels relates to the current best proven approximation to Goldbach's Conjecture. Chen [2] has proven that every sufficiently large even number is the sum of two primes or a prime and a product of two primes. Numbers of the form $p + p^2$ are very special cases of his weaker form. By Theorem 2, the statement "Every even integer $m \geqq 4$ is at level one" is an assertion which lies between [GC] and Chen's Theorem, but clearly closer to the former.

Having explicitly described level one numbers, let us turn our attention to level two. A useful characterization is again obtainable, although one not as simple. If $L(m) \neq 1$, then by Corollary 2.1, we know that m is odd, so summands of m must be of opposite parity. Thus, without loss of generality, we may establish in what follows the notation for partitioning an odd m as $m = 2^i x + y$ with x and y odd, $y \geqq 3$ and $i \geqq 1$.

THEOREM 3 [GC]. *Let $m = 2^i x + y$ be an odd integer such that $L(m) \neq 1$. Then $n = 2^i xy$ is a 2-identifier for m (and thus $L(m) = 2$) if and only if:*

For any divisor d of xy other than x or xy, the number $2^i d + xy/d - 2$ is prime. (4)

Proof. The condition (4) is clearly necessary if $n = 2^i xy$ is to be a 2-identifier for m. Indeed, suppose the latter, and take a divisor d of xy other than x or xy. Let $k = 2^i d + xy/d$. Since $d \neq xy$, both of the factors $2^i d$ and xy/d of n are at least 2, so one has $k\,\mathrm{PF}\,n$. Since $d \neq x$, it follows that $k \neq m$. Since n is a 2-identifier of m, $L(k) = 1$. Note that k is odd. Thus, by Corollary 2.1, $k - 2 = 2^i d + xy/d - 2$ is prime.

Now suppose that (4) is true. To show that n is a 2-identifier for m, take $k \neq m$ such that $k\,\mathrm{PF}\,n$. We must show that $L(k) = 1$. Now k is a sum of two proper factors of n. If each were even, then k would be even also, and $L(k) = 1$ by Corollary 2.1. Thus, assume that k is odd.

Then k is the sum of factors of n having opposite parity, so $k = 2^i d + xy/d$ for some divisor d of xy. Since $k \neq m$, one has $d \neq x$ and since $xy/d \geqq 2$, one has $d \neq xy$. Then applying (4), one has that $k - 2$ is prime, which assures that $L(k) = 1$. This completes the proof.

Theorem 3 indicates that numbers expressible as a power of 2 plus a prime are of special interest.

COROLLARY 3.1 [GC]. *If $m = 2^i + p$ with p an odd prime, then $L(m) \leqq 2$ and $L(m) = 2$ if and only if $i > 1$ and $m - 2$ is not prime.*

Proof. If $L(m) \neq 1$, then Theorem 3 assures that $n = 2^i p$ is a 2-identifier for m. Indeed, the sufficient condition (4) is vacuously satisfied, since in this case, $x = 1$ and $xy = p$, so the only divisors of xy are x and xy.

In light of this corollary, it is of interest to know how many numbers can be expressed in the form $2^i + p$. A theorem of Romanoff [**9**] states that the set S of numbers of the form $a^i + p$, where a is any fixed positive integer and p is prime, has positive asymptotic density within the natural numbers. That is, the limit as $x \to \infty$ of the ratio $|\{y \in S | y < x\}|/x$ exists and is positive.

Our direct computations have shown that over 90% of the odd numbers less than 10,000 are expressible in the form $2^i + p$. However, P. Erdös [**5**] gives an arithmetic progression of odd numbers no term of which is of this form. He also conjectures that there are arbitrarily large gaps between successive numbers of this form. In [**3**], Crocker uses elementary methods to show that $2^{2^n} - 5$ is never expressible as $2^i + p$ for any $n \geqq 3$. Guy [**8**, pp. 24, 140] provides further background and references on this topic.

EXAMPLE 2. There is more to 2-identifiability than what follows from primes plus powers of 2. The following examples illustrate this.

(a) $L(147) = 2$, and $290 = 2 \cdot 5 \cdot 29$ is one of several 2-identifiers. Indeed, 147 PF 290 since $147 = 2 + 5 \cdot 29$. The divisors of $5 \cdot 29$ other than $x = 1$ and $xy = 145$ are 5 and 29, and one has that $2 \cdot 5 + 29 - 2 = 37$ and $2 \cdot 29 + 5 - 2 = 61$. Since 37 and 61 are both prime, $L(147) = 2$ by Theorem 3.

(b) The number $757 = 2^2 \cdot 139 + 3 \cdot 67$ is 2-identifiable, having $111{,}756 = 2^2 \cdot 3 \cdot 67 \cdot 139$ as a 2-identifier which can be shown to be unique. To verify that it is a 2-identifier, one must check the condition (4) of Theorem 3 for the six divisors of $3 \cdot 67 \cdot 139$ other than 139 and $3 \cdot 67 \cdot 139$. In all six cases, the number obtained is a prime. For example, $2^2 \cdot 67 \cdot 139 + 3 - 2 = 37{,}253$, which is a prime.

More examples of these more unusual 2-identifiable numbers are given in TABLE 1.

Theorems 2 and 3 characterize the numbers at levels one and two rather concretely. What can we say beyond this? Here at last E. Lemoine comes on the scene. According to Dickson's *History of the Theory of Numbers* [**4**, vol. I, p. 424], Lemoine made the following conjecture in 1894.

The sum m	The 2-identifier n	Number of d's satisfying (4)
$41 = 2^5 + 3^2$	$288 = 2^5 \cdot 3^2$	1
$217 = 2^2 \cdot 7^2 + 3 \cdot 7$	$4{,}116 = 2^2 \cdot 3 \cdot 7^3$	6
$367 = 2^4 + 3^3 \cdot 13$	$5{,}616 = 2^4 \cdot 3^3 \cdot 13$	6
$599 = 2^3 \cdot 71 + 31$	$17{,}608 = 2^3 \cdot 71 \cdot 31$	2
$757 = 2^2 \cdot 139 + 3 \cdot 67$	$111{,}756 = 2^2 \cdot 139 \cdot 3 \cdot 67$	6
$997 = 2^6 \cdot 3^2 + 421$	$242{,}496 = 2^6 \cdot 3^2 \cdot 421$	4
$2009 = 2^3 \cdot 41 + 41^2$	$551{,}368 = 2^3 \cdot 41^3$	2
$2213 = 2^2 + 47^2$	$8{,}836 = 2^2 \cdot 47^2$	1
$4633 = 2^2 + 3 \cdot 1543$	$18{,}516 = 2^2 \cdot 3 \cdot 1543$	2
$7405 = 2^{10} + 3^2 \cdot 709$	$6{,}534{,}144 = 2^{10} \cdot 3^2 \cdot 709$	4
$8887 = 2^2 \cdot 3 \cdot 193 + 6571$	$15{,}218{,}436 = 2^2 \cdot 3 \cdot 193 \cdot 6571$	6

TABLE 1. **Some level two numbers and their 2-identifiers.**

LEMOINE'S CONJECTURE [LC]. *Every odd number which is at least 7 can be expressed in the form $2p + q$ where p and q are prime.*

We now set forth a refinement of this conjecture which has significant consequences for our problem.

REFINED LEMOINE'S CONJECTURE [RLC]. *For any odd number m which is at least 9, there are odd prime numbers p, q, r and s and positive integers j and k such that $m = 2p + q$, $2 + pq = 2^j + r$ and $2q + p = 2^k + s$.*

Our interest in expressing m as $2p + q$ derives from the fact that $n = 2pq$ has only two P-partners besides m. Expressing these others in the form of a prime plus a power of 2 is clearly of interest in the light of the earlier results. We have checked this conjecture on a computer and found it to be valid for all odd m up to 53,000. Its plausibility rests upon the fact that odd numbers appear to be expressible in the form $2p + q$ in many ways, and that the numbers of the form $2^j + r$ with r prime are adequately dense. Its truth would imply that three is an upper bound for levels of identifiability, as we see in the following theorem.

THEOREM 4 [GC], [RLC]. *For any integer m with $m \geqq 4$, the level of identifiability of m is at most three.*

Proof. Take any such m and suppose that $L(m) > 2$. Then by Corollary 2.1, m is odd. We need only show that m has a 3-identifier. Let p, q, r, s, j and k be the numbers posited for m in [RLC]. Then $n = 2pq$ is an F-partner of m, and its only P-partners are m, $2 + pq$, and $2q + p$. By [RLC], these latter two are expressible as a power of 2 plus a prime, so by Corollary 3.1, their levels are one or two. This makes n a 3-identifier for m, and the proof is complete.

We wish to point out that three could be an upper bound for the levels of identifiability even if [RLC] should prove to be false. The conjecture simply allows for easily identifying a 3-identifier for m if $L(m) > 2$. The fact is, however, that because of the high density of the numbers at levels one and two, for any given F-partner of such an m, it is quite likely that all of its other P-partners will be at level one or two, making it a 3-identifier for m. TABLE 2 lists all of the numbers m less than 10,000 such that $L(m) = 3$; there are only 110 of them. For the smallest of these, namely 149, all but eight of its 74 F-partners are 3-identifiers for it. One of these eight nonidentifiers is $2898 = 2 \cdot 3^2 \cdot 7 \cdot 23$, since this number is an F-partner to $149 = 2 \cdot 3^2 \cdot 7 + 23$ and $331 = 2 \cdot 7 \cdot 23 + 3^2$, both of which are at level three.

331	3181	3341*	5731	6001	7331*		
373	1243	1783	2203	2263	2293	2503	2683
2983	3163	3343	3433	3643	4013*	4063	4153
4543	4573	4663	4813	4843	5143	5323	5923
6073	6193	6283	6403	6463	6673	6853	6883
7393	7583*	7603	7753	7783	7813	8023	8563
8923	9613						
905*	3505	3985	4195	4855	5125	6535	
337	877	907	1477	1597	1777	1807	1867
1927	2377	3187	3637	3697	3817	3877	3967
4327	4567	5077	5467	5557	5737	6247	6547
6637	6757	7177	7267	7297	7387	7417	7747
7867	8087*	8257	8287	8467	8527	8587	9307
9517	9557*	9787	9907				
149*	1199*	1549	1829*	1969	2879*	3299*	3739
6869*	8719	8929					

TABLE 2. **The 3-identifiable numbers less than 10,000, arranged in residue classes modulo 5. (Those congruent to 2 modulo 3 are marked with an asterisk. All others are congruent to 1 modulo 3.)**

Theorem 4 answers one of our questions about how long Prudence and Sam could go on repeating "I don't know your sum" and "I knew you didn't." This exchange could only happen twice. After two times, it is clear that Sam's sum is 3-identifiable, so as was noted above, it is highly likely that Prudence will next say "Now I know your sum," leaving Sam quite in the dark regarding the product. On the other hand, Prudence may have to say "I don't know your sum" a third time. This would happen for example if she had 2898 and Sam had one of its two 3-identifiable P-partners 149 or 331. Her third "I don't know" has given him considerable information. If he had 149, he now knows that she has one of the eight nonidentifiers of his number. Whether there is a case in which, after Prudence's third "I don't know," Sam would, in fact, know her product is an open question. In our terminology, this question translates into:

Does there exist a 3*-identifiable number m such that all but one of the F-partners of m are* 3*-identifiers for it*?

Let us look at how our results apply to analyzing variants on the problems such as the one in (3). As before, Prudence's first "I don't know" tells us that she does not have a 1-identifier for Sam's sum. Since Sam says "Thanks" rather than "I knew you didn't," we conclude that his m is at level one. He thanks Prudence, because she has told him that he can remove from further consideration the 1-identifiers for his number. However, there must be at least two nonidentifying F-partners of his m, or he would have said "I know." The reader is encouraged to continue the analysis, and verify that $m=8$ and $n=16$ is a solution for this puzzle. For a bigger challenge, the reader may wish to find solutions for this one:

P: I don't know your sum.
S: Thank you for that information.
P: I still don't know your sum.
S: Thank you for that information. (5)
P: Thank you for that information.
S: Now I know your product.
P: And now I know your sum.

In addition to taking on variations of the original problem such as (3) and (5), one may wish to construct three-person versions in which the third person (Squire Adam?) is given the sum of the squares of the two numbers, and the three then exchange information. Note that once one of the participants figures out the number of a second one, the number of the third becomes known (because of the equation $x^2+y^2=(x+y)^2-2xy$).

The study of these problems and this theoretical framework can go much farther. For example, we have derived improved ways of screening for potential 2-identifiers. Theorem 3 by itself gives a characterization, but it requires testing all $[m/2]-1$ of the F-partners of a given m and all of their odd divisors. Even with a computer, this is time consuming if m is large. Fortunately, one can often find far more stringent necessary conditions on 2-identifiers. For example, if $m\equiv 1 \bmod 3$, and if m has been partitioned as 2^ix+y with i odd, then $n=2^ixy$ can be a 2-identifier for m only if $x=1$ and y is prime. If $m\equiv 2 \bmod 3$ and i is even, then 2^ixy can be a 2-identifier only if $x=1$ and y is prime or a product of two primes. Thus, the sufficient condition of Corollary 3.1, or a close approximation to it, becomes necessary in these cases. Similar results, although not as restrictive, hold for other cases of the congruence class of m modulo 3 and 5.

Is all of this worthwhile, given that our main results hinge upon unproven conjectures? We think so, for two reasons. First, the results are valid for all $m\leq k$ provided that the conjectures hold for all numbers less than a quadratic function of k. Thus, they are applicable for analyzing reasonably large know/don't know problems since the conjectures are known to hold on large initial segments of the natural numbers. But more importantly, the study has directed our attention to more subtle aspects of the additive theory of prime numbers. Our conjecture [RLC] reflects this, dealing with interaction of sums involving primes whereas [GC] and [LC] deal with such sums only individually. This conjecture and the open questions about numbers at levels two and three are of interest in their own right because of the issues they raise within this fascinating and often baffling additive realm of the prime numbers.

References

[1] A. K. Austin, A calculus for know/don't know problems, this MAGAZINE, 49 (1976) 12–14.

[2] Chen Jing-Run, On the representation of a large even number as the sum of a prime and the product of at most two primes, Sci. Sinica, 16 (1973) 157–176.

[3] Roger Crocker, A theorem concerning primes, this MAGAZINE, 34 (1960/61) 316, 344.

[4] L. E. Dickson, History of the Theory of Numbers, Chelsea, New York, 1952; reprint of the 1919 edition.

[5] P. Erdös, On integers of the form $2^k + p$ and some related problems, Summa Brasil Math., 2 (1947–51) 113–123.

[6] T. Ferguson, Problem 1173, this MAGAZINE, 56 (1983) 177, and Solution, 57 (1984) 180–181.

[7] M. Gardner, Mathematical games, Scientific American, 241 (Dec., 1979) 22.

[8] R. K. Guy, Unsolved Problems in Number Theory, Springer-Verlag, New York, 1981.

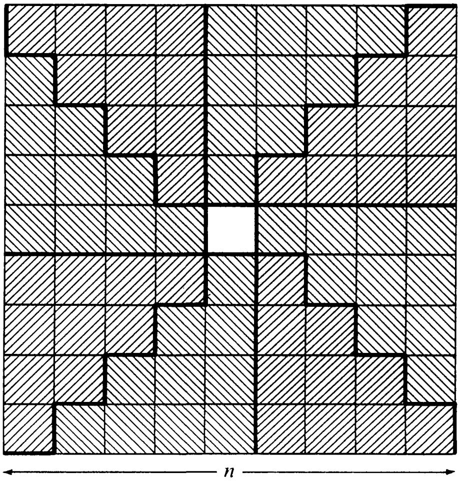
[9] N. P. Romanoff, Über einige Sätze der additiven Zahlentheorie, Math. Ann., 109 (1934) 668–678.

[10] D. Shanks, Solved and Unsolved Problems in Number Theory, Vol. I, Spartan Books, Washington, DC., 1962, 2nd ed., Chelsea, New York, 1978.

[11] D. Sprows, Problem 977, this MAGAZINE, 49 (1976) 96, and Solution, 50 (1977) 268.

Proof without words: Square of an odd positive integer

$$n^2 - 1 = 4\left(\frac{n-1}{2}\right)\left(\frac{n+1}{2}\right) = 8\sum_{i=1}^{(n-1)/2}$$



—EDWIN G. LANDAUER
General Physics Corp.

such sums only individually. This conjecture and the open questions about numbers at levels two and three are of interest in their own right because of the issues they raise within this fascinating and often baffling additive realm of the prime numbers.

References

[1] A. K. Austin, A calculus for know/don't know problems, this MAGAZINE, 49 (1976) 12–14.

[2] Chen Jing-Run, On the representation of a large even number as the sum of a prime and the product of at most two primes, Sci. Sinica, 16 (1973) 157–176.

[3] Roger Crocker, A theorem concerning primes, this MAGAZINE, 34 (1960/61) 316, 344.

[4] L. E. Dickson, History of the Theory of Numbers, Chelsea, New York, 1952; reprint of the 1919 edition.

[5] P. Erdös, On integers of the form $2^k + p$ and some related problems, Summa Brasil Math., 2 (1947–51) 113–123.

[6] T. Ferguson, Problem 1173, this MAGAZINE, 56 (1983) 177, and Solution, 57 (1984) 180–181.

[7] M. Gardner, Mathematical games, Scientific American, 241 (Dec., 1979) 22.

[8] R. K. Guy, Unsolved Problems in Number Theory, Springer-Verlag, New York, 1981.

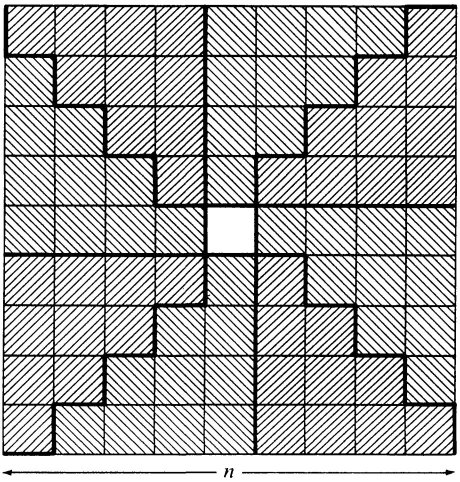
[9] N. P. Romanoff, Über einige Sätze der additiven Zahlentheorie, Math. Ann., 109 (1934) 668–678.

[10] D. Shanks, Solved and Unsolved Problems in Number Theory, Vol. I, Spartan Books, Washington, DC., 1962, 2nd ed., Chelsea, New York, 1978.

[11] D. Sprows, Problem 977, this MAGAZINE, 49 (1976) 96, and Solution, 50 (1977) 268.

Proof without words:
Square of an odd positive integer

$$n^2 - 1 = 4\left(\frac{n-1}{2}\right)\left(\frac{n+1}{2}\right) = 8\sum_{i=1}^{(n-1)/2}$$



—EDWIN G. LANDAUER
General Physics Corp.

Prime-producing Polynomials and Principal Ideal Domains

If a certain polynomial yields "enough" prime values, then a corresponding number ring will be a principal ideal domain, and conversely.

DANIEL FENDEL
San Francisco State University
San Francisco, CA 94132

Consider the well-known polynomial

$$x^2+x+41,$$

which produces prime values for every integer x with $0 \le x \le 39$. A classic problem is to find the constants C which could replace 41. That is, we ask:

For what integers $C \ge 1$ does the polynomial

$$x^2+x+C$$

produce prime values for all integers x with $0 \le x \le C-2$?

(Of course, $C-2$ is the largest upper limit on x for which such an assertion could be true; for if $x=C-1$, then $x^2+x+C=C^2$, which is not prime.)

Interestingly, all such values C are known, and 41 is the largest of them. The values of C which answer the above question are:

$$C=1,2,3,5,11,17, \text{and } 41.$$

There is a natural connection between the polynomials x^2+x+C and imaginary quadratic fields. We can see this by factoring the polynomial over the complex numbers:

$$x^2+x+C=(x+\alpha)(x+\bar{\alpha}),$$

where

$$\alpha=\frac{1+\sqrt{1-4C}}{2}, \qquad \bar{\alpha}=\frac{1-\sqrt{1-4C}}{2};$$

$\bar{\alpha}$ is the complex conjugate of α. For convenience, we set

$$n=4C-1.$$

It is reasonable to look for a relationship between the "prime-producing" character of the polynomial x^2+x+C and factorization in the field $Q(\sqrt{-n})$. In fact, a strong relationship of this type does exist. Specifically, let D_n be the ring of "algebraic integers" in $Q(\sqrt{-n})$. (This will be defined and described in the next section.) We will prove the following (given as Theorem 4 below):

(I) *If D_n is a unique factorization domain (UFD), with $n=4C-1$, then the polynomial x^2+x+C produces prime values for all integers x with $0 \le x \le C-2$.*

Perhaps more surprising is that there is also a connection between these polynomials and the question of whether D_n is a principal ideal domain. We will prove a result of the following type:

(II) *If the polynomial x^2+x+C produces prime values for "enough" integers x, then D_n is a principal ideal domain (PID).*

The "enough" here turns out to be an interval $0 \leq x \leq C^*$, where C^* depends on n, but is always less than or equal to $C-2$. The details are spelled out in Theorem 3. Because of the elementary fact that every PID is a UFD (see [2]), statement (II) is thus a kind of "strong converse" to (I).

Together with a simple discussion of the cases $n \equiv 1$ or 2 mod 4, (I) and (II) constitute an elementary proof of the following well-known result:

COROLLARY. *If D_n is a unique factorization domain, then it is also a principal ideal domain.*

(There is actually a much broader result known for general algebraic number rings, but the proof requires considerable background in ideal theory (see [4]). However, it is *not* true for arbitrary rings that a UFD must be a PID.)

Our results (I) and (II) also allow us to deduce the complete list of values for C given earlier, based on the following very deep theorem of Stark (see [3]):

THEOREM (Stark). *D_n is a principal ideal domain (for positive n) if and only if n is one of the following values:*

$$n = 1, 2, 3, 7, 11, 19, 43, 67, 163.$$

Since we are using $n = 4C-1$, we can ignore the values $n = 1, 2$. The remaining seven values of n give precisely the values of C listed earlier.

The proofs of (I) and (II) are based on the complex norm ϕ, defined as follows:

$$\phi(\gamma) = \gamma\bar{\gamma} = |\gamma|^2, \quad \text{for } \gamma \in \mathbb{C}.$$

There is a simple condition using ϕ (given as Theorem 1 below) for determining which D_n's are euclidean domains "with respect to ϕ." (The precise meaning of this phrase is given later.) If we visualize D_n and the field $K = Q(\sqrt{-n})$ in the complex plane, then Theorem 1 can be expressed in geometric terms as follows:

(III) *D_n is a euclidean domain (ED) with respect to ϕ if and only if it satisfies:*
(*) *if γ is in K, then it is within one unit of some element of D_n.*

Using the elementary fact that every ED is a PID (see [2]), (III) yields *some* of the values of n in Stark's list. But not all: in particular, the last four values ($n = 19, 43, 67, 163$) give rings D_n which are principal ideal domains but not euclidean domains with respect to ϕ. (An elementary proof that D_{19} is not euclidean under any norm is given by Wilson [5].) Therefore something more subtle than (III) is needed to handle PID's.

The key idea in the proof of (II) is the existence of an analogue to Theorem 1 for identifying PID's. This analogue (given as Theorem 2 below) can also be expressed geometrically, as follows:

(IV) *D_n is a principal ideal domain if and only if it satisfies:*
(**) *if γ is in K but not in D_n, then some multiple $\chi\gamma$ of γ (with χ in D_n) is within one unit of, but not equal to, some element of D_n.*

(Note: (IV) extends naturally to arbitrary algebraic number fields. In the extended version, K is any algebraic number field, D is its ring of integers, and "distance" is measured by the field norm.)

The bulk of the proof of (II) consists of an analysis of condition (**) above. This analysis is eventually tied in with our polynomials by the fact that $\phi(x+\alpha) = x^2 + x + C$, for integers x (with α as defined earlier).

Preliminaries

We consider the field $K = Q(\sqrt{-n})$, where Q is the field of rational numbers, and n is a positive, square-free integer. Thus, modulo 4, n is congruent to 1, 2, or 3. The case $n \equiv 3$ mod 4 is of primary importance for this paper, since it corresponds to the situation of the polynomial $x^2 + x + C$, where $n = 4C-1$, in our opening question.

Recall that an element of K is an **algebraic integer** if it is the root of some monic polynomial with integral coefficients. The algebraic integers within K form a ring, which will be denoted by D_n. All congruences considered here will be modulo 4 unless otherwise indicated, so we will write $n \equiv a$ to mean $n \equiv a$ mod 4. We will also use the following standard notation:

Z: *the ring of integers*

(γ): *the ideal generated by an element γ in D_n*

$[a]$: *the largest integer m such that $m \leq a$*

$a|b$: *a is a divisor of b (where a and b are in Z).*

The following lemma gives a concrete description of the ring D_n:

LEMMA 1. *D_n is the set of elements of the form $a + b\alpha$, with a and b in Z, where*

$$\alpha = \begin{cases} \sqrt{-n} & \text{if } n \equiv 1 \text{ or } 2 \\ \dfrac{1+\sqrt{-n}}{2} & \text{if } n \equiv 3. \end{cases}$$

(For a proof, see [**1**].) In terms of the complex plane, Lemma 1 says that the elements of D_n form a lattice, which will look like FIGURE 1 or FIGURE 2, depending on whether $n \equiv 1, 2$ or $\mathrm{n} \equiv 3$.

Elements of K can be written as $a + b\alpha$, with a and b in Q. We can express the norm $\phi(\gamma) = |\gamma|^2$ on K in terms of this description, as follows:

$$\phi(a + b\alpha) = \begin{cases} a^2 + nb^2 & \text{if } n \equiv 1 \text{ or } 2 \\ \left(a + \dfrac{b}{2}\right)^2 + \dfrac{nb^2}{4} = a^2 + ab + \dfrac{n+1}{4}b^2 & \text{if } n \equiv 3. \end{cases}$$

Note that, in the $n \equiv 3$ case, if we set $a = x$ and $b = 1$, we obtain

$$\phi(x + \alpha) = x^2 + x + C, \qquad \text{where } C = \frac{n+1}{4}. \tag{1}$$

The following is a summary of some elementary facts we will need about ϕ:

LEMMA 2. (i) $\phi(\gamma_1\gamma_2) = \phi(\gamma_1)\phi(\gamma_2)$.
(ii) *if $\gamma \neq 0$, then $\phi(\gamma) > 0$.*
(iii) *if $\gamma \in D_n$, then $\phi(\gamma) \in Z$.*
(iv) *if $\gamma \in D_n$, and $\phi(\gamma) = 1$, then γ is a unit.*
(v) *if γ_1 and γ_2 are in D_n, with $(\gamma_1) \subsetneq (\gamma_2)$, then $\phi(\gamma_2) < \phi(\gamma_1)$.*
(vi) *if a, b, c, d, and t are integers, with $a \equiv c$ mod t and $b \equiv d$ mod t, then $\phi(a + b\alpha) \equiv \phi(c + d\alpha)$ mod t.*
(vii) *if $n \equiv 3$ mod 4 and $x \in Z$, then $\phi(x + \alpha) = \phi(-1 - x + \alpha)$.*

(Verification of these properties of ϕ is left to the reader.) We also need the following result, which says, in effect, that elements of $D_n \setminus Z$ cannot be "small":

LEMMA 3. *Suppose $\gamma \in D_n \setminus Z$.*
(i) *If $n \equiv 1$ or 2, then $\phi(\gamma) \geq n$.*
(ii) *If $n \equiv 3$, then $\phi(\gamma) \geq (n+1)/4$.*

Proof. Write γ as $a + b\alpha$, so $b \neq 0$. Thus (i) is obvious. If $|b| = 1$, then $(a + b/2)^2 \geq 1/4$, so (ii) follows. But if $|b| > 1$, then $\phi(\gamma) \geq nb^2/4 \geq n$, and (ii) follows as well.

Finally, we have the following simple consequence.

LEMMA 4. *If $n > 3$ with $n \equiv 3$ and $0 \leq t \leq \sqrt{n/3}$, then the equation $t = \phi(x + \alpha)$ has no integral solution for x.*

This follows from Lemma 3, (ii), since $\sqrt{n/3} < (n+1)/4$ for $n > 3$, and $x + \alpha$ is in $D_n \setminus Z$.

Conditions for euclidean and principal ideal domains

We say that a ring D of complex numbers is a euclidean domain (ED) (with respect to the norm ϕ) if

(i) $\phi(\gamma)$ is an integer for all γ in D,

and

(ii) (division algorithm) if γ_1 and γ_2 are in D, with $\gamma_2 \neq 0$, then there are elements in δ and η in D satisfying $\gamma_1 = \gamma_2\delta + \eta$, and such that $\phi(\eta) < \phi(\gamma_2)$.

The following theorem is a formal statement of result (III) from the introduction.

THEOREM 1. *The following are equivalent*:
(i) D_n *is a euclidean domain.*
(ii) *For each* $\gamma \in K$, *there exists a* $\delta \in D_n$ *such that* $\phi(\gamma - \delta) < 1$.

Proof. (i) → (ii): Suppose $\gamma \in K$, and let t be an integer such that $t\gamma \in D_n$, and divide $t\gamma$ by t using the division algorithm. This gives $t\gamma = t\delta + \eta$, with δ and η in D_n and $\phi(\eta) < \phi(t)$. Then $\phi(\gamma - \delta) = \phi(\eta/t) < 1$.

(ii) → (i): First note that $\phi(\gamma) \in Z$ for all γ in D_n, by Lemma 2, (iii). Next, suppose that γ_1 and γ_2 are in D_n, with $\gamma_2 \neq 0$. Set $\gamma = \gamma_1/\gamma_2$, and choose $\delta \in D_n$ as provided so that $\phi(\gamma - \delta) < 1$, and set $\eta = \gamma_1 - \gamma_2\delta$. Then $\gamma_1 = \gamma_2\delta + \eta$, and $\phi(\eta) = \phi(\gamma_2)\phi(\gamma - \delta) < \phi(\gamma_2)$, as desired.

Using Theorem 1 and FIGURES 1 and 2, it is fairly routine to show the following:

COROLLARY 1. D_n *is a euclidean domain* (*with respect to* ϕ) *if and only if* n *is one of the following values*:

$$n = 1, 2, 3, 7, 11.$$

We will need the cases $n = 1$ and $n = 2$ to complete the discussion of the situation where $n \equiv 1$ or 2. The cases $n = 3$ and 7 will allow us to avoid problems with later inequalities.

We now give the analogue of Theorem 1 for principal ideal domains. (The following is (IV) from the introduction.)

THEOREM 2. *The following are equivalent*:
(i) D_n *is a principal ideal domain.*
(ii) *For each* $\gamma \in K \setminus D_n$, *there exist* χ *and* δ *in* D_n *such that* $0 < \phi(\chi\gamma - \delta) < 1$.

Proof. (i) → (ii): Suppose $\gamma \in K \setminus D_n$, and let t be an integer such that $t\gamma \in D_n$. Let I be the ideal of D_n generated by $t\gamma$ and t. By assumption, there exists $\beta \in I$ with $I = (\beta)$. Choose χ and δ in D_n with $\beta = \chi(t\gamma) - \delta t$. Since $\gamma \notin D_n$, we have $t\gamma \notin (t)$, so $(t) \subsetneq (\beta)$. By Lemma 2, (v), we have $\phi(\beta) < \phi(t)$. Since $\beta \neq 0$, we have $0 < \phi(\beta/t) = \phi(\chi\gamma - \delta) < 1$, as desired.

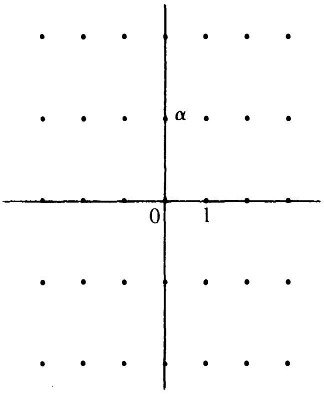


FIGURE 1. $n \equiv 1$ **or 2 mod 4.**

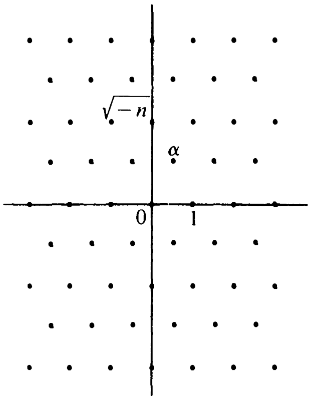


FIGURE 2. $n \equiv 3$ **mod 4.**

(ii) $\to$ (i): Let I be a nonzero ideal of D_n, and choose $\beta \in I$, $\beta \neq 0$, with $\phi(\beta)$ minimal. Thus $(\beta) \subseteq I$. Suppose $I \neq (\beta)$, so there exists $\theta \in I \setminus (\beta)$. Let $\gamma = \theta/\beta \in K \setminus D_n$, and choose χ and δ in D_n as described in (ii), so $0 < \phi(\chi\gamma - \delta) < 1$. Then $\chi\theta - \delta\beta = (\chi\gamma - \delta)\beta \in I \setminus \{0\}$, and so $0 < \phi(\chi\theta - \delta\beta) = \phi(\chi\gamma - \delta)\phi(\beta) < \phi(\beta)$, contradicting the choice of β. Thus, $I = (\beta)$, so D_n is a principal ideal domain.

The polynomial $x^2 + x + C$ and principal ideal domains

Our goal in this section is to prove the following more precise version of (II) from the introduction (recall $n = 4C - 1$).

THEOREM 3. *If $x^2 + x + C$ produces prime values for all integers x with $0 \le x \le \left[\frac{1}{2}\sqrt{n/3}\right]$, then D_n is a principal ideal domain.*

Thus, the C^* of (II) is actually $[(1/2)\sqrt{n/3}]$. Clearly $C^* < (n-7)/4$ $(= C - 2)$ for large n; in fact, this holds for $n \ge 11$. Corollary 1 already tells us that D_3 and D_7 are PID's, and we shall assume $n \ge 11$.

Our results will therefore give us the following curious situation. The primality of $x^2 + x + C$ over the short interval $0 \le x \le [(1/2)\sqrt{n/3}]$ will guarantee that D_n is a PID, and hence also a UFD. We will see (Theorem 4) that this in turn guarantees the primality of the polynomial over the generally longer interval $0 \le x \le (n-7)/4$!

In the proof of Theorem 3, we will use the identity (1),

$$\phi(x + \alpha) = x^2 + x + C$$

together with the criterion for PID's given by Theorem 2. Thus, following Theorem 2, we consider an arbitrary $\gamma \in K \setminus D_n$. We must find elements χ, $\delta \in D_n$ such that $0 < \phi(\chi\gamma - \delta) < 1$.

The following technical lemma is based on a famous approximation theorem of Dirichlet. It holds for any field $K = Q(\sqrt{-n})$, $n \equiv 3 \bmod 4$, and any $\gamma \in K$, whether or not the ring D_n is a PID. We defer the proof to the end of our article.

LEMMA 5. *There is a positive integer t, with $t \le \sqrt{n/3}$, and an element δ in D_n, such that $\phi(t\gamma - \delta) < 1$.*

We shall now make two attempts to satisfy condition (ii) of Theorem 2, first with $\chi = t$ (as in Lemma 5), and if that fails, with $\chi = t\bar{\gamma}$. If both of these fail, we shall show that the polynomial $x^2 + x + C$ takes a composite value somewhere in the interval $0 \le x \le C^*$, contradicting the assumption of Theorem 3. Here are the details:

Let t be the smallest integer satisfying Lemma 5, and δ as provided there. If $t\gamma$ is not in D_n, then we also have $0 < \phi(t\gamma - \delta)$, and so we have fulfilled condition (ii) of Theorem 2, using $\chi = t$. So we now assume $t\gamma \in D_n$. This implies that $t\bar{\gamma}$ is also in D_n, and so we can use it as a new candidate for χ. Thus let $\chi = t\bar{\gamma}$. Then $\chi\gamma = (1/t)\phi(t\gamma)$, which is a rational number, and so $\chi\gamma$ must in fact be less than one unit from some ordinary integer δ in D_n. Thus once again we will have satisfied (ii) of Theorem 2, unless $\chi\gamma = \delta$, i.e., $\chi\gamma \in Z$. This can only happen if $t | \phi(t\gamma)$.

The following lemma tells us what we need in order to prevent that:

LEMMA 6. *If $t|\phi(t\gamma)$, then $\phi(x + \alpha)$ is composite for some integer x with $0 \le x < t/2$.*

Proof. Since $t\gamma \in D_n$, we can write $t\gamma = a + b\alpha$, with a, $b \in Z$. We first show that b and t are relatively prime, as follows: any prime dividing t must also divide $\phi(t\gamma)$ by hypothesis, but $\phi(t\gamma) = a^2 + ab + ((n+1)/4)b^2$. Thus any prime which divides both b and t must also divide a^2, and hence a. This would mean that a, b, and t would have a common factor, contradicting the minimality of t.

Now, since b and t are relatively prime, there exists $y \in Z$ with $yb \equiv 1 \bmod t$. We then find $x \in Z$, with $ya \equiv x \bmod t$; we can choose x so that $-t/2 \le x < t/2$. Thus $\phi(yt\gamma) = \phi(ya + yb\alpha) \equiv \phi(x + \alpha) \bmod t$ (see Lemma 2, (vi)). By assumption, $t|\phi(t\gamma)$, and so clearly $t|\phi(yt\gamma)$, and hence $t|\phi(x + \alpha)$. But $t \neq \phi(x + \alpha)$ by Lemma 4 (we have $n > 3$ here). On the other hand, Theorem 2 provides that $\gamma \notin D_n$, but we are assuming $t\gamma \in D_n$, and so $t \neq 1$. Thus $\phi(x+\alpha)$ must be composite.

Finally, we can improve the restriction on x as follows: if $-t/2 \leq x < 0$, then we let $x^* = -1 - x$, which satisfies $0 \leq x^* < t/2$. Since $\phi(x+\alpha) = \phi(-1-x+\alpha)$ (Lemma 2, (vii)), we have that $\phi(x^*+\alpha)$ is also composite, completing the proof.

Thus, to get D_n to be a PID, we need only assure that the conclusion of Lemma 6 is false. Using $t \leq \sqrt{n/3}$, and the identity $\phi(x+\alpha) = x^2 + x + C$, this is precisely the hypothesis of Theorem 3.

The polynomial $x^2 + x + C$ and unique factorization domains

Before looking at our specific situation, we mention an elementary result about UFD's in general. Recall that an element w of a ring is called **irreducible** if a factorization $w = uv$ implies that u or v is a unit. We will need the following well-known result.

LEMMA 7. *If a ring D is a unique factorization domain, and an irreducible element $w \in D$ divides a product of elements in D, then w divides one of the factors.* (For a proof, see [**2**].)

The main result of this section is the following (this is (I) from the introduction):

THEOREM 4. *Suppose that $n \equiv 3$. If D_n is a unique factorization domain, then $x^2 + x + C$ produces prime values for all integers x with $0 \leq x \leq C - 2$ (where $C = (n+1)/4$).*

It turns out that we can take care of the cases $n \equiv 1, 2$ with the same basic analysis. The result in that case is the following.

THEOREM 5. *Suppose $n \equiv 1$ or 2. If $n > 2$, then D_n is not a unique factorization domain.*

Corollary 1 tells us that D_1 and D_2 are ED's, and hence PID's and UFD's. Using that fact and Theorem 5 if $n \equiv 1$ or 2, and Theorems 3 and 4 if $n \equiv 3$, we get the following consequence, mentioned in the introduction:

COROLLARY 2. *If D_n is a unique factorization domain, then it is also a principal ideal domain.*

We now turn to the proofs of Theorems 4 and 5, initially handling all cases together. We noted in Lemma 3 that there is a lower bound for $\phi(\gamma)$ if γ is in $D_n \setminus Z$. For convenience in handling the different cases, we set

$$L = \begin{cases} n & \text{if } n \equiv 1 \text{ or } 2 \\ \dfrac{n+1}{4} & \text{if } n \equiv 3. \end{cases}$$

Thus, if $\gamma \in D_n \setminus Z$, then $\phi(\gamma) \geq L$. From this we get the following:

LEMMA 8. *If p is a prime in Z, with $p < L$, then p is irreducible in D_n.*

Proof. Suppose $p = \gamma_1\gamma_2$, with $\gamma_1, \gamma_2 \in D_n$, and neither a unit. Then γ_1 and γ_2 are not integers, since p is a prime, so $p^2 = \phi(p) = \phi(\gamma_1)\phi(\gamma_2) \geq L^2$, which is a contradiction.

LEMMA 9. *If D_n is a UFD and $a \in Z$, then $\phi(a+\alpha)$ has no prime factors less than L.*

Proof. Suppose p is such a prime, so it is irreducible by Lemma 8. Then $p | (a+\alpha)$ or $p | \overline{(a+\alpha)}$ by Lemma 7 since $\phi(a+\alpha) = (a+\alpha)\overline{(a+\alpha)}$. If $n \equiv 1$ or 2 then $\overline{a+\alpha} = a - \alpha$; if $n \equiv 3$ then $\overline{a+\alpha} = a + 1 - \alpha$. In either case, p divides neither $a + \alpha$ nor $\overline{a+\alpha}$, since the coefficient of the basis element α is ± 1.

We leave it to the reader to verify the following simple inequality:

LEMMA 10. *If $n \equiv 3$ and $0 \leq x \leq (n-7)/4$, then $\phi(x+\alpha) < L^2$.*

Our main results are now easy.

Proof of Theorem 4. Suppose $\phi(x+\alpha) = x^2 + x + C$ is not prime, with x in the given range of values. Then $\phi(x+\alpha) < L^2$, by Lemma 10, and so $\phi(x+\alpha)$ has a prime factor less than L, contradicting Lemma 9.

Proof of Theorem 5. We have $\phi(n+\alpha) = n^2 + n$, so $\phi(n+\alpha)$ has the prime factor 2. But $2 < L$ by assumption (here $L = n$). Thus Lemma 9 says D_n cannot be a UFD.

Proof of Lemma 5. The following result concludes the proof of Theorem 3.

LEMMA 5. *Suppose $n \equiv 3$. For any $\gamma \in K$, there is a positive integer t, with $t \leq \sqrt{n/3}$, and an element δ in D_n, such that $\phi(t\gamma - \delta) < 1$.*

To prove Lemma 5, we write $\gamma = a + b\alpha$ and set $m = [\sqrt{n/3}\,] + 1$. Our final lemma tells how to choose t:

LEMMA 11. *Let m be an integer ≥ 2, and $b \in Q$. Then there exists $t \in Z$, with $1 \leq t \leq m-1$, and $m_1 \in Z$, with $|tb - m_1| \leq 1/m$.*

Proof. The proof uses the "pigeonhole principle." Let $((x))$ denote the fractional part of x, i.e., $((x)) = x - [x]$. Set $b_j = ((jb))$, $j = 1, \ldots, m-1$, and $I_j = [j/m, (j+1)/m]$, $j = 0, \ldots, m-1$. If some b_t is in either I_0 or I_{m-1}, then tb is within $1/m$ of an integer, as desired. If not, then we have $m-1$ b_j's and only $m-2$ remaining intervals, so two b_j's must be in the same interval. Thus, some b_r and b_s are in the same interval, with $1 \leq r < s \leq m-1$. Then $(s-r)b$ is within $1/m$ of some integer, so $t = s - r$ satisfies the stated condition.

We now complete the proof of Lemma 5. Choose t and m_1 as in Lemma 11, and set $c = tb - m_1$. Then choose $m_2 \in Z$ as close as possible to $ta + c/2$, (so that $|ta + c/2 - m_2| \leq 1/2$), and set $\delta = m_2 + m_1\alpha$. Then

$$\begin{aligned}\phi(t\gamma - \delta) &= \phi\left[(ta - m_2) + (tb - m_1)\alpha\right]\\ &= \phi\left[(ta - m_2) + c\alpha\right]\\ &= \left(ta - m_2 + \frac{c}{2}\right)^2 + \frac{n}{4}c^2\\ &\leq \frac{1}{4} + \frac{n}{4}\cdot\frac{1}{m^2}\\ &< \frac{1}{4} + \frac{n}{4}\cdot\frac{3}{n} = 1,\end{aligned}$$

as desired.

I wish to express my thanks to the referee whose suggestions were very helpful in preparing the final draft of this article.

References

[1] William W. Adams and Larry Joel Goldstein, Introduction to Number Theory, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1976, p. 216.

[2] Thomas Hungerford, Algebra, Springer-Verlag, New York, 1974, Section III.3.

[3] Harold Stark, A complete determination of the complex quadratic fields of class-number one, Michigan Math J., 14 (1967) 1–27.

[4] Edwin Weiss, Algebraic Number Theory, McGraw-Hill, New York, 1963, chapter 4.

[5] J. C. Wilson, A principal ideal ring that is not a euclidean ring, this MAGAZINE, 46 (1973) 34–38.

NOTES

The Slice Group in Rubik's Cube

David Hecker
Ranan Banerji
Saint Joseph's University
Philadelphia, PA 19131

The different maneuvers on Rubik's cube can be thought of as a set of transformations forming a subgroup of the group of all permutations of the cube's 54 facelets. In this paper, we shall do a complete study of the small subgroup of this group generated by turning only the center layers of the cube. This is known as the **slice group**. Our purpose in this study is to use the cube to illustrate several fundamental group theoretic techniques.

Before an accurate statement of the problem can be made, we must introduce some notation, most of which is standard in the cube literature [4]. We shall describe all maneuvers as if they were carried out with respect to the cube held in a fixed position, referring to the six faces by the "colors" F, B, R, L, U, and D for front, back, right, left, up, and down. The cube will be assumed to start with all six faces having solid colors, known as the **clean state** of the cube (see FIGURE 1). A clockwise 90° turn of any of these six faces (clockwise looking at the face "from outside") will be denoted by T_F, T_B, etc. (see FIGURE 2). We shall also talk about a 90° clockwise turn of the entire cube looking at the corresponding face from outside and denote these by C_F, C_B, etc. Every possible maneuver M can be written as a finite sequence of these fundamental moves, hence these moves generate the group of all possible transformations of the cube. The inverse of any move M, denoted by M^{-1}, is the maneuver required to undo the effect of the move M. Note that the actual twists of the cube needed to accomplish this will not be unique. To avoid this problem, when speaking of a move or maneuver on the cube, we refer only to its effective **permutation** of the cube's 54 facelets, and not to the actual twists required. Two different sequences of twists having the same effect are considered to be the same move. Repeating any move a number of times will be expressed with the usual exponential notation.

In addition to referring to the cube's facelets, we shall also refer to the 27 subcubes of the cube,

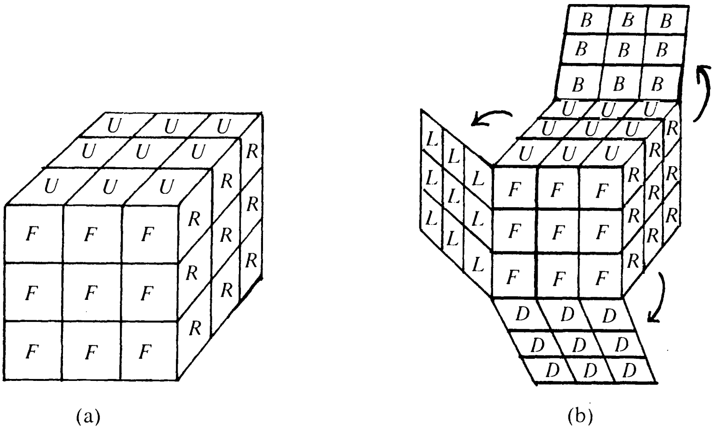


FIGURE 1. **(a) Rubik's cube in the clean state. (b) Rubik's cube in the clean state with hidden sides displayed.**

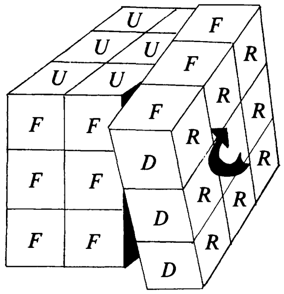


FIGURE 2. The move T_R.

often called **cubies** in the literature [4]. One of these cubies lies in the interior of the cube and cannot be seen. Eight others appear on the corners of the cube; these each have 3 facelets showing. Twelve are edge cubies with 2 facelets showing, and the remaining six lie at the center of each face and have only one facelet visible. The cubies are important since the facelets on each cubie stay on that cubie so that a large number of the 54! permutations in the group of all permutations of the 54 facelets are prohibited. Also, every move takes corner cubies to corner positions, edge cubies to edge positions, and center cubies to center positions, thus further limiting the possible permutations. The position of any facelet or cubie when the cube is in the clean state is called the **home position** of that facelet or cubie. The clean state is completely characterized by all facelets being in their home position. We think of this position as representing the identity permutation on the cube.

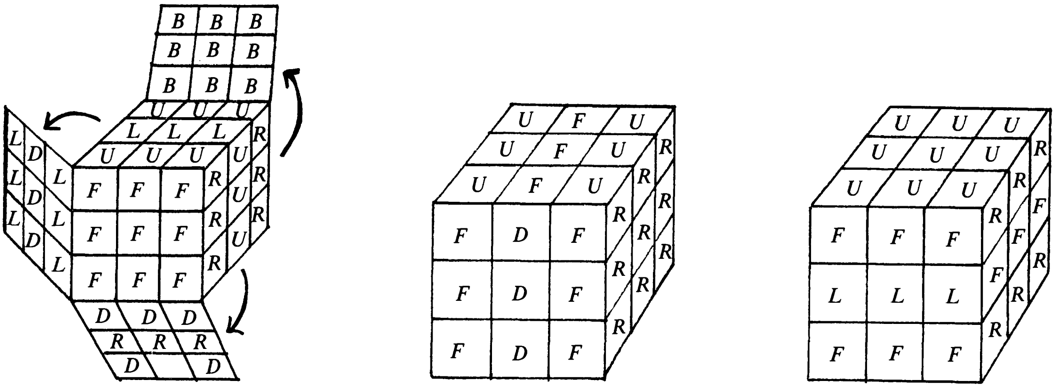
The slice group

The **slice group** is the group of transformations of the cube generated by the following three move sequences:

$$T_BT_F^{-1}C_B^{-1}, \quad T_LT_R^{-1}C_L^{-1}, \quad T_UT_D^{-1}C_U^{-1}. \qquad (1)$$

These are illustrated in FIGURE 3. One can think of them in two ways. First, consider them as turning two opposite faces "parallelly," one clockwise and one counterclockwise, then turning the whole cube to return those two faces to their original positions. Alternately, one can consider the overall effect, which is to turn the center slice of the cube clockwise 90° when looking at the F, R, and D faces respectively.

Our discussion will be from this latter point of view. We shall denote the three sequences in (1) by the single letter F, R, and D, which will stand for turns of the corresponding center slice. We



$F : T_BT_F^{-1}C_B^{-1}$ $R : T_LT_R^{-1}C_L^{-1}$ $D : T_UT_D^{-1}C_U^{-1}$

FIGURE 3. The three moves which generate the slice group.

have thus given these letters two meanings, as we have also used them to represent colors of the faces. The meaning in a given instance should be clear from the context.

To discover the structure of the slice group G of permutations generated by the moves F, R, and D, we analyze the actions of slice moves on different sets of cubies, thus constructing homomorphisms from G to known groups.

Recall that if H is any group of permutations on a set X, and if Y is a subset of X such that for every $h \in H$ and $y \in Y$, we have $h(y) \in Y$, then Y is said to be **closed** under the action of H. When we have such a closed subset, a homomorphism is induced from H into the group of permutations of Y: simply map h in H to the permutation it induces on the set Y. In our case, G is a group of permutations on the set of cubies and facelets. We will find subsets of cubies that are closed under G, thus inducing homomorphisms of the type just described.

Returning to the cube, first notice that every corner cubie is left fixed, or unmoved, by each slice move, thus the slice group induces the identity permutation on the set of corner cubies and their facelets. This is convenient, since the colors on these corner cubies act as benchmarks in case we forget which face is front, back, etc.

Next, slice moves always move center cubies to center positions, hence the set of six center cubies is closed under the action of G. We denote by τ the homomorphism this induces from G into S_6, the group of permutations on six objects (in this case, the center cubies). The image $\tau(G)$ in S_6 ignores the action of G on all cubies except the centers. By examining the cube, one notes that each slice move has the same effect on the centers as some rigid motion of the entire cube. For example, the move F takes centers to the same positions that C_F does. Therefore, τ maps G into T, the group of rigid motions (rotations) of the cube (a subgroup of S_6). The map τ is actually onto T since T is generated by C_F, C_R, and C_D, and $\tau(F) = C_F$, $\tau(R) = C_R$, and $\tau(D) = C_D$. It is well known that T is isomorphic to S_4, and thus has 24 elements. (The isomorphism with S_4 is unimportant in what follows. The interested reader can construct the isomorphism by thinking of T as acting on the set of 4 diagonal line segments linking opposing corners of a rigid cube.)

The twelve edge cubies can be partitioned into three subsets of four each, E_F, E_R, and E_D, where each set contains the edge cubies on the center slice parallel to the F, R, and D faces, respectively. By manipulating the cube, one notices that the generators F, R, and D (and hence every slice move in G) take edge cubies in these sets to edge cubies in the same set. Hence, as above, we get homomorphisms ϕ_F, ϕ_R, and ϕ_D, each mapping G into the group of permutations of E_F, E_R, and E_D, respectively. Again, by examining the cube, one notices that F induces a 4-cycle on the elements of E_F, while R and D each induce the identity permutation on E_F. Hence $\phi_F(F)$ is a 4-cycle, $\phi_F(R) = \phi_F(D) = id$, and we see that $\phi_F(G)$ is generated by $\phi_F(F)$ and is isomorphic to Z_4. For any element g of G, $\phi_F(g)$ is the number of times mod 4 that F appears in any sequence of moves generating g, with F^{-1} counted as -1 (or 3 since $F^{-1} = F^3$). A similar analysis can be made of the homomorphisms ϕ_R and ϕ_D.

We thus have four homomorphisms of G: one onto T and three onto Z_4. By combining the four homomorphisms, one obtains a homomorphism from G into $T \times Z_4 \times Z_4 \times Z_4$ given by

$$g \to (\tau(g), \phi_D(g), \phi_R(g), \phi_F(g)). \tag{2}$$

The kernel of this homomorphism is clearly the intersection of the kernels of the four homomorphisms defining it, which is the set of elements in G leaving all center cubies and all elements of E_D, E_R, and E_F fixed. Such a permutation must leave the cube in its clean state, and is thus the identity. Therefore, the homomorphism is injective and G is isomorphic to its image, a subgroup of $T \times Z_4 \times Z_4 \times Z_4$.

Next, we notice that each generator of the slice group acts as the identity on two sets of edge cubies, as a 4-cycle on one set of edge cubies, and as a 4-cycle inside T, the group of motions of the centers. Therefore, overall, each generator acts as a product of two disjoint 4-cycles, yielding an even permutation. It follows that G is isomorphic to a subgroup of the *even* permutations in $T \times Z_4 \times Z_4 \times Z_4$. We will show that the image of G contains all such even permutations. (Note that parity in $T \times Z_4 \times Z_4 \times Z_4$ is inherited from S_{18}, the group of all permutations of the 18 center and edge cubies, of which $T \times Z_4 \times Z_4 \times Z_4$ is a subgroup.)

Using (2), G can be identified as a subgroup of $T \times Z_4 \times Z_4 \times Z_4$, and τ is the projection onto the first coordinate of the product. Hence, $\ker \tau$ can only contain elements of the form (id, a, b, c) in $T \times Z_4 \times Z_4 \times Z_4$, with $a+b+c$ even for the reasons discussed above. The following moves generate the indicated elements:

(i)	$RF^{-1}DF$	$(id,1,1,0)$
(ii)	$RDFD^{-1}$	$(id,0,1,1)$
(iii)	$FDR^{-1}D^{-1}FD^{-1}RD$	$(id,0,0,2)$.

Remember that the last three components of these elements refer to the cyclic permutation of the edge cubies in the D, R, and F planes, respectively. It can be shown that the three moves (i), (ii), (iii) generate every element of the form (id, a, b, c) with $a+b+c$ even. The procedure for doing this is similar to that of expressing the vector (a, b, c) as a linear combination of $(1,1,0)$, $(0,1,1)$, and $(0,0,2)$ as done in linear algebra, only here we work with integers and do arithmetic mod 4. Therefore, $\ker \tau$ is precisely the set of elements of this form, and thus has 32 elements. $\operatorname{Im} \tau$ has 24 elements since τ is onto T. Hence G has $24 \times 32 = 768$ elements.. But the set of even permutations in $T \times Z_4 \times Z_4 \times Z_4$ also has 768 elements, so G must equal this set.

Let us use this analysis of the structure of G to solve a cube that has been mixed up through slice moves. For example, consider the cube oriented as in FIGURE 4. (Should the reader wish to follow this discussion on an actual cube, the move $R^2FR^2F^{-1}DF^2R^{-1}$ will put a cube in the clean state into this position.) The first requirement is that we be able to recognize the position's expression as an element of $T \times Z_4 \times Z_4 \times Z_4$. One finds the first coordinate by writing down the permutation in S_6 of the center cubies. For the other three coordinates, one examines the edge cubies in E_D, E_R, and E_F, respectively, and counts the number of D, R, and F moves that were required to put the cubies in their current position. The result of this computation for FIGURE 4 is $([(F, L, U)(B, R, D)], 1, 3, 2)$.

To solve the cube, first make slice moves mimicking the rigid motions of the cube to bring the center cubies to their home positions. At most two 90° or 180° turns will always be sufficient to do this. In our example, the move F^{-1} puts the right and left centers in their home position (see FIGURE 5(a)). The move R will then return the remaining centers to their home position, as in FIGURE 5(b). The cube is now in the position $(id,1,0,1)$. The purpose of this maneuver is to put the cube's positional representation in $\ker \tau$. The advantage here is that $\ker \tau$, having only 32

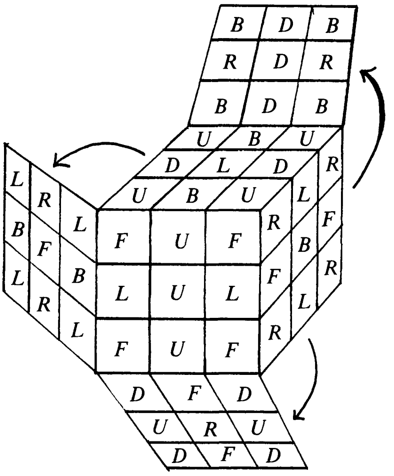


FIGURE 4. The cube in position $([(F, L, U)(B, R, D)], 1, 3, 2)$.

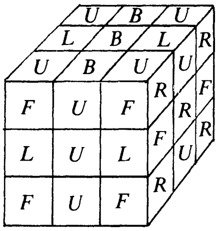


FIGURE 5 (a)

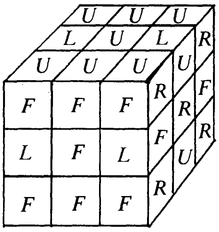


FIGURE 5 (b)

elements, is a much simpler group to work with than G. To complete the solution, we now use the generators for $\ker \tau$ given by equations (i), (ii), and (iii) to produce the inverse element of the current cube position so that the cube can be restored to the clean state. The inverse of $(id, 1, 0, 1)$ is $(id, 3, 0, 3)$, which can be obtained by move (i) 3 times followed by (ii) and then (iii). Denoting this sequence of moves by M we write $M = \text{(i)}^3\text{(ii)(iii)}$. Hence, starting from FIGURE 4, we can make the following sequence of moves to restore the cube:

$$F^{-1}RM = F^{-1}R(RF^{-1}DF)(RF^{-1}DF)(RF^{-1}DF)(RDFD^{-1})(FDR^{-1}D^{-1}FD^{-1}RD).$$

This is definitely not the shortest possible solution, but it works. One method of shortening the answer is to use the fact that $\text{(i)}^3 = \text{(i)}^{-1} = F^{-1}D^{-1}FR^{-1}$; this yields the solution

$$F^{-1}R(F^{-1}D^{-1}FR^{-1})(RDFD^{-1})(FDR^{-1}D^{-1}FD^{-1}RD). \tag{3}$$

Note that $R^{-1}R$ appears in this expression and can be cancelled out; however, very little else can be done to shorten this process without a whole new approach to the problem.

The method used here is typical of cube-solving strategies. One finds a sequence of nested subgroups of the group of allowable positions; in our case,

$$G \supset \ker \tau \supset 1.$$

Then, for each position in the sequence, one finds moves for a representative in each coset of the smaller subgroup in the larger. One calculates the cube's position, then performs the moves corresponding to the known coset representative of the position's inverse to reduce the cube's state to one in the next subgroup down the sequence. In our example, the first such reduction was made by performing the move $F^{-1}R$, a representative of the coset over $\ker \tau$ of the inverse of the cube's original position. All the published cube solutions (see references) use this strategy, although different solutions use different subgroup series. For example, Taylor and Rylands [8] solve a slice group problem by utilizing the series

$$G \supset \ker \Theta \supset 1$$

where Θ is the projection of G onto $Z_4 \times Z_4 \times Z_4$. This solution first puts all edge cubies in their home positions to obtain a "spot pattern." Then generators of $\ker \Theta$ are given to solve that pattern.

In the case of the slice group, more efficient solutions could be found by an exhaustive search (a good exercise for those with a background in computer programming) since there are only 768 possible positions. This, however, is an unrealistic approach for solving a generally scrambled cube, since there are about 10^{18} possible different cube positions [1]. The series of subgroups approach seems to be the only one to generate practical algorithms for solving larger cube problems. For further information on this approach, and some interesting subgroup series used for larger cube problems, we refer the reader to Frey and Singmaster [4].

The oriented slice group

Some of the cubes on the market today have symbols or designs pasted on the facelets instead of solid colors. Some even have pictures, and one must unscramble all six pictures to solve the cube. In solving a scrambled cube of this type, one might notice that some of the center facelets have been rotated with respect to the other designs on the faces on which they lie. Move sequences are known [8] that yield rotations of individual center cubies. We will study the rotations which are possible using slice moves.

To analyze the effect of slice moves on center cubies, we shall redefine the identity transformation of the cube to include fixed orientation of the centers. This, of course, redefines the term "clean state" for the cube as well (see FIGURE 6). We then get a larger group G^* of possible cube positions, the **oriented slice group**. If one does not have a cube with designs or pictures, one can experiment with G^* by pasting pieces of mailing labels over the center and upper right hand corner of each face, and then drawing arrows on the labels, both in the same direction, to signify orientation. The arrows on the corners provide benchmarks of orientation since the corner facelets remain fixed under slice moves. In what follows, we will assume that the arrows have been drawn on the cube in the directions indicated in FIGURE 6.

To describe the position of the cube after some slice moves, one must first describe the positions of all the center facelets and edge cubies with some element of G, then describe the orientation of each of the six center facelets. We represent the orientation of a given center as an element of Z_4 in the following way. One finds the center facelet in question on the cube, then compares the direction of the arrow pasted on it with the direction of the arrow on the corner cubie of the face on which the center facelet currently resides. The element in Z_4 counts the number of 90° clockwise rotations the center has made from the standard direction noted on the corner cubie. Using one copy of Z_4 for each center facelet, the cube's position can be represented as an element of the set $G \times Z_4 \times Z_4 \times Z_4 \times Z_4 \times Z_4 \times Z_4 = G \times (Z_4)^6$, where the last six coordinates represent the orientation of the U, D, F, B, R, and L colored centers, respectively. The representations of the three generators of the oriented slice group, using this notation are given in TABLE 1.

It is important to note that we have yet to define a group operation on the set $G \times (Z_4)^6$. We will write the representation of a slice move in $G \times (Z_4)^6$ as $(\rho,(A,B,C),(a,b,c,d,e,f))$, where ρ is in T, a subgroup of S_6, (A,B,C) is in $Z_4 \times Z_4 \times Z_4$, and (a,b,c,d,e,f) is in $(Z_4)^6$. Note that $(\rho,(A,B,C))$ comprises the G component of the slice move. Then given two slice moves,

$$g = (\rho,(A,B,C),(a,b,c,d,e,f))$$

and

$$h = (\sigma,(X,Y,Z),(u,v,w,x,y,z)),$$

we define their composition (the group operation in $G \times (Z_4)^6$) to be

$$hg = \big(\sigma\rho,((A,B,C) + (X,Y,Z)),((a,b,c,d,e,f) + \rho^{-1}(u,v,w,x,y,z))\big)$$

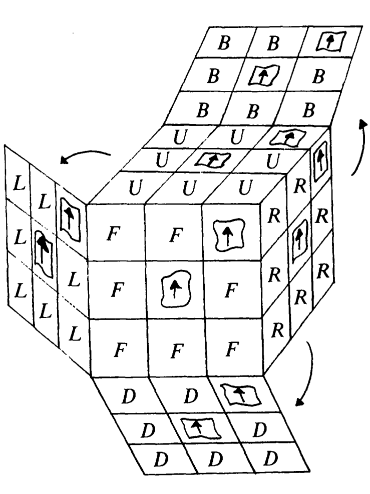


FIGURE 6. An oriented cube in the clean state.

	$\longleftarrow$ G $\longrightarrow$									
Move	Permutation of Center Facelets	Cycles of Edge Planes			Rotations of Center Facelets					
		E_D	E_R	E_F	U	D	F	B	R	L
D	(F,R,B,L)	1	0	0	0	0	0	2	2	0
R	(U,B,D,F)	0	1	0	0	0	0	0	0	0
F	(U,R,D,L)	0	0	1	1	1	0	0	1	1

TABLE 1

where additions are componentwise, mod 4. The purpose of applying ρ^{-1} before adding the last six components is to make sure the orientation changes are applied to the proper center facelets. The representation of a position in $G \times (Z_4)^6$ assumes that moves were made starting from the clean state. However, composing moves requires the second move to start from where the first one left off, so the center cubies are not starting from their home positions. Since ρ indicates the centers' positions after the move g, applying ρ^{-1} before adding permutes the orientation changes being made by h so that they are applied to the proper facelets' coordinates. The resulting group structure is called a **semidirect product** of groups. We refer the reader to the literature [5] for more information on this topic.

It can be shown that G^* can be represented as a subgroup of the set $G \times (Z_4)^6$ with the group operation described above. In fact, G^* is the subgroup generated by the moves R, D, and F whose representations are given in TABLE 1.

Let π: $G^* \to G$ be the homomorphism which projects G^* onto G, that is, π focuses on cubie position and ignores center facelet orientations. Then $\ker \pi$ is the set of slice moves that leave all cubies in their home position, but may rotate center facelets. Note that in $\ker \pi$, the composition of moves corresponds to ordinary addition in $(Z_4)^6$ since the G component is the identity element. This greatly simplifies matters.

The following two moves, α and β, are in $\ker \pi$. Their last six coordinates in our representation are given.

$$\begin{aligned} \alpha &= RF^{-1}DFR^{-1}F^{-1}D^{-1}F && (0,0,2,2,0,0) \\ \beta &= RF^{-1}DFD^{-1}FR^{-1}F^{-1} && (3,1,0,0,3,1) \end{aligned} \tag{4}$$

One can construct similar moves by conjugating these with rigid motions of the cube (think of holding the cube with different faces forward and up and using the same twists) to obtain

$$\begin{aligned} \alpha^* &= C_U\alpha C_U^{-1} && (0,0,0,0,2,2), \\ \beta^* &= C_U\beta C_U^{-1} && (3,1,1,3,0,0), \\ \alpha' &= C_L\alpha C_L^{-1} && (2,2,0,0,0,0), \\ \beta' &= C_L\beta C_L^{-1} && (0,0,1,3,3,1). \end{aligned} \tag{5}$$

The six moves in (4) and (5) generate $\ker \pi$. We outline a proof of this, but leave the details to the reader.

First, show that for any element of G^*, the number of odd numbers in the last six coordinates is either 0 or 4. (This is true of the generators. Show that this property is preserved under composition in G^*. You will need other properties of the generators and the fact that opposite centers always remain opposite to each other.) Next, prove that within $\ker \pi$, the sum over coordinates corresponding to orientations of opposite faces is 0 mod 4. (This is only true in $\ker \pi$ and is more difficult to prove. Consider the generators R, D, and F without the final rigid motion given in their definition, and remember that in $\ker \pi$, all centers end up in their home position.) Then count and find that there are 32 six-tuples having these two properties, and check that all 32 can be generated using α, α^*, β, β^*, α' and β'.

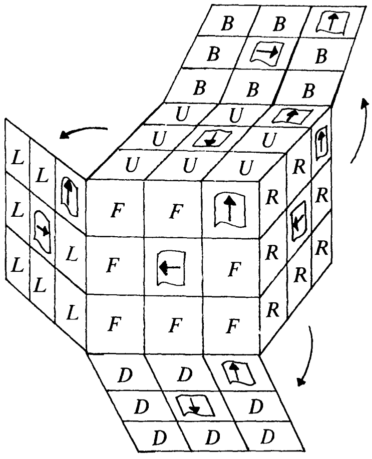


FIGURE 7. **The cube in position (*id*, (0,0,0), (2,2,3,1,3,1)).**

From the information above, we find that

$$|G^*| = |\ker\pi|\cdot|\mathrm{Im}\,\pi| = |\ker\pi|\cdot|G| = 32\times 768 = 24576.$$

This is the total number of distinct possible positions of an oriented Rubik's cube scrambled by slice moves.

Although we do not have a simple description of which elements of $G\times(Z_4)^6$ are in G^*, we do have an algorithm for solving an oriented cube problem by using a series of subgroups. If we let $\tau'\colon G^*\to T$ be the composition $G^*\xrightarrow{\pi} G\xrightarrow{\tau} T$, then the series

$$G^*\supset\ker\tau'\supset\ker\pi\supset 1$$

provides our solution. The passage from G^* to $\ker\pi$ through $\ker\tau'$ is precisely the method of solution given previously, which ignores orientation; that is, first get the centers in their home positions (pass to $\ker\tau'$), then move the edge cubies to their home positions leaving centers fixed (pass to $\ker\pi$) using the moves (i), (ii), and (iii). Now express the inverse of the current position (in $\ker\pi$) as a product of the generators α, α^*, α', β, β^*, and β' and perform these moves to orient the center facelets, thus solving the cube. For example, suppose an oriented cube is put into the position of FIGURE 4 by making the move $R^2FR^2F^{-1}DF^2R^{-1}$ on a cube in the clean state. One solves this cube problem by first ignoring the orientation arrows and following the solution technique described for an unoriented cube. According to our previous calculations, this requires us to make the move given in formula (3). This will leave the cube in the position $(id,(0,0,0),(2,2,3,1,3,1))$, as shown in FIGURE 7. Since this position is in $\ker\pi$, its inverse is $(id,(0,0,0),(2,2,1,3,1,3))$, and this inverse can be expressed as a product of the known generators of $\ker\pi$. The composition $\alpha^*\alpha'\beta'$ is one of many that work. Making this move will restore the cube to the clean state.

One of us (Ranan Banerji) is supported part-time by the National Science Foundation under Grant MCS-821 7964 to Saint Joseph's University.

References

[1] C. Bandelow, Inside Rubik's Cube and Beyond, Birkhauser, Boston, Mass., 1982.
[2] E. Berlekamp, J. H. Conway, and R. Guy, Winning Ways, Academic Press, 1982, pp. 760–768, 808–809.
[3] T. Davis, Teaching mathematics with Rubik's cube, Two-Year College Math J., 13 (1982) 178–185.
[4] A. H. Frey and D. Singmaster, Handbook of Cubik Math, Enslow Publishers, Hillside, N.J., 1982.
[5] S. MacLane and G. Birkhoff, Algebra, Macmillan Co., New York, 1967, p. 461.
[6] J. G. Nourse, The Simple Solution of Rubik's Cube, Bantam Books, New York, 1981.
[7] D. Singmaster, Notes on Rubik's Magic Cube, Enslow Publishers, Hillside, N.J., 1980.
[8] D. Taylor and L. Rylands, Cube Games, Holt, Rinehart, and Winston, New York, 1981.

A Bug's Shortest Path on a Cube

WEIXUAN LI
Changsha Railway Institute
Changsha, Hunan, China

EDWARD T. H. WANG
Wilfrid Laurier University
Waterloo, Ontario, Canada N2L 3C5

In any introductory course on combinatorics, the problem of finding the number of shortest paths from one corner of a city to the opposite corner is almost always introduced. As is well known, if the city is viewed as a grid of $m+1$ streets going N–S and $n+1$ streets going E–W (i.e., m blocks by n blocks), then the number of shortest paths is

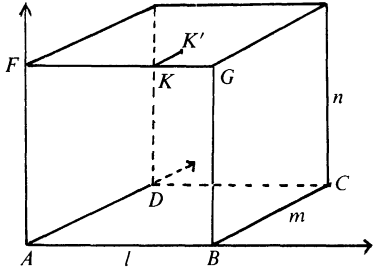
$$f(m,n)=\binom{m+n}{m}=\frac{(m+n)!}{m!n!}.$$

This can be readily obtained by identifying each shortest path either as a selection of m objects from $m+n$ objects or as a permutation of $m+n$ objects consisting of exactly m of the symbols E (east) and n of the symbols N (north).

Now one naturally wonders about the 3-dimensional analogue of this problem. Imagine for example, that a bug situated at one corner of an $n\times n\times n$ Rubik's cube wants to reach the opposite corner of the opposite face in such a way that its path consists of line segments of integer length each of which is parallel to one of the axes. What is the number of such shortest paths? If the bug can "chew" its way through the interior of the cube, then the same argument used in the 2-dimensional version immediately gives the answer $(3n)!/(n!)^3$. But if the bug can only crawl on the surface of the cube and because the surface is slippery, must crawl along the edges or the grooves between the small cubes, to determine the number, $f(n)$, of such shortest paths turns out to be a much harder problem than the 2-dimensional version. Though it is obvious that $f(1)=6$, to compute the value of $f(2)$ by brute force might be quite a challenging job. (Any one with $2\times 2\times 2$ Rubik's cube is encouraged to give it a try.) To determine the value of $f(3)$, with or without the use of an ordinary Rubik's cube, seems to be a formidable task. For the interest of the readers, we will not reveal these two values until the end of the article. To answer this question, we consider the more general problem in which the cube is replaced by a rectangular box.

Let Ω be the $l\times m\times n$ rectangular prism in space with vertices at $A=(0,0,0)$, $B=(l,0,0)$, $C=(l,m,0)$, $D=(0,m,0)$, $E=(0,m,n)$, $F=(0,0,n)$, $G=(l,0,n)$ and $H=(l,m,n)$. See FIGURE 1, where l, m, and n are positive integers. A bug wants to go from A to H on the surface of Ω in such a way that its path consists of line segments of integer lengths each of which is parallel to one of the axes. What is the number $f(l,m,n)$ of different shortest paths? (It is clear that the length of such a path is $l+m+n$.) This is question (1) asked in [1].

FIGURE 1



THEOREM.

$$f(l,m,n)=2\left\{\binom{l+m+n}{l}+\binom{l+m+n}{m}+\binom{l+m+n}{n}-\binom{l+m}{l}-\binom{m+n}{m}-\binom{n+l}{n}\right\}.$$

First Proof. We partition the set $\mathscr{F}$ of all shortest paths T into families $\mathscr{F}_i$ as follows: define $T\in\mathscr{F}_i$ if T passes through the interior of exactly i faces. It is evident that $0\leqslant i\leqslant 2$. Let $c_i=|\mathscr{F}_i|$. Then clearly $c_0=6$. If $T\in\mathscr{F}_1$ and goes through the interior of the face $ABCD$, then it clearly must go from A to C and then along the edge CH. The number of such paths is $\binom{l+m}{l}-2$ since we must exclude the two paths $A\to B\to C\to H$ and $A\to D\to C\to H$. Taking each one of the three faces containing A into account, we see that

$$c_1=2\left\{\binom{l+m}{l}+\binom{m+n}{m}+\binom{n+l}{n}-6\right\}.$$

Consider now $T\in\mathscr{F}_2$. Note that $c_2=c(l;m,n)+c(m;n,l)+c(n;l,m)$ where $c(l;m,n)$ is the number of paths of $\mathscr{F}_2$ passing through the interior of two faces of dimension $l\times m$ and $l\times n$ respectively. Suppose T goes through the interior of both faces $ABGF$ and $EFGH$. Let $K=(i,0,n)$ be the point on FG where the bug exits the face $ABGF$, $1\leqslant i\leqslant l-1$. Then the second "part" of T must start from $K'=(i,1,n)$. The number of shortest partial paths from A to K is $\binom{n+i}{n}-1$ (excluding $A\to F\to K$) and the number of shortest partial paths from K' to H is $\binom{l+m-i-1}{m-1}$. Therefore,

$$\begin{aligned}c(l;m,n)&=2\sum_{i=1}^{l-1}\left\{\binom{n+i}{n}-1\right\}\binom{l+m-i-1}{m-1}\\&=2\left\{\sum_{i=1}^{l-1}\binom{n+i}{n}\binom{l+m-i-1}{m-1}-\sum_{i=1}^{l-1}\binom{l+m-i-1}{m-1}\right\}.\end{aligned}\tag{1}$$

Now, by a known formula ([**2**], p. 64, formula (1)),

$$\sum_{k=0}^{a}\binom{a-k}{r}\binom{b+k}{s}=\binom{a+b+1}{r+s+1}.$$

Hence,

$$\begin{aligned}\sum_{i=1}^{l-1}\binom{n+i}{n}\binom{l+m-i-1}{m-1}&=\sum_{i=0}^{l}\binom{n+i}{n}\binom{l+m-i-1}{m-1}-\binom{l+m-1}{m-1}-\binom{n+l}{n}\\&=\sum_{i=0}^{l+m-1}\binom{l+m-1-i}{m-1}\binom{n+i}{n}-\binom{l+m-1}{l}-\binom{n+l}{l}\\&=\binom{l+m+n}{m+n}-\binom{l+m-1}{l}-\binom{n+l}{l}.\end{aligned}\tag{2}$$

(In the second equality above, we used the fact that $\binom{l+m-1-i}{m-1}=0$ if $i>l$.) Also, from ([**2**], p. 64, formula (7)),

$$\begin{aligned}\sum_{i=1}^{l-1}\binom{l+m-i-1}{m-1}&=\sum_{i=1}^{l-1}\binom{l+m-i-1}{l-i}=\sum_{j=1}^{l-1}\binom{m+j-1}{j}\\&=\sum_{j=0}^{l-1}\binom{m+j-1}{j}-1=\binom{m+l-1}{l-1}-1.\end{aligned}\tag{3}$$

From (1), (2), and (3), and using the identity

$$\binom{l+m-1}{l}+\binom{l+m-1}{l-1}=\binom{l+m}{l}$$

we obtain

$$c(l;m,n)=2\left\{\binom{l+m+n}{l}-\binom{l+m}{l}-\binom{n+l}{l}+1\right\}.$$

Adding up similar expressions for $c(m;n,l)$ and $c(n;l,m)$, we get

$$c_2=2\left\{\binom{l+m+n}{l}+\binom{m+n+l}{m}+\binom{n+l+m}{n}-2\binom{l+m}{l}-2\binom{m+n}{m}-2\binom{n+l}{n}+3\right\}.$$

Therefore,

$$\begin{aligned}f(l,m,n)&=c_0+c_1+c_2\\&=2\left\{\binom{l+m+n}{l}+\binom{l+m+n}{m}+\binom{l+m+n}{n}-\binom{l+m}{l}-\binom{m+n}{m}-\binom{n+l}{n}\right\}.\end{aligned}$$

Now we give a second proof which is more "combinatorial" in the sense that it does not involve those tedious computations in the above proof.

Second Proof. Using the same notations as in the first proof, the number of shortest paths through the faces $ABGF$ and $EFGH$ is $\binom{l+m+n}{l}$ since we could view the two faces as constituting a planar grid of dimension $l\times(m+n)$. Summing over six pair of faces (i.e., ($ABGF$, $EFGH$), ($ABGF$, $BCHG$), ($ABCD$, $BCHG$), ($ABCD$, $CHED$), ($ADEF$, $EFGH$), ($ADEF$, $CHED$)), we have

$$2\left\{\binom{l+m+n}{l}+\binom{l+m+n}{m}+\binom{l+m+n}{n}\right\}$$

shortest paths, some of which are counted more than once. Indeed, paths in $\mathscr{F}_0$ have been counted three times and those in $\mathscr{F}_1$, twice; e.g., the path $A\to F\to G\to H$ is counted by the three pairs of faces ($ABGF$, $EFGH$), ($ABGF$, $BCHG$), and ($ADEF$, $EFGH$); and a path via the interior of the face $ABGF$ followed by the edge GH is counted by the two pairs of faces ($ABGF$, $EFGH$), and ($ABGF$, $BCHG$). Thus we have

$$3c_0+2c_1+c_2=2\left\{\binom{l+m+n}{l}+\binom{l+m+n}{m}+\binom{l+m+n}{n}\right\}. \tag{4}$$

The number of shortest paths through the faces $ABGF$ and along the edge GH is $\binom{n+l}{n}$. Summing over the six faces we get $2\left\{\binom{l+m}{l}+\binom{m+n}{m}+\binom{n+l}{n}\right\}$ shortest paths, some of which are counted more than once. Indeed, each path in $\mathscr{F}_0$ has been counted twice. Thus we have

$$2c_0+c_1=2\left\{\binom{l+m}{l}+\binom{m+n}{m}+\binom{n+l}{n}\right\}. \tag{5}$$

Our result now follows by subtracting (5) from (4).

In particular, if our prism is a cube (i.e., $l=m=n$), then we get the answer to the question that was raised in the beginning of this note:

$$f(n)=f(n,n,n)=6\left\{\binom{3n}{n}-\binom{2n}{n}\right\}.$$

Thus, for example, $f(1)=6$, $f(2)=54$ and $f(3)=384$.

This joint article was written when the first author was visiting Wilfrid Laurier University in July, 1984. The hospitality of this Institute is greatly appreciated.

Research supported by the Natural Sciences and Engineering Research Council of Canada under grant A9121.

References

[1] J. F. Lucas, Paths and Pascal numbers, TYCMJ, 14 (1983) 329–341.

[2] Alan Tucker, Applied Combinatorics, John Wiley and Sons, Inc., 1980.

Missing More Serves May Win More Points

LEONARD GILLMAN
The University of Texas at Austin
Austin, TX 78712

This note embellishes David Gale's engaging piece about serving strategies in tennis [**1**]. It includes a review of the needed concepts, so familiarity with [**1**] is not required.

The server in tennis ordinarily cultivates an arsenal of two serves: r, rapid and risky; and s, slow and safe. The customary strategy is to employ one's risky serve first, then (when necessary) one's safe serve. Why is that good? If the proportion of bad serves (of either kind) should go up, the server frets. Why fret?

Nonmathematicians fail these questions, because the answers require formulating and combining conditional probabilities, concepts that defeat them: McEnroe is happy to have got in his first serve "when he needed to"; coaches count the number of aces; reporters quote isolated statistics; players explain that they use a safe second serve to prevent a double fault; announcers report that Connors is missing his first serve and "it's hurting him."

Let w_x denote the probability that the server wins the point on serving a particular serve $x(=r$ or $s)$. To win, two things must happen: one's serve must be good, and one must win the ensuing play. Let p_x denote the probability that the serve is good, and let π_x be the probability that, if it is good, the server goes on to win the point. Then

$$w_x = p_x \pi_x. \tag{1}$$

In this simple model we are assuming that p_x and π_x are constant; e.g., we ignore the possible element of surprise in suddenly switching from second serve safe to second serve risky. Note that p_x is controlled by the server alone, while π_x depends on both players. We impose the realistic condition that both probabilities lie strictly between 0 and 1.

In our notation, w_r is the probability that the server wins the point with a rapid serve. Since the server is favored to win (in a decently matched game), we will have $w_r > 0.50$. (No?) Likewise, $w_s > 0.50$. (Yes?)

Let W_{xy} denote the probability that the server wins the point using strategy xy (first serve x, second serve y). Since the second serve exists only if the first serve was bad,

$$W_{xy} = w_x + (1 - p_x) w_y. \tag{2}$$

We apply (2) to the four strategies *rr*, *rs*, *sr*, *ss*. Of course we assume

$$p_r < p_s \tag{3}$$

(as prejudged by the terminology), and

$$\pi_r > \pi_s, \tag{4}$$

since the rapid serve is harder for the receiver to handle than the slow one. (If the inequalities ran in the same direction there would be nothing to talk about.) Note that, as far as one can tell from (1), (3), and (4), both $w_s > w_r$ and $w_s < w_r$ are possible; it turns out, moreover, that among tennis players at all levels, both are reasonable.

Are there any obvious choices among the strategies—is a particular one always best; is one of the four never best; is one of them always worst?

It is clear from (2) (or from thinking about the situation without the formula) that the second serve should be chosen so as to maximize w_y. Thus the relation

$$w_s > w_r$$

is equivalent to

$$W_{rs} > W_{rr} \text{ and to } W_{ss} > W_{sr}. \tag{5}$$

Comparing first serves, we find from (2) that the relation

$$w_r > w_s[1-(p_s-p_r)]$$

is equivalent to (6)

$$W_{rs} > W_{ss}.$$

Notice that the factor in brackets is < 1.

How do *rs* and *sr* compare? From (2), (1), and (4), we see that

$$W_{rs} - W_{sr} = p_r p_s(\pi_r - \pi_s) > 0. \tag{7}$$

Thus, *sr* is never the best strategy, and so should never be used. We can now classify all possibilities. If (5) and (6) both hold then *rs* is the best strategy; if (5) fails, then *rr* is best (or, in the special case $w_r = w_s$, tied with *rs*); and if (6) fails, then (5) holds and *ss* is best [**1**].

The model can be useful in other settings. The numbers π_x and W_{xy} might represent expected values, measuring economic gain or artistic enjoyment or other rewards. One application is to a common dating pattern in which the boys are the ones who phone the girls. Here r and s represent different girls, and p_x is the probability that x will accept a date from a particular boy. (This time, p_x is not controlled by one "player" alone.) If for simplicity we rule out strategies *rr* and *ss*, and if we assume (4), then (7) provides the complete solution. The result extends to any number of choices: e.g., if $\pi_r > \pi_s > \pi_t$, then the optimal phoning strategy is *rst*.

The traditional tennis strategy *rs* turns out to be never worse than second best—and the untraditional strategy *sr* is never better than second worst. To see this, consider the 4! possible orderings; for convenience, call the strategies 1, 2, 3, 4, in alphabetical order. The result (7) cuts the 24 possibilities down to 12, and the equivalence (5) reduces this number to 6, in all of which the assertion is true: 1234, 2143, 2413, 2431, (4213), and 4231. Further juggling with inequalities eliminates 4213. The other five orderings are all possible; in fact, Borg and McEnroe exhibited four of them at Wimbledon.

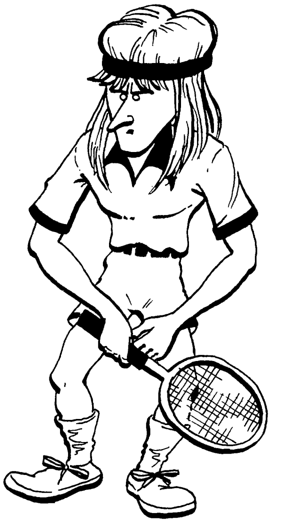
The last remark is based on data I recorded from the televised men's finals at Wimbledon and New York in 1980 and 1981. (I imagine that's the last time I watched any television.) I sat drawing tick marks, like a zombie. The ball is often hard to see, the call of a play may be inaudible, and the zombie can easily get mixed up. So the data are surely imperfect; but I think they give the main flavor. The numbers p_x and π_x that appear in TABLE 1 are the observed frequency ratios, and the other entries are computed from them. (I ignore such effects as players' tiring and assume the probabilities to be constant throughout the match.) Given the uncertainties surrounding the data, I saw no point in listing three decimal places to break two-place ties. (I tried for women's data too, but what I got were always too fragmentary to be of use, because the sexist network kept interrupting the women's finals to show the men's *semi*finals.)

Both players played strategy *rs* (though admittedly some of McEnroe's second serves were pretty fast). Apparently, Borg would have done slightly better at Wimbledon with *ss* in 1980 (when he won) and with *rr* in 1981 (when he lost). (Carl Morris has called my attention to the article [**2**], which tests the statistical significance of such comparisons.)

Some apparent paradoxes baffle people untrained in mathematical thinking and may even perplex some mathematics students.

1. In the matches reported here, *the values of* w_r *are all under* 0.50 (and go as low as 0.34).

2. It is a fact of tennis that p_r and π_r are not independent variables. Suppose you decide to smash your risky serve a little harder, or aim closer to the corner. You will surely decrease your probability p_r. But the serves that do go in will be more difficult to return, and you have increased your probability π_r. If as a result the product $p_r\pi_r$ $(= w_r)$ increases, you have gained. McEnroe's 1980 data are consistent with this (though additional factors must have helped account for the huge increase in π_r). *A riskier first serve may increase the probability of winning the point on your first serve.* Likewise, reducing p_s by taking more chances might increase π_s so as to increase the



	Borg				McEnroe			
	1980		1981		1980		1981	
	W	NY	W	NY	W	NY	W	NY
p_r	.62	.45	.61	.54	.63	.60	.59	.54
π_r	.69	.83	.76	.63	.69	.80	.77	.69
w_r	.43	.38	.46	.34	.43	.49	.45	.37
p_s	.89	.85	.88	.86	.96	.89	.96	.85
π_s	.68	.72	.50	.52	.66	.67	.52	.55
w_s	.61	.62	.44	.45	.63	.59	.50	.47
W_{rr}	.59	.58	.64	.50	.59	.68	.63	.54
W_{rs}	.66	.71	.63	.55	.67	.72	.66	.59
W_{sr}	.65	.67	.50	.50	.65	.65	.52	.53
W_{ss}	.67	.71	.49	.51	.66	.66	.52	.54

TABLE 1. **Armchair data on two tennis matches between Bjorn Borg and John McEnroe. W: Wimbledon; NY: New York (U.S. Open).**

product $p_s\pi_s$ $(=w_s)$. Borg's data for both years are examples. *A riskier second serve may increase the probability of winning the point on your second serve.* Hence *a riskier second serve may increase the probability of winning the point.* (Cf. (5).)

3. It is a fact of mathematics that the probability of winning the point—*taking both serves into account*—is expressed by formula (2). For definiteness, look at W_{rs}:

$$W_{rs} = w_r + (1 - p_r)\,w_s. \tag{8}$$

Behold the factor $1 - p_r$. When p_r goes down, it goes up, and the second term on the right goes up: a riskier first serve increases the probability of winning the point on the second serve. If the first term w_r also increases (as explained in the preceding paragraph), then W_{rs} goes up. Hence *a riskier first serve may increase the probability of winning the point.* This can happen even if w_r goes down, as in Borg's volatile 1980.

References

[1] David Gale, Optimal strategy for serving in tennis, this MAGAZINE, 44 (1971) 197–199.

[2] S. L. George, Optimal strategy in tennis: a simple probabilistic model, J. Royal Statist. Soc. Ser. C, 22 (1973) 97–104.

An Example of an Error-Correcting Code

MARK RABENSTEIN, *student*
McKernan Junior High School
Edmonton, Alberta, Canada T6G 0K1

I am a student in grade 8. Recently, I went to an enrichment program run by Andy Liu of the University of Alberta. The topic we studied is called "error-correcting codes."

The problem goes like this. A secret agent has to send a message back to headquarters. He uses a transmitter which sends a string of 0's and 1's. Unfortunately, from time to time, a 1 gets changed into a 0 while the message is on its way, or vice versa. So he has to send some extra digits to make sure there is no misunderstanding. Fortunately, no more than k digits in the expanded or encoded message are changed at one time.

We studied many interesting schemes for encoding a message. The first one is really simple. Just repeat the message $2k+1$ times, and the copy that appears at least $k+1$ times is the correct one. However, this requires lots of digits, and this is not good for a secret agent.

When I thought things over, I did not see why it was necessary to repeat the message $2k+1$ times. If there are no more than k mistakes and the message is repeated $k+1$ times, one of the copies must be correct! The only problem is: How can we tell which is the correct one?

Well, to tell whether a copy has one mistake (or any odd number of mistakes), there is a simple way. Add a 1 to the original message if it has an odd number of 1's, and add a 0 otherwise. This way, the number of 1's in the encoded message is always even.

This extra digit is called a "parity-check digit," parity meaning odd or even. If an odd number of mistakes is made in the encoded message, then the number of 1's in it will be odd and not even as it is supposed to be, and in this case we can tell something is wrong. Of course, the method fails for an even number of mistakes.

Let us go back two paragraphs and see how parity-check digits can be of help there. As I said, repeat the message $k+1$ times. Now add a parity-check digit to each of the last k copies. I will show that this works.

NEW IMPROVED METHOD

```
1. ENCODING:       1100 11000 11000                    Ratio of
                                                       message lengths
                                                       = 14:20
2. TRANSMISSION:   1000 11001 11000
                                                       How does this
3. DECODING:       (a)  1000 | 11001 | 11000           ratio improve
                                 ×                     as K becomes
                   (b)  1000 | 11000                   large?
                                 ✓
                   (c)  11000
```

Check each of the last k copies of the received message to see if anything is wrong. Suppose we find a copy with an odd number of mistakes. Well, throw it out! With each such copy goes at least one mistake. In the worst case, everything is gone except the first copy. It must be correct because all the mistakes have been thrown out.

Suppose we are left with $l+1$ copies (the last l copies having parity-check digits). We know that there are at most l mistakes, and they come in pairs in the last l copies.

What does this mean? This means that at least half of these l copies contain no mistakes. We should be able to tell what the correct message is, unless there is a two-way tie. In that case, we still have the first copy, which must be correct because all the mistakes have been used up.

So my scheme does work. Of course, it still needs lots of digits, unlike some of the really clever schemes I learned in the enrichment program.

Remarks by A. Liu: The code presented in this note is apparently new. The reader may supply a more formal proof. The "apology" in the last paragraph is really not necessary in that the ease of encoding and decoding for this scheme offsets its lack of sophistication. Its rate of information is asymptotically nearly twice that of the repetition codes [**1**].

Codes with higher rates (the "really clever schemes") were known early on in the history of error-correcting codes (see for example [**2**], [**3**], and [**5**]). A recent publication [**6**] gives an interesting historical account and shows the interrelationship of error-correcting codes with other areas of mathematics. A definitive treatise [**4**] details the state of the art as well as listing over one thousand references.

References

[1] F. L. Alt, A Bell Telephone Laboratories' computing machine (I), Math Tables and Aids to Comput. (Math. Comput.), 3 (1948/49) 1–13.

[2] M. J. E. Golay, Notes on digital coding, Proc. IRE (IEEE), 37 (1949) 657.

[3] R. W. Hamming, Error detecting and correcting codes, Bell System Tech. J., 29 (1950) 147–160.

[4] F. J. MacWilliams and N. J. A. Sloane, The Theory of Error-Correcting Codes, North Holland Pub. Co., Amsterdam, 1977.

[5] C. E. Shannon, A mathematical theory of communication, Bell System Tech. J., 27 (1948) 379–423, 623–656.

[6] T. M. Thompson, From Error-Correcting Codes through Sphere Packings to Simple Groups, Mathematical Association of America, Washington, DC, 1983.

The Mathematics of Double Entry Bookkeeping

David P. Ellerman
Boston College
Chestnut Hill, MA 02167

For several centuries, the double entry system has been the bookkeeping system used by most sizable business enterprises throughout the world. It is little known in mathematics and it is virtually unknown in accounting that the double entry system is based on a well-known mathematical construction of undergraduate algebra, the group of differences, in which the integers are represented as equivalence classes of ordered pairs of natural numbers.

The T-accounts of double entry bookkeeping are precisely the ordered pairs of the group of differences construction. With the one exception of a paragraph by D. E. Littlewood (see below), the author has not found a single mathematics book, elementary or advanced, popular or esoteric, which notes that this construction is the theoretical basis of a mathematical technique applied, day in and day out, in the mundane world of business for over five centuries. And even though the construction is standard fare in an undergraduate modern algebra course, the connection with double entry bookkeeping is totally absent in the accounting literature (see References; [4] gives a review of the literature in mathematical accounting).

Check each of the last k copies of the received message to see if anything is wrong. Suppose we find a copy with an odd number of mistakes. Well, throw it out! With each such copy goes at least one mistake. In the worst case, everything is gone except the first copy. It must be correct because all the mistakes have been thrown out.

Suppose we are left with $l+1$ copies (the last l copies having parity-check digits). We know that there are at most l mistakes, and they come in pairs in the last l copies.

What does this mean? This means that at least half of these l copies contain no mistakes. We should be able to tell what the correct message is, unless there is a two-way tie. In that case, we still have the first copy, which must be correct because all the mistakes have been used up.

So my scheme does work. Of course, it still needs lots of digits, unlike some of the really clever schemes I learned in the enrichment program.

Remarks by A. Liu: The code presented in this note is apparently new. The reader may supply a more formal proof. The "apology" in the last paragraph is really not necessary in that the ease of encoding and decoding for this scheme offsets its lack of sophistication. Its rate of information is asymptotically nearly twice that of the repetition codes [**1**].

Codes with higher rates (the "really clever schemes") were known early on in the history of error-correcting codes (see for example [**2**], [**3**], and [**5**]). A recent publication [**6**] gives an interesting historical account and shows the interrelationship of error-correcting codes with other areas of mathematics. A definitive treatise [**4**] details the state of the art as well as listing over one thousand references.

References

[1] F. L. Alt, A Bell Telephone Laboratories' computing machine (I), Math Tables and Aids to Comput. (Math. Comput.), 3 (1948/49) 1–13.

[2] M. J. E. Golay, Notes on digital coding, Proc. IRE (IEEE), 37 (1949) 657.

[3] R. W. Hamming, Error detecting and correcting codes, Bell System Tech. J., 29 (1950) 147–160.

[4] F. J. MacWilliams and N. J. A. Sloane, The Theory of Error-Correcting Codes, North Holland Pub. Co., Amsterdam, 1977.

[5] C. E. Shannon, A mathematical theory of communication, Bell System Tech. J., 27 (1948) 379–423, 623–656.

[6] T. M. Thompson, From Error-Correcting Codes through Sphere Packings to Simple Groups, Mathematical Association of America, Washington, DC, 1983.

The Mathematics of Double Entry Bookkeeping

DAVID P. ELLERMAN
Boston College
Chestnut Hill, MA 02167

For several centuries, the double entry system has been the bookkeeping system used by most sizable business enterprises throughout the world. It is little known in mathematics and it is virtually unknown in accounting that the double entry system is based on a well-known mathematical construction of undergraduate algebra, the group of differences, in which the integers are represented as equivalence classes of ordered pairs of natural numbers.

The T-accounts of double entry bookkeeping are precisely the ordered pairs of the group of differences construction. With the one exception of a paragraph by D. E. Littlewood (see below), the author has not found a single mathematics book, elementary or advanced, popular or esoteric, which notes that this construction is the theoretical basis of a mathematical technique applied, day in and day out, in the mundane world of business for over five centuries. And even though the construction is standard fare in an undergraduate modern algebra course, the connection with double entry bookkeeping is totally absent in the accounting literature (see References; [**4**] gives a review of the literature in mathematical accounting).

The encounters between mathematics and double entry bookkeeping have been so sparse that the highlights can be easily specified. A description of double entry bookkeeping was first published by the Italian mathematician Luca Pacioli in 1494 [**9**]. The system had been developed in Italy during the fourteenth century. Although Pacioli's system was governed by precise rules, his presentation was in a practical and nonmathematical form.

As an abstract mathematical construction, the group of differences seems to have been first published by Sir William Rowan Hamilton in 1837 [**6**]. Hamilton presented the ordered-pairs construction of the integers as a prelude to his ordered-pairs treatment of the complex numbers. He made no mention of bookkeeping although accountants had, at that time, been using the intuitive algebra of the ordered pairs called "*T*-accounts" for about four centuries.

Arthur Cayley (1821–1895) was one of the few later mathematicians who wrote about double entry bookkeeping. In the year before his death, he published a small pamphlet entitled *The Principles of Book-keeping by Double Entry* in which he wrote:

> The Principles of Book-keeping by Double Entry constitute a theory which is mathematically by no means uninteresting: it is in fact like Euclid's theory of ratios an absolutely perfect one, and it is only its extreme simplicity which prevents it from being as interesting as it would otherwise be. [**1**, Preface]

In the pamphlet, Cayley did not present a mathematical formulation, but only described double entry bookkeeping in the practical informal terms familiar to Cayley from his fourteen years of work as a lawyer. However, in his presidential address to the British Association for Advancement of Science, Cayley hinted that the "notion of a negative magnitude" is "used in a very refined manner in bookkeeping by double entry" [**2**, p. 434].

Another brief but insightful observation was made in a semipopular work by D. E. Littlewood in which he noted that the ordered pairs in the group of differences construction function like the debit and credit balances in a bank account.

> The bank associates two totals with each customer's account, the total of moneys credited and the total of moneys withdrawn. The net balance is then regarded as the same if, for example, the credit amounts of £102 and the debit £100, as if the credit were £52 and the debit £50. If the debit exceeds the credit the balance is negative.
>
> This model is adopted in the definition of signed integers. Consider pairs of cardinal numbers (a, b) in which the first number corresponds to the debit, and the second to the credit. A definition of equality is adopted such that
>
> $$(a,b)=(c,d)$$
>
> if and only if $a+d=b+c$. [**8**, p. 18]

Some modern accounting theorists believe that the mathematical treatment of double entry bookkeeping must involve transaction matrices. The presentation of transactions involving scalars can be facilitated using a square array or table of scalars usually called a "transactions matrix." These transactions tables were first used by the English mathematician Augustus DeMorgan [**3**], and have been popularized a century later by the American mathematician John Kemeny and his colleagues in an influential text [**7**].

Transactions tables have, however, retarded the development of a mathematical formulation of double entry bookkeeping. As will be seen below, double entry bookkeeping lives in group theory, *not* in matrix algebra. When double entry bookkeeping is mathematically formulated using the group of differences, it can be generalized to new systems of accounting such as vector accounting (or even "fraction accounting" where multiplication replaces addition). But the relatively superficial use of matrix algebra involved in "transactions matrices" does not generalize to these new domains [**4**, Chapter 12, section 2, "Transactions Matrices"]. Hence transactions tables will not be used here.

In what follows, we give an elementary introduction to the modern mathematical formulation of double entry bookkeeping. The basic concepts of double entry bookkeeping are introduced in their natural mathematical context.

The construction of the Pacioli group

Double entry bookkeeping is based on the construction of the integers (positive and negative) as equivalence classes of ordered pairs of natural numbers. The ordered pairs of this construction correspond to the T-accounts of double entry bookkeeping. The left-hand entry in the ordered pair corresponds to the **debit** side of the T-account, and the right-hand entry to the **credit** side. The notation $[d//c]$ for a T-account is drawn from Pacioli himself.

> At the beginning of each entry, we always provide "per", because, first, the debtor must be given, and immediately after the creditor, the one separated from the other by two little slanting parallels (virgolette), thus, //,... . [**9**, p. 43]

Thus a general T-account with a debit entry of d and a credit entry of c will be represented as

$$[d//c]=\begin{array}{c|c}\text{Debits}&\text{Credits}\\\hline d&c\end{array}.$$

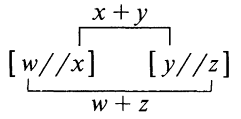
Since the label "T-account" will be used later in a specific accounting context, the general ordered pairs $[d//c]$ will be called ***T*-terms**. The additive group of integers is obtained by defining addition on T-terms and then imposing an equivalence relation compatible with the addition operation. Given T-terms $[w//x]$ and $[y//z]$, their **sum** is defined as the T-term obtained by adding debit to debit and credit to credit, i.e.,

$$[w//x]+[y//z]=[w+y//x+z].$$

The **zero *T*-term** $[0//0]$ is the T-term which 'acts' like zero in the sense that adding it to any T-term makes no difference, i.e.,

$$[d//c]+[0//0]=[d//c].$$

Given two T-terms $[w//x]$ and $[y//z]$, the **cross-sums** are the two numbers obtained by adding the debit in one to the credit in the other:



Cross-sums: $x+y \qquad w+z$.

Two T-terms are set **equal** if their cross-sums are equal, i.e.,

$$[w//x]=[y//z]\quad\text{if and only if}\quad x+y=w+z.$$

This equivalence relation is compatible with addition in the sense that if

$$[w//x]=[w'//x']\quad\text{and}\quad[y//z]=[y'//z']$$

then

$$[w+y//x+z]=[w'+y'//x'+z'].$$

Thus addition is well-defined on equivalence classes independently of their representatives. The notation $[d//c]$ will henceforth be used to represent the equivalence class of the ordered pair.

The numbers occurring in a T-term can never be negative, but we can still define the negative of a T-term without negative numbers. The negative of a T-term $[y//z]$ is another T-term such that when added to $[y//z]$ the sum is the zero T-term. It suffices to reverse the debit and credit entries. Hence we define the negative or **inverse** of a T-term $[y//z]$ as its 'reverse' $[z//y]$, since

$$[y//z]+[z//y]=[y+z//y+z]=[0//0].$$

This completes the definition of the ordered-pairs construction of the integers from the natural numbers. In view of the connection with double entry bookkeeping, we will call it the **Pacioli group**.

The double entry method of bookkeeping

The double entry method uses the Pacioli group to perform additive algebraic operations on equations. First we must translate or **encode** equations into the Pacioli group. A T-term equal to the zero T-term $[0//0]$ will be called a **zero-term**. For example, $[x//x]$ is a zero-term even when x is nonzero. The translation of equations into the Pacioli group is very simple: equations between nonnegative numbers correspond to zero-terms. That is, for any nonnegative numbers w and y,

$$w=y$$

if and only if

$$[w//0]+[0//y]=[w//y]=[0//0].$$

In more general terms, given any equation where all numbers are nonnegative such as $w+\cdots+x=y+\cdots+z$, we encode each left-hand-side number as a **debit-balance** T-term, such as $[w//0]$, and we encode each right-hand-side number as a **credit-balance** T-term, such as $[0//y]$. Then the original equation holds if and only if the sum of the encoded T-terms is a zero-term:

$$w+\cdots+x=y+\cdots+z$$

if and only if

$$[w//0]+\cdots+[x//0]+[0//y]+\cdots+[0//z]$$

is a zero-term.

This translation or encoding of equations into zero-terms works even if the original equation contains negative numbers since any equation can be transformed to one all of whose terms are positive by transferring the negative numbers to the other side.

In double entry bookkeeping, transactions must be recorded in such a way as to maintain the truth of an equation such as the balance sheet equation:

$$\text{Assets} = \text{Liabilities} + \text{Net Worth}.$$

That is, transactions must be recorded by valid algebraic operations which transform equations into equations. In the Pacioli group, an equation translates into a zero-term, so a valid algebraic operation would be an operation that transforms zero-terms (equations) into zero-terms (equations). But there is *only one such operation: add a zero-term*. Zero plus zero equals zero. Thus a transaction must be represented by a zero-term to be added to the zero-term representing the balance sheet equation.

In bookkeeping, the **double entry principle** is that each transaction must be recorded with equal debits and credits. The mathematical basis for this principle is simply that transactions are represented by zero-terms (so the debits must equal the credits in the transaction). In double entry bookkeeping, the zero-terms arising as the representations of equations (e.g., the balance sheet equation) and of transactions will be called respectively **equational zero-terms** and **transactional zero-terms**.

Any valid (additive) algebraic operation on an equation then boils down to one scheme:

$$\begin{array}{l} \text{original equational zero-term} \\ \underline{+\ \text{transactional zero-term}\quad} \\ =\text{final equational zero-term.} \end{array}$$

In bookkeeping, there are many transactions to record, but each is still represented by a transactional zero-term. The result of adding many zero-terms to the original equational zero-term still yields another zero-term, the final equational zero-term.

It remains to specify how to reverse the translation process, how to **decode** zero-terms as equations. A zero-term, such as an equational zero-term, is a sum of T-terms. It is not itself an equation with a left- and right-hand side. Indeed, the T-terms can be shuffled around in any order. To decode a zero-term into an equation, one can use any criterion one wishes to divide the T-terms into two sets, L and R. Then construct an equation as follows: if a T-term $[d//c]$ is in the set L, decode it as the number $d-c$ on the left-hand side of the equation, and if $[d//c]$ is in the set R, decode it as $c-d$ on the right-hand side of the equation. This procedure will always yield a valid equation, given a zero-term. For example, consider the zero-term

$$[6//2]+[7//2]+[1//7]+[13//16].$$

Let the set L be, say, $[6//2]$ and $[1//7]$, and thus the set R contains the remaining T-terms, $[7//2]$ and $[13//16]$. The set L decodes as $6-2+1-7$ on the left-hand side, and R decodes as $2-7+16-13$ on the right-hand side, so we have the equation:

$$6-2+1-7=2-7+16-13.$$

In bookkeeping, the T-accounts in the final equational zero-term would be put in the sets L and R according to the side of the initial balance sheet equation from which the accounts were originally encoded.

A numerical example of double entry bookkeeping

Double entry bookkeeping is used to update the balance sheet equation

$$\text{Assets} = \text{Liabilities} + \text{Net Worth}$$

to show the effects of economic transactions. We develop a numerical example using very broad accounts (such as the three above). For simplicity, we will not use temporary or flow accounts such as Revenue or Expenses. The mathematical structure is the same regardless of the accounts used.

The initial equation must have all accounts expressed as positive numbers:

$$\begin{array}{ccccc} \text{Assets} & & \text{Liabilities} & & \text{Net Worth} \\ 15000 & = & 10000 & + & 5000. \end{array} \tag{1}$$

It is the position of the account in the all-positive equation that identifies the account as a left-hand side (LHS) or **debit-balance account**, or as a right-hand side (RHS) or **credit-balance account**.

The equation is encoded as an equational zero-term by encoding each debit-balance account as a debit-balance T-term $[d//0]$, and each credit-balance account as a credit-balance T-term $[0//c]$. Hence balance sheet equation (1) yields the equational zero-term

$$\begin{array}{ccccc} \text{Assets} & & \text{Liabilities} & & \text{Net Worth} \\ [15000//0] & + & [0//10000] & + & [0//5000]. \end{array}$$

These T-terms with accounting labels, like "Assets" and "Liabilities," attached to them may properly be called ***T*-accounts**. Since only plus signs appear between the T-terms, the plus signs may be left implicit and the T-terms may be reshuffled in any order. This yields what accountants call the

Ledger

$$\begin{array}{ccc} \text{Assets} & \text{Liabilities} & \text{Net Worth} \\ [15000//0] & [0//10000] & [0//5000]. \end{array}$$

Thus a ledger is just an abbreviated form (without plus signs) of an equational zero-term, i.e.,

"Ledger" = "Equational Zero-term."

The sum of all the credit entries in a zero-term must equal the sum of the debit entries. In the case of the ledger, that summation is precisely what is called the **trial balance** in accounting.

According to the double entry principle, each transaction is recorded by an equal debit d and credit c so it would be represented in the Pacioli group by the zero-term $[d//0]+[0//c]$ obtained by encoding the equation $d=c$. **Debiting** any X to an account means adding the debit T-term $[X//0]$ to the T-account, and **crediting** X to an account means adding the credit T-term $[0//X]$ to the T-account. Hence the transaction represented by the zero-term $[d//0]+[0//c]$ would be recorded by adding the debit part $[d//0]$ and the credit part $[0//c]$ to the appropriate T-accounts.

It is a common mistake of nonaccountants to think that "debit" means "negative." But a debit-balance account like Assets does not have a negative balance. To debit an account does not necessarily mean to subtract from the balance in the account. That is only true for credit-balance accounts. Debiting a debit-balance account means *adding* to the account's balance.

Each transaction is recorded by adding the debit and credit parts of the transactional zero-term to the appropriate *T*-accounts. In accounting, the **journal** is the listing of the debit and credit parts of the transactional zero-terms and the *T*-accounts to which they are added. Thus

"Journal" = "List of Transactional Zero-terms."

Consider three simple transactions:

(1) $1200 is expended on inputs used in production,
(2) $1500 of outputs is produced and sold, and
(3) $800 of principal is paid on a loan.

Transaction (1) covers both the expenditure of $1200 to purchase inputs and the use of those inputs in production. Thus the debit-balance account of Assets is to be reduced by crediting it with $1200. The new assets produced are not recorded until transaction (2), so the other end of transaction (1) is a reduction in the credit-balance account of Net Worth by debiting it $1200. Transaction (2) records the production and sale of $1500 of new assets with an addition to the debit-balance account of Assets (a $1500 debit) and an addition to the credit-balance account of Net Worth (a $1500 credit). Transaction (3) records an $800 loan payment by equally reducing the Assets and Liabilities accounts, i.e., by crediting Assets and debiting Liabilities by $800.

	Journal	
Trans.	Accounts and Description	Debit Credit
	Net Worth	[1200 // 0]
1	Assets	[0 // 1200]
	$1200 expended on inputs	
	Assets	[1500 // 0]
2	Net Worth	[0 // 1500]
	$1500 of outputs produced and sold	
	Liabilities	[800 // 0]
3	Assets	[0 // 800]
	$800 principal payment on a loan	

The debit and credit parts of the transactional zero-terms can then be added or "**posted**" to the appropriate *T*-accounts in the ledger. Thus

"Posting to the Ledger" = "Adding the Transactional Zero-terms to the Equational Zero-term."

Assets	Liabilities	Net Worth
[15000// 0]	[0 //10000]	[0 //5000]
[0 //1200]		[1200// 0]
[1500 // 0]		[0 //1500]
[0 // 800]	[800// 0]	
[16500//2000]	[800//10000]	[1200//6500]

This yields the updated ledger or equational zero-term:

Assets	Liabilities	Net Worth
[16500//2000]	[800//10000]	[1200//6500].

Finding the balance in a *T*-account means subtracting as much as possible from each side so long as neither side becomes negative. The resulting *T*-term is said to be in **reduced form**. The quick way to put an account in reduced form is to take the minimum of the debit and credit entries, and subtract it from both sides. A *T*-account with a debit or credit entry of zero is in reduced form. The other possibly nonzero entry in the T-account in reduced form is called the **balance** of the account. Putting the above T-accounts in reduced form yields:

Assets	Liabilities	Net Worth
[14500//0]	[0//9200]	[0//5300].

It remains only to decode the equational zero-term (= ledger with plus signs between the T-accounts) to obtain the final balance sheet equation. That involves selecting the two sets L and R of T-terms to be decoded respectively on the LHS and RHS of the final equation. The selection is obvious; decode each T-account onto the side of the equation that it originally came from. The Assets account is in the set L, and Liabilities and Net Worth are in the set R. Hence the ledger decodes to yield the:

Final Balance Sheet Equation

Assets		Liabilities		Net Worth
14500	=	9200	+	5300.

Why the double entry system?

In conclusion, let us consider the distinctive features of the double entry system of bookkeeping. It is often thought that the characteristic feature of the system is the double aspect of the entries, the fact that at least two accounts are affected by each transaction. But the double entry system is a system of *recording* transactions. The fact that two or more accounts are affected is a characteristic of the transaction itself, not of the recording method. The same two or more accounts would be affected by any method of recording transactions which is based on updating a *complete* accounting equation. This is the general mathematical fact, having nothing in particular to do with accounting, that two or more terms must be changed when changes are made in an equation. Given an equation

$$a+b+\cdots+c=x+y+\cdots+z$$

it is impossible to change only one term, such as b, and still maintain the validity of the equation.

Another method of recording transactions, without using the double entry machinery of debits and credits, would be to directly perform additions and subtractions of positive numbers to the accounts in the balance sheet equation. But then there is no quick check on the plausibility of the transaction used to record a business event. Suppose that an event was formulated as the 'transaction' of adding $100 to both Liabilities and Net Worth. Is that possible? Some thought is required to see that this formulation of the business event could not possibly be correct. Much more thought would be necessary for a multiple (two or more) entry transaction in an accounting system with hundreds of accounts. Yet the check is immediate in the double entry system. Liabilities and Net Worth are both credit-balance accounts so the proposed 'transaction' is a double credit in violation of the double entry principle.

A third method of recording transactions, which does involve a quick check, is obtained by moving all the terms in the balance sheet equation to one side:

$$-\text{Assets}+\text{Liabilities}+\text{Net Worth}=0.$$

Then each bona fide transaction would be recorded by adding equal positive and negative amounts to the accounts (some of which would have a negative balance). If the positive and negative amounts did not sum to zero, it could not represent a bona fide transaction. The problem with this accounting method is the counterintuitiveness of having negative balances in some accounts, e.g., Assets. An increase in such accounts would be represented by adding a negative number to the (negative) balance in the accounts, and a decrease would be represented by adding a positive number to the (negative) balance.

All three of these methods are logically sound, and all three record a transaction with "double entries" affecting two (or more) accounts. The popularity of the double entry method down through the centuries seems to be based on the fact that it provides a quick check on the plausibility of a transaction (the double entry principle of equal debits and credits) and that it uses the symmetric T-accounts (the ordered pairs of the group of differences construction) to represent the balance in each account as a positive number.

References

[1] A. Cayley, The Principles of Book-keeping by Double Entry, Cambridge University Press, Cambridge, 1894.

[2] ______, Presidential address to the British Association for the Advancement of Science, The Collected Mathematical Papers of Arthur Cayley, Vol. XI, Cambridge University Press, Cambridge, 1896.

[3] Augustus DeMorgan, On the main principle of book-keeping, Elements of Arithmetic, James Walton, London, 1869.

[4] David Ellerman, Economics, Accounting, and Property Theory, D. C. Heath, Lexington, Mass. 1982.

[5] John B. Geijsbeek, Ancient Double-Entry Bookkeeping, Scholars Book Company, Houston, 1914.

[6] Sir William Rowan Hamilton, Theory of conjugate functions, or algebraic couples; with a preliminary and elementary essay on algebra as the science of pure time, Trans. Royal Irish Academy, XVII (1837) 293–422. Also reprinted in The Mathematical Papers of Sir William Rowan Hamilton, III Algebra, H. Halberstam and R. E. Ingram (eds), Cambridge University Press, Cambridge, 1967.

[7] J. Kemeny, A. Schleifer, J. L. Snell, and G. Thompson, Finite Mathematics with Business Applications, Prentice-Hall, Englewood Cliffs, N.J., 1962.

[8] D. E. Littlewood, The Skeleton Key of Mathematics, Harper Torchbooks, New York, 1960 (orig. published in 1949).

[9] L. Pacioli, Summa de Arithmetica, Geometrica, Proporcioni et Proporcionalita, trans. by J. B. Geijsbeek as Ancient Double-Entry Bookkeeping, reprinted by Scholar Book Company (orig. published in 1494).

Differential Equations and Analytic Geometry

R. Spigler
Courant Institute of Mathematical Sciences
New York University
New York, NY 10012

In this note we give an example in which analytic geometry proves useful in the study of linear dependence and independence of certain quadratic combinations of linearly independent functions. By taking some care, we can apply the result to certain ordinary differential equations.

Often in physics we meet pairs of quantities such as voltage and current, electric and magnetic fields, or pressure and velocity, whose product also has an important physical meaning. In each of these examples it represents the power or energy per unit time. We may sometimes wish to describe such a product of two quantities when the two quantities satisfy two given linear homogeneous ordinary differential equations (ODE's), but *without solving* these equations. For example, we may wish to obtain the power series expansion for the product, without passing through the construction of the series for the two associated quantities. It would therefore be useful to construct first the linear homogeneous ODE satisfied by such a product and then to solve *it* by series. Moreover, the knowledge of Nth order ODE's whose solutions are products of solutions of two lower-order equations would reduce the study of certain higher-order equations to the separate (easier) analysis of lower-order equations.

In [**9**] the following was proved.

Theorem. *Given two linear homogeneous ordinary differential equations, whose orders are m and n, and with smooth coefficients on some real open interval, it is possible to construct a linear homogeneous ODE, such that among its solutions one finds all the products of solutions of the two given ODE's, and its order N is the lowest possible*; *moreover,* $m+n-1 \leqslant N \leqslant mn$.

plausibility of a transaction (the double entry principle of equal debits and credits) and that it uses the symmetric T-accounts (the ordered pairs of the group of differences construction) to represent the balance in each account as a positive number.

References

[1] A. Cayley, The Principles of Book-keeping by Double Entry, Cambridge University Press, Cambridge, 1894.

[2] ______, Presidential address to the British Association for the Advancement of Science, The Collected Mathematical Papers of Arthur Cayley, Vol. XI, Cambridge University Press, Cambridge, 1896.

[3] Augustus DeMorgan, On the main principle of book-keeping, Elements of Arithmetic, James Walton, London, 1869.

[4] David Ellerman, Economics, Accounting, and Property Theory, D. C. Heath, Lexington, Mass. 1982.

[5] John B. Geijsbeek, Ancient Double-Entry Bookkeeping, Scholars Book Company, Houston, 1914.

[6] Sir William Rowan Hamilton, Theory of conjugate functions, or algebraic couples; with a preliminary and elementary essay on algebra as the science of pure time, Trans. Royal Irish Academy, XVII (1837) 293–422. Also reprinted in The Mathematical Papers of Sir William Rowan Hamilton, III Algebra, H. Halberstam and R. E. Ingram (eds), Cambridge University Press, Cambridge, 1967.

[7] J. Kemeny, A. Schleifer, J. L. Snell, and G. Thompson, Finite Mathematics with Business Applications, Prentice-Hall, Englewood Cliffs, N.J., 1962.

[8] D. E. Littlewood, The Skeleton Key of Mathematics, Harper Torchbooks, New York, 1960 (orig. published in 1949).

[9] L. Pacioli, Summa de Arithmetica, Geometrica, Proporcioni et Proporcionalita, trans. by J. B. Geijsbeek as Ancient Double-Entry Bookkeeping, reprinted by Scholar Book Company (orig. published in 1494).

Differential Equations and Analytic Geometry

R. Spigler

Courant Institute of Mathematical Sciences
New York University
New York, NY 10012

In this note we give an example in which analytic geometry proves useful in the study of linear dependence and independence of certain quadratic combinations of linearly independent functions. By taking some care, we can apply the result to certain ordinary differential equations.

Often in physics we meet pairs of quantities such as voltage and current, electric and magnetic fields, or pressure and velocity, whose product also has an important physical meaning. In each of these examples it represents the power or energy per unit time. We may sometimes wish to describe such a product of two quantities when the two quantities satisfy two given linear homogeneous ordinary differential equations (ODE's), but *without solving* these equations. For example, we may wish to obtain the power series expansion for the product, without passing through the construction of the series for the two associated quantities. It would therefore be useful to construct first the linear homogeneous ODE satisfied by such a product and then to solve *it* by series. Moreover, the knowledge of Nth order ODE's whose solutions are products of solutions of two lower-order equations would reduce the study of certain higher-order equations to the separate (easier) analysis of lower-order equations.

In **[9]** the following was proved.

THEOREM. *Given two linear homogeneous ordinary differential equations, whose orders are m and n, and with smooth coefficients on some real open interval, it is possible to construct a linear homogeneous ODE, such that among its solutions one finds all the products of solutions of the two given ODE's, and its order N is the lowest possible; moreover,* $m+n-1 \leqslant N \leqslant mn$.

The lower estimate is not trivial. In [**8**] the same result was proved for the case of ODE's with constant coefficients and some related cases, by using completely different techniques. In any case, as long as we are concerned just with the vector spaces spanned by the solutions of the ODE's we can also invoke a Theorem due to I. Connell [**2**].

In the special case that the two given ODE's both are the *same third order* ODE ($m = n = 3$), we can give a direct and elementary proof of the Theorem based on purely geometric considerations (which focus on intersections of conics in Euclidean 2-space). Moreover, we obtain an improvement of the estimate on N in the Theorem. Our result is of interest since there are certain third-order linear ODE's satisfied by some elliptic functions (see [**5**, p. 510, n. 3.10], [**3**, pp. 553–554]). The equation for the products then is satisfied by certain squares and products of elliptic functions.

For the special case $m = n = 2$, several direct proofs of the Theorem show that $N = 3$ or 4 (see [**10**, pp. 147–149], [**9**, pp. 136–138, 144]). For the cases m and $n > 3$, the method we shall follow below would require discussing systems of quadratic forms in several variables, a subject studied in algebraic geometry.

Let us now consider a third-order linear homogeneous ODE, with real-valued smooth coefficients on some open interval I of the real line. Let $u_1(t)$, $u_2(t)$, $u_3(t)$ be a system of *linearly independent solutions* of the equation. Then the corresponding Wronskian is $W(u_1(t), u_2(t), u_3(t)) \neq 0$ in I.

We want to discuss the order of the linear homogeneous ODE whose solutions include $(c_1 u_1 + c_2 u_2 + c_3 u_3)^2$, c_i, $i = 1,2,3$ being arbitrary constants, so consider the set

$$\begin{array}{ccc} u_1^2 & u_1 u_2 & u_1 u_3 \\ u_2 u_1 & u_2^2 & u_2 u_3 \\ u_3 u_1 & u_3 u_2 & u_3^2 \end{array} \tag{1}$$

of the functions which are candidates to span the vector space of solutions of the ODE "for the products."

In order to reduce our discussion to that of the linear independence or dependence of the functions $u_i u_j, i, j = 1,2,3$ in (1), we recall first that the statements "Wronskian identically zero" and "linear dependence" *are not equivalent*. More precisely, suppose that $u_1(t), u_2(t), \ldots, u_n(t)$ are n functions, continuous together with their first $(n-1)$ derivatives in some real open interval I.

If $u_1(t), u_2(t), \ldots, u_n(t)$ are linearly dependent in I, then $W(u_1(t), u_2(t), \ldots, u_n(t)) \equiv 0$ in I. Conversely, if $W(u_1(t), u_2(t), \ldots u_n(t)) \equiv 0$ in I, it is not true that $u_1(t), u_2(t), \ldots u_n(t)$ are linearly dependent in I. The first and simplest example of this fact was probably given by G. Peano [**6**], [**7**]: take $u_1 = t^2$, $u_2 = t|t|$ in the interval $(-a, a)$, $a > 0$. However, if $W(u_1(t), u_2(t), \ldots, u_n(t)) \equiv 0$ in I, there exists an interval $J \subseteq I$, such that $u_1(t), u_2(t), \ldots, u_n(t)$ are linearly dependent in J, (see, e.g. [**4**, p. 48]). It follows that the condition $W(u_1(t), u_2(t), \ldots, u_n(t)) \neq 0$ in I is equivalent to the linear independence *in every subinterval J* of I.

By symmetry, the nine functions $u_i u_j, i, j = 1,2,3$, reduce to six, as is clear from (1), so that the upper bound for N is lowered to 6 in this case. As $m + n - 1 = 5$, we have only to prove that $N \geqslant 5$.

First of all, certainly $N \geqslant 3$. In fact, we can show the functions $u_1^2, u_1 u_2, u_1 u_3$ are linearly independent in every subinterval J^* of J, where J is any of the subintervals of I where $u_1(t)$ does not vanish. For if

$$\sum_{i=1}^{3} c_i u_1(t) u_i(t) = u_1(t) \sum_{i=1}^{3} c_i u_i(t) \equiv 0 \quad \text{in any } J^* \subseteq J,$$

then since $u_1(t), u_2(t), u_3(t)$ form a fundamental system of solutions for the original third-order equations and $u_1(t)$ cannot vanish except at most at finitely many points in every compact subinterval of I, and nowhere in J, we must have $c_i = 0$, $i = 1, 2, 3$. Let us restrict ourselves to the interval J.

We only need to prove that the case $N = 4$ *cannot* occur. The proof will be given by contradiction. Suppose that there are, among the functions in (1), *four* linearly independent solutions of the equation "for the products." Without loss of generality, let them be u_1^2, u_1u_2, u_1u_3 (which *are* linearly independent), and u_2u_3. Then any other function in (1) can be written as a linear combination of these, and in particular,

$$\begin{aligned} u_2^2 &= c_{21}u_1^2 + c_{22}u_1u_2 + c_{23}u_1u_3 + c_{24}u_2u_3 \\ u_3^2 &= c_{31}u_1^2 + c_{32}u_1u_2 + c_{33}u_1u_3 + c_{34}u_2u_3 \end{aligned} \qquad t \in J, \tag{2}$$

where $\Sigma_{j=1}^4 |c_{ij}| > 0$, $i = 2, 3$. The equations in (2) can be interpreted as the equations of two conics, in homogeneous coordinates. Setting $x = u_2/u_1$, $y = u_3/u_1$ (recall that u_1 vanishes at most at finitely many points in every compact subinterval of I and nowhere in J), we get

$$\begin{cases} x^2 = c_{21} + c_{22}x + c_{23}y + c_{24}xy \\ y^2 = c_{31} + c_{32}x + c_{33}y + c_{34}xy. \end{cases} \tag{3}$$

It follows that the study of system (3) is reduced to discussing the intersection of two conics in the (ordinary) plane. Here any two distinct conics meet at most at four points, or, in the case of two degenerate conics, can have a straight line in common.

If the conics meet at no point, the system (3) has no (real) solutions and, therefore, $N > 4$ (four linearly independent solutions do not suffice to span the whole set of solutions of the equations for the products).

If the conics intersect at some point $(\bar{x}, \bar{y})$ then (3) has the solution

$$\frac{u_2}{u_1} \equiv \bar{x}, \qquad \frac{u_3}{u_1} \equiv \bar{y},$$

and thus $u_1(t)$ and $u_2(t)$, as well as $u_1(t)$ and $u_3(t)$, would be linearly dependent in J. In this case, all points (x, y) such that

$$ax + by + c = 0,$$

where a, b, c are constants and $|a| + |b| > 0$, would be solutions to system (3). This means that $au_2(t) + bu_3(t) + cu_1(t) \equiv 0$ in J, so that $u_1(t), u_2(t), u_3(t)$ would be linearly dependent in J. This is false and, therefore, we cannot have $N = 4$. Q.E.D.

Thus $N = 5$ or 6, the precise value of N depending on the particular third-order differential equation chosen in the class of all third-order linear homogeneous ODE's with smooth coefficients in some real open interval. That is, for a given equation in such a class, N will be 5 *or* 6.

For example, the equation with constant coefficients

$$z''' - (a+2)z'' + 2az' = 0, \qquad 0 < a < 2,$$

has the fundamental system of solutions $\{1, e^{ax}, e^{2x}\}$. For any fixed value of a, $0 < a < 1$ or $1 < a < 2$, the corresponding equation for the products has order $N = 6$ and the fundamental system $\{1, e^{ax}, e^{2x}, e^{2ax}, e^{(a+2)x}, e^{4x}\}$. For $a = 1$ we obtain an equation of order $N = 5$ and the fundamental system $\{1, e^x, e^{2x}, e^{3x}, e^{4x}\}$.

Another example is given by the equation

$$z''' - (a+3)z'' + (3a+2)z' - 2az = 0, \qquad 1 < a < 2,$$

which has the fundamental system of solutions $\{e^x, e^{ax}, e^{2x}\}$. For any fixed value of the parameter $a, 1 < a < 3/2$ or $3/2 < a < 2$, the equation for the products has order $N = 6$, having the fundamental system $\{e^{2x}, e^{(a+1)x}, e^{3x}, e^{2ax}, e^{(a+2)x}, e^{4x}\}$, while N drops to 5 when $a = 3/2$.

We note in closing that our result can be obtained in the framework of projective geometry. We may think of u_1, u_2, u_3 in (2) as projective coordinates in the projective complex plane. In such a plane any two conics meet exactly at four points, provided that the intersections are counted with their multiplicity, unless they degenerate in a pair of (distinct or coinciding) straight lines and one of them be in common. The rest follows above. For the use of projective coordinates in studying oscillations of solutions of a single third-order ODE, see [**1**].

The interpretation of solutions of linear homogeneous ODE's of nth order as curves in a $(n-1)$-dimensional space was well known (see [**11**, ch. II,III]). When t varies (in J), $(u_1(t), u_2(t), u_3(t))$ above represent parametrically a curve in the space (u_1, u_2, u_3). Alternatively, this curve can be described by the nonhomogeneous coordinates $(x(t), y(t))$, where $x(t) = u_2(t)/u_1(t)$, $y(t) = u_3(t)/u_1(t)$, in the projective plane (x, y).

References

[1] G. D. Birkhoff, On the solutions of ordinary linear homogeneous differential equations of the third order, Annals Math., 12 (2) (1911) 103–127; also Collected Mathematical Papers, Vol. I, Amer. Math. Soc., New York, 1950, pp. 49–73.

[2] I. Connell, An estimate for the dimension of the product of two vector spaces, Linear and Multilinear Algebra, 4 (1977) 273–275.

[3] G. H. Halphen, Traité des Fonctions Elliptiques et de leurs Applications, Gauthier-Villars, Paris, 1888, 2nd vol.

[4] W. Hurewicz, Lectures on Ordinary Differential Equations, The Technology Press, M.I.T., and J. Wiley & Sons, New York, 1958.

[5] E. Kamke, Differentialgleichungen: Lösungsmethoden und Lösungen, Chelsea, New York, 1948.

[6] G. Peano, Sur le déterminant Wronskien, Mathesis, 9 (1889) 75–76.

[7] ______, Sur les Wronskiens, Mathesis, 9 (1889) 110–112.

[8] R. Spigler, An application of group theory to matrices and to ordinary differential equations, Linear Algebra and its Applications, 44 (1982) 143–151.

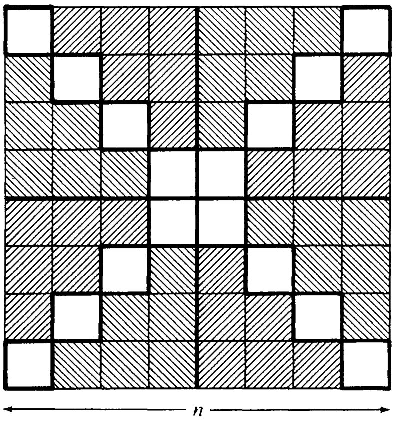
[9] ______, The linear differential equation whose solutions are the products of solutions of two given differential equations, Jour. Math. Anal. Appl., 98 (1984) 130–147.

[10] G. N. Watson, A Treatise on the Theory of Bessel Functions, 2nd ed., Cambridge Univ. Press, Cambridge, England, 1958.

[11] E. J. Wilczynski, Projective Differential Geometry of Curves and Ruled Surfaces, Chelsea Publ. Co., New York, 1905.

Proof without words: Square of an even positive integer

$$n^2 - 2n = 4\left(\frac{n}{2} - 1\right)\left(\frac{n}{2}\right) = 8 \sum_{i=1}^{n/2-1} i$$



—Edwin G. Landauer
General Physics Corp.

plane any two conics meet exactly at four points, provided that the intersections are counted with their multiplicity, unless they degenerate in a pair of (distinct or coinciding) straight lines and one of them be in common. The rest follows above. For the use of projective coordinates in studying oscillations of solutions of a single third-order ODE, see [**1**].

The interpretation of solutions of linear homogeneous ODE's of nth order as curves in a $(n-1)$-dimensional space was well known (see [**11**, ch. II,III]). When t varies (in J), $(u_1(t), u_2(t), u_3(t))$ above represent parametrically a curve in the space (u_1, u_2, u_3). Alternatively, this curve can be described by the nonhomogeneous coordinates $(x(t), y(t))$, where $x(t) = u_2(t)/u_1(t)$, $y(t) = u_3(t)/u_1(t)$, in the projective plane (x, y).

References

[1] G. D. Birkhoff, On the solutions of ordinary linear homogeneous differential equations of the third order, Annals Math., 12 (2) (1911) 103–127; also Collected Mathematical Papers, Vol. I, Amer. Math. Soc., New York, 1950, pp. 49–73.

[2] I. Connell, An estimate for the dimension of the product of two vector spaces, Linear and Multilinear Algebra, 4 (1977) 273–275.

[3] G. H. Halphen, Traité des Fonctions Elliptiques et de leurs Applications, Gauthier-Villars, Paris, 1888, 2nd vol.

[4] W. Hurewicz, Lectures on Ordinary Differential Equations, The Technology Press, M.I.T., and J. Wiley & Sons, New York, 1958.

[5] E. Kamke, Differentialgleichungen: Lösungsmethoden und Lösungen, Chelsea, New York, 1948.

[6] G. Peano, Sur le déterminant Wronskien, Mathesis, 9 (1889) 75–76.

[7] ______, Sur les Wronskiens, Mathesis, 9 (1889) 110–112.

[8] R. Spigler, An application of group theory to matrices and to ordinary differential equations, Linear Algebra and its Applications, 44 (1982) 143–151.

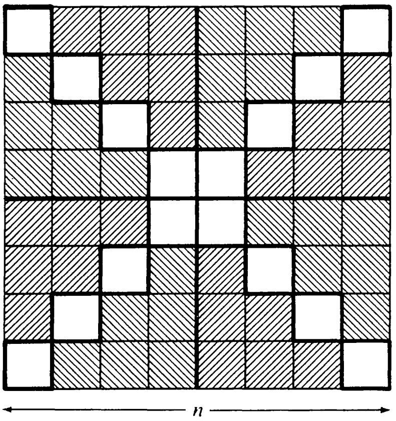
[9] ______, The linear differential equation whose solutions are the products of solutions of two given differential equations, Jour. Math. Anal. Appl., 98 (1984) 130–147.

[10] G. N. Watson, A Treatise on the Theory of Bessel Functions, 2nd ed., Cambridge Univ. Press, Cambridge, England, 1958.

[11] E. J. Wilczynski, Projective Differential Geometry of Curves and Ruled Surfaces, Chelsea Publ. Co., New York, 1905.

Proof without words: Square of an even positive integer

$$n^2 - 2n = 4\left(\frac{n}{2} - 1\right)\left(\frac{n}{2}\right) = 8 \sum_{i=1}^{n/2-1} i$$



—Edwin G. Landauer
General Physics Corp.

PROBLEMS

LEROY F. MEYERS, Editor
G. A. EDGAR, Associate Editor
The Ohio State University

LOREN LARSON, Editor-elect
St. Olaf College

Proposals

To be considered for publication, solutions should be received by February 1, 1986.

1221. Let k and n be fixed integers with $1 < k < n$. Which subgroup of the symmetric group S_n is generated by the k-cycles $(x, x+1, \ldots, x+k-1)$ for $x = 1, 2, \ldots, n-k+1$? [*William A. McWorter, Jr., The Ohio State University.*]

1222. Given the nonnegative integer k, find the number of odd binomial coefficients $\binom{m}{r}$ with $0 \leqslant m < 2^k$. [*Marta Sved, The University of Adelaide, Australia.*]

1223. Let $(a_k)_{k=1}^{\infty}$ be a strictly increasing sequence of positive integers such that $\sum_{k=1}^{\infty}(1/a_k)$ converges.

(a) For each positive integer n, set

$$f(n) = \sum_{\substack{k=1 \\ a_k \neq n}}^{\infty} \frac{1}{|n - a_k|}.$$

Prove that $\liminf_{n \to \infty} f(n) = 0$. Give an example where $\lim_{n \to \infty} f(n) = 0$ fails.

(b) Prove that

$$\liminf_{l \to \infty} \sum_{k=1}^{l-1} \frac{1}{a_l - a_k} = 0 \quad \text{and} \quad \liminf_{l \to \infty} \sum_{k=l+1}^{\infty} \frac{1}{a_k - a_l} = 0.$$

Give an example where

$$\liminf_{l \to \infty} \sum_{\substack{k=1 \\ k \neq l}}^{\infty} \frac{1}{|a_k - a_l|} = 0$$

fails. [*Paul Erdös, Hungarian Academy of Sciences.*]

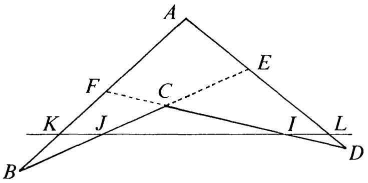
ASSISTANT EDITORS: DANIEL B. SHAPIRO and WILLIAM A. MCWORTER, JR., *The Ohio State University.*

*We invite readers to submit problems believed to be new. Proposals should be accompanied by solutions, if at all possible, and by any other information that will assist the editors. A problem submitted as a Quickie should have an unexpected, succinct solution. An asterisk (*) next to a problem number indicates that the proposer did not supply a solution.*

Solutions should be written in a style appropriate for Mathematics Magazine. Each solution should begin on a separate sheet containing the solver's name and full address. It is not necessary to submit duplicate copies.

New proposals and solutions to problems numbered 1221 and higher should be sent to Loren Larson, Mathematics Dept., St. Olaf College, Northfield, MN 55057. Solutions to proposals numbered 1220 and lower should be sent to Leroy F. Meyers, Mathematics Department, The Ohio State University, 231 W. 18th Ave., Columbus, Ohio 43210.

1224. Consider the nonconvex quadrilateral $ABCD$ as in the picture. Let I and J be chosen so that $DI = CF$ and $BJ = CE$. Let K and L be the points where the line IJ intersects AB and AD, respectively. Show that $KJ = IL$. [*Aristomenis Siskakis, Purdue University, Indianapolis.*]



1225. Show that

$$\sum_{n=1}^{\infty} \log^2\left(1+\frac{x}{n}\right) > 4\left[\log\Gamma(1+x) - 2\log\Gamma\left(1+\frac{x}{2}\right)\right]$$

for real $x > 0$. [*Robert E. Shafer, Berkeley, California.*]

Quickies

Answers to the Quickies are on p. 245.

Q698. Let D be a (circular) disk and C its boundary. Let p and q be two points on C. Show that there exists an ellipse E included in D which approximates C as closely as we wish and meets C precisely in the points p and q. [*Jan Mycielski, University of Colorado.*]

Q699. Prove that $(2n)!/n!(n+1)!$ is an integer for $n \geqslant 0$. [*Harry D. Ruderman, The Bronx, New York.*]

Q700. Evaluate

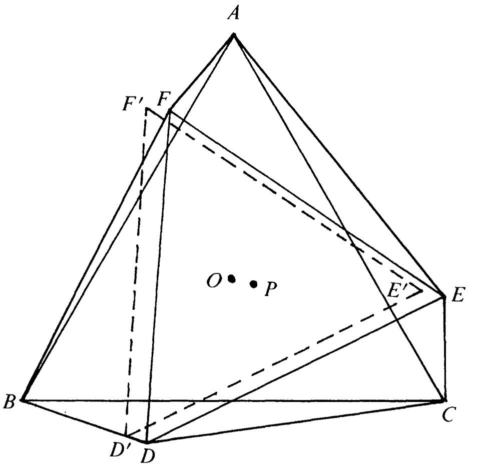
$$(n+1)\int_0^x \frac{(x-t)^n}{(1-t)^{n+2}}\,dt \quad \text{for} \quad x<1 \text{ and } n = 0,1,2,\ldots.$$

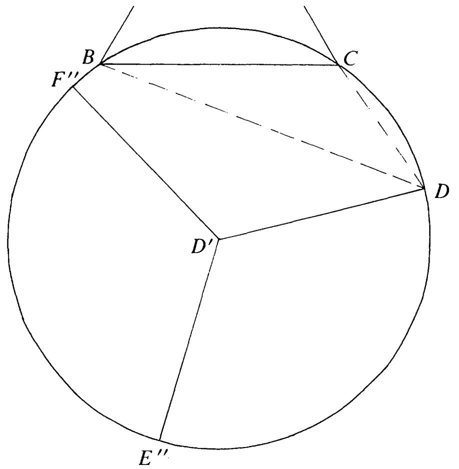
[*J. Schlosser and J. Howard, New Mexico Highlands University.*]

Solutions

Two Triangles in a Hexagon January 1983

1161. Two equilateral triangles are placed so that their intersection is a hexagon (not necessarily regular). The vertices of the equilateral triangles are connected to form an outer hexagon. Show that if three alternate angles of the outer hexagon are equal, then the triangles have the same center. [*Roger Izard, Dallas, Texas.*]





Solution: Let the centres of the equilateral triangles ABC and DEF be O and P, respectively, and let the angles at D, E, and F of the hexagon $AFBDCE$ be equal. Suppose that $O \neq P$. Then translate triangle DEF by the vector $\overrightarrow{PO}$ to form the equilateral triangle $D'E'F'$ with centre O. The respective images of A, C, E', and E under a clockwise rotation of 120° about O are C, B, D', and a new point E''. The respective images of A, B, F', and F under a counterclockwise rotation of 120° about O are B, C, D', and a new point F''. Since the segments $D'D$, $E'E$, and $F'F$ are parallel and equal, their images as a result of the rotations are three radii $D'D$, $D'E''$, $D'F''$ of a circle K with centre D' which make angles of 120° with each other. Since all vertices of the equilateral triangle $DE''F''$ lie on one side of BC, so does the circumcentre D' of that triangle. Since $\angle BE''C = \angle CEA = \angle BDC = \angle AFB = \angle BF''C$, the arc of K containing D, E'', and F'' and cut off by BC contains all points on one side of BC from which BC subtends the same angle as from D. Thus the segment BC is a chord of K which does not intersect the radii $D'D$, $D'E''$, and $D'F''$, and so must subtend an angle of less than 120° from D′, so that $\angle BDC < 60°$. But DE and DF intersect the open segment BC, so that $\angle BDC > 60°$. This contradiction shows that P and O must be the same.

THE COPS
Ottawa, Canada

Also solved by the proposer. There were two incomplete solutions in which the problem was reduced to an equivalent one which is just as hard to solve.

Lattice Point Triangles September 1984

1196. (a) Prove that the area of a triangle whose vertices are integer lattice points in the plane is always half an (even or odd) integer.

(b) Suppose three lattice points are chosen at random. What is the probability that the area of the triangle they determine is an integer? More precisely, if the points are chosen from a large rectangle, does the probability that the area is an integer converge as the dimensions of the rectangle grow without bound? [*Eric C. Nummela, New England College.*]

Solution I: (a) Pick's formula for the area of a polygon with vertices that have integer coordinates establishes this immediately. [*Ed. note.* The area of such a polygon is the number of interior lattice points, plus half the number of boundary lattice points, minus 1.]

HARRY D. RUDERMAN
The Bronx, New York

Solution II: (a) If the vertices are (a_1, b_1), (a_2, b_2), and (a_3, b_3), the area is well known to be the absolute value of

$$\frac{1}{2}\begin{vmatrix} a_1 & b_1 & 1 \\ a_2 & b_2 & 1 \\ a_3 & b_3 & 1 \end{vmatrix} = \frac{1}{2}\left[(a_2 - a_1)(b_3 - b_1) - (a_3 - a_1)(b_2 - b_1)\right].$$

This is always half an integer.

(b) Let the vertices be chosen at random in the rectangle $|a_i| \leqslant M, |b_i| \leqslant N$. We do not change the area A by translating the origin to (a_1, b_1), and in the translated coordinate system we have $2A = |a'_2 b'_3 - a'_3 b'_2|$. Hence

$$P(2A \text{ is even}) = P\big(a'_2 b'_3 \text{ is even and } a'_3 b'_2 \text{ is even}\big) + P\big(a'_2 b'_3 \text{ is odd and } a'_3 b'_2 \text{ is odd}\big),$$

and as $M, N \to \infty$, this sum converges to

$$\frac{3}{4} \cdot \frac{3}{4} + \frac{1}{4} \cdot \frac{1}{4} = \frac{5}{8}.$$

G. A. HEUER
Concordia College

Also solved by S. F. Barger, Robert E. Bernstein, Ada Booth, Stephen D. Bronn, Jordi Dou (Spain), Howard Eves, Alberto Facchini (Italy), Cornelius Groenewoud, Jerrold W. Grossman, Ray Haertel & Jack McCown, L. Kuipers (Switzerland), Andrew J. Lazarus (student), J. C. Linders (The Netherlands), Andy Martin & Chuck Parsons (students), Vania D. Mascioni (student, Switzerland), Mike Molloy (student, Canada), William A. Newcomb, Bill Olk (student), John Oman, Richard Parris, George Schillinger, Harry Sedinger, Robert E. Shafer, Jan Söderkvist (student, Sweden), Christine A. Sprengel & Lorraine L. Foster, Michael Vowe (Switzerland), John Ward (student), Steven R. Weston (student), and the proposer. Also solved (part (a) only) by Kenneth Bernstein and Steven Davis.

Groenewoud and *Parris* computed explicitly the probability that a lattice triangle in a finite rectangle has integer area. Parris's formula for a rectangle of m dots by n dots is

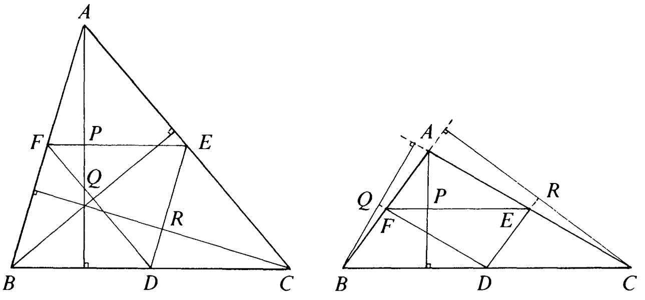
$$P(A \text{ is an integer}) = 1 - \frac{3}{8}\left(1 - \frac{r(m)}{m^2}\right)\left(1 - \frac{r(n)}{n^2}\right),$$

where $r(x)$ is the remainder on dividing x by 2. *Sprengel & Foster* found that $P(2A$ is divisible by the prime $p)$ $\to (p^2 + p - 1)/p^3$ as $m, n \to \infty$. *Shafer* noted that a lattice triangle in three dimensions need not have rational area.

Collinear Mid-Altitudes — September 1984

1197. Characterize the triangles of which the midpoints of the altitudes are collinear. [*Hüseyin Demir, Middle East Technical University, Ankara, Turkey.*]

Solution I: The midpoints of the altitudes of a triangle are collinear if and only if the triangle is right-angled.



Proof. We note first that the altitudes of a triangle all lie inside the triangle if it is acute-angled, while if it has an obtuse angle, two of the altitudes lie outside the triangle.

Let P, Q, and R be the midpoints of the altitudes from the vertices A, B, and C, respectively, of triangle ABC. If D, E, and F are the midpoints of the sides BC, CA, and AB, then P, Q, and R lie, respectively, on EF, FD, and DE, produced if necessary. By Pasch's axiom applied to the triangle DEF, the points P, Q, and R are collinear if and only if two of them coincide with two of D, E, F, in other words lie on the sides of triangle ABC. This occurs if and only if triangle ABC is right-angled.

J. H. WEBB
University of Cape Town
South Africa

Solution II: The altitudes of a triangle are concurrent at the orthocentre. This is the only property of the altitudes that we need make use of; the answer to the problem is just a special case of the following more general result.

Let D, E, F be points on the side-lines (i.e., lines containing the sides) BC, CA, AB, respectively, of the triangle ABC, such that AD, BE, CF are concurrent at a point P. Then the midpoints of AD, BE, CF are collinear if and only if P coincides with a vertex of triangle ABC or lies on one of its side-lines.

Proof. If P coincides with a vertex, suppose $P = A$ without loss of generality. Then $E = F = A$, and D lies anywhere on the side-line BC; the midpoints of AD, BE, CF are collinear on a line parallel to BC.

If P is not a vertex, we use oblique coordinate axes AB and AC, with suitable units of measurement along the axes so that A, B, C have coordinates (0,0), (2,0), (0,2), respectively. Let P have coordinates (p,q). Then the coordinates of D, E, F are $(2p/(p+q),\ 2q/(p+q))$, $(0, 2q/(2-p))$, $(2p/(2-q), 0)$; we require $p+q \neq 0$, $2-p \neq 0$, $2-q \neq 0$, since otherwise at least one of D, E, F is undefined (for instance, AP is parallel to BC if $p+q=0$). The coordinates of the three midpoints are $(p/(p+q),\ q/(p+q))$, $(1, q/(2-p))$, $(p/(2-q), 1)$; these midpoints are collinear if and only if

$$\begin{vmatrix} p/(p+q) & q/(p+q) & 1 \\ 1 & q/(2-p) & 1 \\ p/(2-q) & 1 & 1 \end{vmatrix} = \frac{2pq(2-p-q)}{(p+q)(2-p)(2-q)} = 0.$$

This occurs if and only if $p = 0$ or $q = 0$ or $p + q = 2$, i.e., if and only if P lies on a side-line of the triangle (in which case two of the midpoints coincide).

Now the orthocentre of a triangle cannot lie on a side-line of the triangle unless it coincides with a vertex, i.e., unless the triangle is right-angled. Hence the midpoints of the altitudes are collinear if and only if the triangle is right-angled.

J. F. RIGBY
University College
Cardiff, Wales

Also solved as in solution I by Jordi Dou (Spain), Howard Eves, Syrous Marivani, Mike Molloy (student, Canada), Richard Parris, Cem Tezer (Turkey), and Michael Woltermann; as in the generalized solution II (but using the theorems of Ceva and Menelaus) by Cem Tezer (Turkey, second solution); using analytic geometry by S. F. Barger, Kenneth Bernstein, Ragnar Dybvik (Norway), Cornelius Groenewoud, Boulkhodra Hacene, L. Kuipers (Switzerland), Hubert J. Ludwig, Bill Olk (student), John Oman, Harry Sedinger, Robert S. Stacy (West Germany), John S. Sumner, Michael Vowe (Switzerland), Jihad Yamout (student), and Robert L. Young; using barycentric or similar coordinate systems by O. Bottema (The Netherlands), J. T. Groenman (The Netherlands, two solutions), J. C. Linders (The Netherlands), and the proposer; using conjugate complex coordinates by Howard Eves (second solution) and Stephanie Sloyan; and using vector analysis by Leonard D. Goldstone and Harry D. Ruderman.

A Case of (Euler's) Identity

1198. Prove the identity

$$\sum_{k=0}^{m-1} \csc^2\left(\frac{k\pi}{m}+\alpha\right) = m^2\csc^2(m\alpha),$$

provided that $m\alpha$ is not an integral multiple of π. [*Russell Euler*, *Northwest Missouri State University*.]

Solution I: The identity is an immediate consequence of the identity

$$m\cot(m\alpha) = \sum_{k=0}^{m-1} \cot\left(\frac{k\pi}{m}+\alpha\right), \tag{1}$$

from which it is obtained by differentiating with respect to α. We will give a proof of (1), taken from Leonhard Euler, *Introductio in Analysin Infinitorum*, §§249–250.

From $\cos(m\alpha) \pm i\sin(m\alpha) = (\cos\alpha \pm i\sin\alpha)^m$ we obtain

$$2\cos(m\alpha) = (\cos\alpha + i\sin\alpha)^m + (\cos\alpha - i\sin\alpha)^m$$

and

$$2i\sin(m\alpha) = (\cos\alpha + i\sin\alpha)^m - (\cos\alpha - i\sin\alpha)^m.$$

Division then yields

$$\frac{1}{i}\cot(m\alpha) = \frac{(\cot\alpha + i)^m + (\cot\alpha - i)^m}{(\cot\alpha + i)^m - (\cot\alpha - i)^m} = \frac{2\Sigma_k(-1)^k\binom{m}{2k}(\cot\alpha)^{m-2k}}{2i\Sigma_k(-1)^k\binom{m}{2k+1}(\cot\alpha)^{m-2k-1}},$$

that is,

$$\cot(m\alpha) = \frac{(\cot\alpha)^m - \binom{m}{2}(\cot\alpha)^{m-2} + \cdots}{m(\cot\alpha)^{m-1} - \binom{m}{3}(\cot\alpha)^{m-3} + \cdots},$$

or

$$x^m - mx^{m-1}\cot(m\alpha) - \binom{m}{2}x^{m-2} + \cdots = 0, \tag{2}$$

where $x = \cot\alpha$. Now (2) is also valid if we set $x = \cot(\alpha + k\pi/m)$ for any integer k, since $\cot(m\alpha + k\pi) = \cot(m\alpha)$. Hence the m roots of (2), considered as an equation in x, are $\cot(\alpha + k\pi/m)$ for $k = 0, 1, \ldots, m-1$. (Note that these roots are all different, for the arguments lie within one period of the cotangent function.) And now it is seen from (2) that the sum of these roots is equal to $m\cot(m\alpha)$, as was to be shown.

J. C. Linders
Eindhoven, The Netherlands

Solution II: Both sides of the identity represent periodic meromorphic functions of α having the same poles and the same principal parts: $(\alpha - \alpha_n)^{-2}$, where $\alpha_n = n\pi/m$. They therefore differ by a periodic entire function $f(\alpha)$. From the estimate

$$|\sin(x+iy)|^2 = (\sin^2 x)(\cosh^2 y) + (\cos^2 x)(\sinh^2 y) \geqslant \sinh^2 y$$

we see that $f(\alpha) \to 0$ uniformly in α' as $\alpha'' \to \pm\infty$, where α' and α'' are the real and imaginary parts of α. Therefore f is an everywhere bounded entire function, hence a constant, by Liouville's theorem, hence zero (this having already been found as its limiting value as $\alpha'' \to \pm\infty$).

William A. Newcomb
Lawrence Livermore National Laboratory

Also solved, or references provided, by Kenneth Bernstein, Adanır Bilyer (Turkey), David Callan, Antonio Córdova Y. (student, Chile), Chico Problem Group, Boulkhodra Hacene (student), H. Kappus (Switzerland), L. Kuipers (Switzerland), Sai Chong Kwok, Syrous Marivani, Vania D. Mascioni (student, Switzerland), Jerry Metzger, Roger B. Nelsen, David Paget (Australia), Richard Parris, Bob Prielipp, Robert E. Shafer, J. M. Stark, Michael Vowe (Switzerland), Robert J. Wagner (two solutions), Gordon Williams, and the proposer.

Most solvers quoted or derived the formula

$$\sin(m\alpha) = 2^{m-1}\prod_{k=0}^{m-1}\sin\left(\alpha+\frac{k\pi}{m}\right)$$

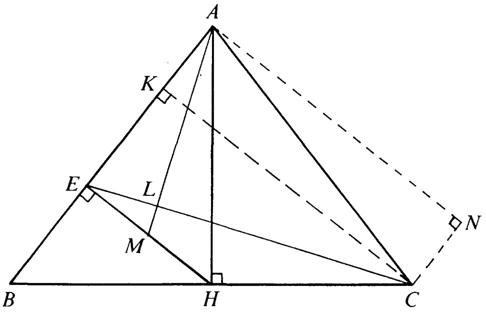
(L. Euler, *op. cit.*, §240) and then used logarithmic differentiation to obtain (1). *Shafer* and *Williams* used the explicit formula related to Solution II,

$$\csc^2\theta = \sum_{n=-\infty}^{\infty}\frac{1}{(\theta-n\pi)^2}$$

(Euler, §174). All of these formulas can be found in Bromwich, *An Introduction to the Theory of Infinite Series*, second edition, ch. IX, and some of them can be found in various compilations of formulas, such as Hansen, Jolley, and Gradshtein & Ryzhik. *Wagner's* second solution used properties of the gamma function. *Prielipp* referred to the similar *Monthly* problem 5486, vol. 75 (1968), pp. 421–422, and to *Fibonacci Quarterly* problem H-349, vol. 22 (1984), pp. 190–191.

Perpendicular Lines in an Isosceles Triangle — September 1984

1199. In the isosceles triangle ABC, with $AB = AC$, let H be the foot of the altitude from A, let E be the foot of the perpendicular from H to AB, and let M be the midpoint of EH. Show that $AM \perp EC$. [*Aristomenis Siskakis, University of Illinois.*]



Solutions I *and* II: I. Let CK be the altitude from C. In the similar right triangles EHA and KBC, the corresponding medians AM and CE make equal angles with the hypotenuses HA and BC. Let these medians intersect at L. Then the quadrangle $HCAL$ is cyclic. Hence $\angle ALC = \angle AHC = 90°$.

II. We use harmonic pencils. Let CK be the altitude from C, and construct the rectangle $AKCN$. Since the segment HE, parallel to AN, is bisected by AM, we have $(AB, AH; AM, AN) = -1$. Similarly, since the segment BK, parallel to CN, is bisected by CE, we have $(CK, CB; CE, CN) = -1$. In the two harmonic pencils, three lines are perpendicular to corresponding lines. Hence the fourth lines, namely, AM and CE, are perpendicular.

HÜSEYIN DEMIR
Middle East Technical University
Ankara, Turkey

Also solved by sixty-two others (including the proposer and eight students), who submitted seventy-three solutions.

Joseph Konhauser located the problem in the *Monthly*, problem E1476, with three published solutions in v. 69 (1962), p. 233. Four other solvers of that problem forgot to mention the fact when submitting solutions to this problem. *P. J. Pedler* (Australia) and *J. H. Webb* (South Africa) found the problem in Loren C. Larson, *Problem*

Solving Through Problems, p. 27, and *Geoffrey A. Kandall* found it in M. N. Aref & W. Wernick, *Problems and Solutions in Elementary Geometry*, p. 32, ex. 92. *O. Bottema* (The Netherlands) and *Webb* provided converses. (1) If ABC is any triangle, then $AM \perp EC$ if and only if $AB = AC$. (2) If E is any point on the line AB, then $AM \perp EC$ if and only if either $HE \perp AB$ or A is the midpoint of BE. (Other points are defined as in the problem statement.)

Comments

1014 (proposed May 1977; correction September 1977, p. 221; solution November 1979).

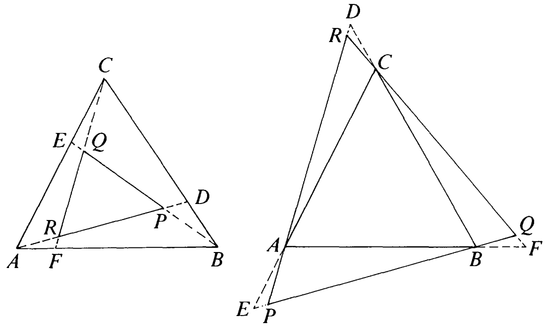
To avoid a counterexample (based on $BD + CE = BC$), the proposer and solver (*K.R.S. Sastry*, Ethiopia) suggested the following restatement.

Given a triangle ABC, the points D, E, F are chosen on the lines BC, CA, AB, respectively, all closer to (or all farther from) the vertices B, C, A than C, A, B, such that $AD = BE = CF$. The lines AD, BE, CF intersect to form ΔPQR, with P on AD and BE, Q on BE and CF, and R on CF and AD.

(a) Show that ΔPQR is equilateral if and only if ΔABC is.

(b) Express the area of ΔPQR in terms of the area of ΔABC.

Relabeled diagrams for the two cases are shown below.



1037 (proposed March 1978; solution November 1979).

In the third line of the solution of (a), insert "$+F_2$" after "F_{n+1}", insert an additional "$+F_3$" at the end of the next line, and replace the last "F_{n+2}" by "F_{n-2}" in the following line.

1041 (proposed May 1978; solution November 1979).

The Schwarz inequality used in the solution should have three terms. After (1) has been proved, equality will hold in (A) iff $a_1 = b_1 = c_1$ and $a_2 = b_2 = c_2$. Also, p_i should be defined as $2s_i$. The counterexample given for (2) is degenerate; a nondegenerate example is given by $a_i = b_i = 100c_i$.

1048 (proposed September 1978; solution November 1979).

The last display in the solution should read:

$$\sum_{k=N+1}^{\infty} \frac{n(a_{k+1}-a_k)}{a_k a_{k-1} \cdots a_{N+1}} < n\varepsilon \sum_{k=N+1}^{\infty} \frac{a_k}{a_k a_{k-1} \cdots a_{N+1}} < n\varepsilon \sum_{k=0}^{\infty} \frac{1}{(a_{N+1})^k} < 2n\varepsilon,$$

which is less than 1 if $\varepsilon < 1/(2n)$ and $N > N(\varepsilon)$.

1118* (proposed March 1981; comment May 1982).

Stanley Rabinowitz let his computer test Peter Ørno's conjecture that if P_n is the smallest set containing the nth prime p_n and containing every prime divisor of $p_n q + 1$ whenever it contains q, then P_n is a finite set. The computer supported the conjecture for $n \leqslant 100$, except for $p_{88} = 457$ and $p_{100} = 541$, where numbers were encountered which were too large (10^{12}) to be factored by the program. The smallest and largest sets were P_3 with 6 elements and P_{97} with 569 elements.

1156 (proposed November 1982; solution January 1984).

Gao Ling (China) has proved the following generalization. Let a, b, c, and F be the three sides and area of triangle ABC, and let a', b', c', and F' be the corresponding quantities for triangle $A'B'C'$. If $a \geqslant b \geqslant c$, $a' \leqslant b' \leqslant c'$, and r is any positive real number, then

$$a'^r(-a^r+b^r+c^r)+b'^r(a^r-b^r+c^r)+c'^r(a^r+b^r-c^r) \geqslant 3(16FF'/3)^{r/2}, \qquad (1)$$

with equality if and only if both triangles are equilateral. Ling also asks if (1) holds even without the restrictions $a \geqslant b \geqslant c$ and $a' \leqslant b' \leqslant c'$.

1168 (proposed March 1983; solution March 1984).

Benny N. Cheng (student) notes that a "dual" theorem can be proved similarly. Let P be a variable point on side BC of triangle ABC. Segment AP meets the excircle opposite A in two points, Q and R, with R being farther from A. Then the ratio AR/AP is a maximum when P is the point of contact of the incircle with side BC.

Answers

Solutions to the Quickies on p. 238.

Q698. Let C be the equator of a sphere S. Clearly there exists a circle C' on S, as close as we wish to C, such that $C' \cap C = \{p, q\}$. The orthogonal projection of C' into the equatorial plane is the desired ellipse E.

Q699. We prove that $n+1$ divides $(2n)!/(n!)^2 = \binom{2n}{n}$. Indeed, $(2n+1)\binom{2n}{n} = (n+1)\binom{2n+1}{n+1}$, and $n+1$ is relatively prime to $2n+1$. Alternatively, the assertion is clearly true if n is 0 or 1. Then for $n > 1$ we have

$$\binom{2n}{n} - \frac{(2n)!}{n!(n+1)!} = \binom{2n}{n}\left(1 - \frac{1}{n+1}\right) = \binom{2n}{n-1}.$$

Ed. note. The numbers $(2n)!/n!(n+1)!$ are called Catalan numbers, and more information and references can be found in this MAGAZINE, 57 (1984) 195–208 as well as in many books on combinatorics.

Q700. By division with remainder (or the formula for the sum of a finite geometric series), we have

$$\frac{1}{1-x} = \sum_{i=0}^{n} x^i + \frac{x^{n+1}}{1-x} \qquad \text{for } x \neq 1,$$

and by Maclaurin's theorem with integral remainder for $1/(1-x)$ we have

$$\frac{1}{1-x} = \sum_{i=0}^{n} x^i + (n+1)\int_0^x \frac{(x-t)^n}{(1-t)^{n+2}}\,dt \quad \text{for } x < 1.$$

Hence

$$(n+1)\int_0^x \frac{(x-t)^n}{(1-t)^{n+2}}\,dt = \frac{x^{n+1}}{1-x} \quad \text{for } x < 1.$$

REVIEWS

Paul J. Campbell, Editor
Beloit College

Assistant Editor: Eric S. Rosenthal, West Orange, NJ. Articles and books are selected for this section to call attention to interesting mathematical exposition that occurs outside the mainstream of the mathematics literature. Readers are invited to suggest items for review to the editors.

Senechal, Marjorie, and Galiulin, R. V., *An introduction to the theory of figures: the geometry of E. S. Fedorov*, Structural Topology No. 10 (1984) 5-22.

Summary and review of the neglected and never-translated 1885 book by Federov, in which he systematized geometrical crystallography and derived the five convex parallelohedra (polyhedra which fill space when arranged face-to-face in parallel position). (Note: Structural Topology is a little-known interdisciplinary journal on "geometry applied to problems of structure and morphology in architecture, design and engineering." All articles appear in both English and French. For further information: Le revue Topologie structurale, Dépt. de math., UQAM C.P. 8888, Succ. "A", Montréal, Québec, Canada, H3C 3P8).

Peterson, Ivars, *The fivefold way for crystals*, Science News 127 (23 March 1985) 188-189.

A new aluminum-manganese alloy exhibits five-fold (icosahedral) symmetry at the microlevel. Such structure cannot extend to the macro level in a strictly periodic fashion; instead, the crystal lattice must follow a three-dimensional version of Roger Penrose's aperiodic tiling patterns.

Kolata, Gina, *Prestidigitator of digits: Persi Diaconis has a magical way with statistics*, Science 85 (April 1985) 66-72.

Intriguing profile of statistician and magician Persi Diaconis, who -- among his other achievements -- can make a coin come up heads every time. (See also the earlier interview in Albers and Alexanderson's Mathematical People.)

Reid, Robert J. O., *The physics of boomerangs*, Mathematical Spectrum 17:2 (1984/85) 48-57.

Presents qualitative and quantitative analysis of boomerang motion, including precession. The physics and mathematics are simple (torque, moment of inertia, calculus) and the results satisfyingly surprising: the radius of the boomerang's orbit is independent of the spin velocity, of the length of the arms, the number of arms, and of the forward velocity! Varying the factors that do matter allows one to custom-design boomerangs for indoors and outdoors.

Golomb, Solomon, *The invincible primes*, The Sciences 25:2 (March-April 1985) 50-57.

From the definition of a prime to the best guesses on outstanding conjectures, Golomb covers the topic for the general reader.

Dreyfus, Hubert, and Dreyfus, Stuart, *Mindless machines: computers don't think like experts, and never will*, The Sciences 24:6 (November-December 1984) 18-22.

"Will expert systems duplicate, or even surpass, the achievements of human experts? ... [N]ow it seems all such attempts are doomed to failure because of a fundamental misunderstanding among computer scientists as to how human experts operate ... The traditional view holds that beginners begin with specific cases, and as they become more proficient they abstract and develop more and more sophisticated rules. But a better explanation seems to be that skills are acquired in just the opposite way, that one progresses from abstract rules to specific cases ... The expert is simply not following rules ... [but] recognizing thousands of special cases."

Tape, Walter, *The topology of mirages*, Scientific American 252:6 (June 1985) 120-129, 136.

"Topological methods are impressive when they are applied to mirages: they provide insight that does not require quantitative knowledge of the complex atmospheric conditions responsible ... Conversely, mirages can furnish graphic illustrations of topological ideas." Tape discusses transfer mappings, their degrees, and the odd-number theorem. Splendid photographs and graphics.

Devlin, Keith, *The golden age of mathematics*, New Scientist 106 (18 April 1985) 30-33.

Keith Devlin (Lancaster) cites the three major mathematical accomplishments of 1984: the disproof of the Mertens conjecture (whose truth would have implied the Riemann Hypothesis); the proof of the Bieberbach conjecture (by Louis de Branges); and finally -- "the real shocker of the year" -- the proof of the Riemann Hypothesis by Hideya Matsumoto (Paris). HOWEVER, Devlin warns that "many mathematicians remain skeptical about Matsumoto's claim." Among them is Matsumoto himself, who two months before Devlin's article had acknowledged a gap in his argument. The Riemann Hypothesis remains unproven; Devlin should have been more cautious (as were, e.g., mathematical journals and the Los Angeles Times).

Knight, Gordon, *The geometry of Maori art -- weaving patterns*, New Zealand Mathematics Magazine 21:3 (1984) 80-86.

All 12 symmetry groups for weaving patterns find instances in Maori art. (In the article the author notes he was unable to find an example of p4; but in a subsequent communication to B. Grünbaum, he enclosed an example of a Kete pattern, Whakatane Museum Collection #244, depicted as Pattern 4 in Mick Prendergast, Rarange Whakairo, Maori Plaiting Patterns, Cormandel Pr., 1984.)

David, Edward E., Jr., *The federal support of mathematics*, Scientific American 252:5 (May 1985) 45-51, 144.

Version for a general audience of the so-called "David report," published earlier as Renewing U.S. Mathematics: Critical Resource for the Future, with extracts in Notices AMS. The main point is that federal research support for mathematics has declined by one-third over the past 15 years. David, President of the Exxon Research and Engineering Company, originally proposed increasing funding from $78M to $180M over five years; but the federal budgets for 1985 and for 1986 continue the "pattern of undernourishment."

Cockcroft, Wilfred, *Does mathematics still count?*, New Scientist 106 (9 May 1985) 28-30.

Three years ago in England a national inquiry was undertaken into the teaching of mathematics. The current progress report speaks little about curriculum, but it notes that "it is resources, not rhetoric, that count." New resources have been devoted to in-service training of mathematics teachers, roughly $3.5 million per year -- this, in a country with one-fourth the population of the U.S.

Olivastro, Dominic, *Object lessons: in pursuit of pi,* The Sciences 25:3 (May-June 1985) 58-60.

This first column in a new series recounts some history and puzzles about pi.

Batty, Michael, *Fractals -- geometry between dimensions*, New Scientist 106 (4 April 1985) 31-35.

Yet another popular introduction to fractals, featuring computer-generated pictures produced on the BBC Micro. The author quotes physicist John A. Wheeler: "It is possible to believe that no one will be considered scientifically literature tomorrow who is not ... familiar with fractals."

Albers, Donald J., and Alexanderson, G. L. (eds.), Mathematical People: Profiles and Interviews, Birkhäuser, 1985; xviii + 372 pp, $24.95.

For the past few years the (Two-Year) College Mathematics Journal has featured interviews with well-known mathematicians. All those interviews (15 of them) are collected here, together with a good stock of unpublished interviews and profiles. New additions are D. Blackwell, P. Diaconis, R. Graham, P. Hilton, B. Mandelbrot, M. Rees, R. Smullyan, and O. Taussky-Todd, plus a reminiscence of S. Lefschetz. The book is a "must" for every mathematics department commons room, introducing students to the variety of "real people" who make the mathematical world so interesting.

Rucker, Rudy, The 4th Dimension: Toward a Geometry of Higher Reality, Houghton Mifflin, 1984; xi + 228 pp, $17.95.

Thoroughly engrossing adventure into "dimensions beyond our material plane." There is not one fourth dimension, but many, says Rucker; in mind-dazzling fashion, he uses geometry as a springboard to metaphysics. Far more substantial than most other books on the fourth dimension, this one grapples with many of the big ideas of physics, philosophy, and mysticism, in a journey both entertaining and provocative.

Gardner, Martin (ed.), The Sacred Beetle and Other Great Essays in Science, Prometheus, 1984; xv + 427 pp.

Expanded and revised edition of Great Essays in Science (1957). Gardner's goal is not to teach science but "to spread before the reader ... a sumptuous feast of great writing -- absorbing, thought-disturbing pieces that have something important to say about science and say it forcibly and well." Contributors vary from Robert Louis Stevenson to J. Robert Oppenheimer, from Francis Bacon to Carl Sagan.

Macdonald, Ian D., Alex: a 1-Act Play about Euclid, Polygonal, 1984; 32 pp, $3.95.

"Since almost nothing is known of the life of Euclid, it is a very suitable subject for a play ... One may assume that he was not so easily permitted to get on with the writing of "The Elements" but rather had his share of problems and distractions from government officials, well-meaning friends and relatives, a young lady ..." Here's a "mathematical" play for a high-school audience; some of the British dialect will need changes (or translation) for Americans.

Hirsch, Christian R., and Zweng, Marilyn J., The Secondary School Mathematics Curriculum, 1985 Yearbook, NCTM, 1985; vi + 250 pp, $14.50.

Rich compendium of new directions for high-school mathematics. As Z. Usiskin remarks in the introductory essay, it is time for another revolution; and he points to the fact that dramatic change can take place quickly, if the public and teachers have the resolve.

TEAM Modules: Hours of Daylight, Highway Slope Design, and Aircraft Sidestep Maneuver. Videotapes, each with booklet containing student manual and instructor's guide; plus diskette with Apple Basic programs for the first and third modules. MAA, 1984; free loan (contact national office or Section TEAM coordinator.

"TEAM" stands for "Teaching Experiential Applied Mathematics," and NSF funding has allowed the MAA to produce these splendid modules on how mathematical modeling is done. Each module focuses on a real problem and features two briefings by the person who had to solve it: one on the problem itself and its setting, and the second on a proposed solution. The student manuals summarize relevant facts and formulas, and students should build their own models before hearing the one proposed. The mathematics needed varies by problem; calculus and junior-senior maturity are presumed. The overall enterprise is nicely conceived and mostly well-done. John Jobe (Oklahoma State) serves as interviewer, and asks probing questions that are just the ones that will be on students' minds. Two quibbles: the computer programs are unacceptable (no comments, no structure), and the videotapes suffer from unlively editing and so are in part simply filmed lectures (i.e., they suffer from public TV syndrome). These three modules all use continuous mathematics and deterministic models; some subsequent modules should feature discrete mathematics and stochastic modeling.

Brookshear, J. Glenn, Computer Science: An Overview, Benjamin/Cummings, 1985; xiii + 448 pp. (P).

Books for introductory computer science courses tend to either emphasize programming in a particular language, or else focus on "computer literacy" plus a little programming. Brookshear's book achieves the remarkable feat of successfully addressing to both programming and nonprogramming audiences a comprehensive overview of the scope and key issues of computer science and can be used either as the main text for a nonsuperficial literacy course or as a valuable supplement to a programming course. However, an important additional ingredient at the introductory level is consideration of the social, ethical, and economic aspects of computer use. Fortunately, other supplementary texts address precisely these matters.

Wetzel, Gregory F., and Bulgren, William G., The Algorithmic Process: An Introduction to Problem Solving, SRA, 1985; ix + 292 pp.

Developing algorithms and coding them in a particular syntax are the two main types of problems in an intro CS course; of the two, the latter largely exercises already-acquired low-level pattern-matching skills, while the former requires ingenuity, creativity, and organization of thinking. This text asserts that understanding algorithm development is the more important, and it attempts to separate that topic from language syntax by relegating the latter to another text. Incongruously, however -- and probably at the publisher's insistence -- the last fourth of the book is occupied with (under-documented) programs translating the text's algorithms into Pascal and Fortran-77.

Mitchell, Charlie, Math Anxiety: What It Is and What to Do About It, Action Pr. (Box 25738, Tempe, AZ 85282), 1984; x + 96 pp, $7.95 (P).

"Math anxiety is a classically conditioned response that has been generalized and is reinforced by avoidance ..." This book analyzes math anxiety at the psychological and behavior levels: "The most effective place to begin your attack on math anxiety is at the physical symptom level because it is the physical state that disables the thinking process." The book teaches relaxation, desensitization, cognitive restructuring, and effective math studying techniques. It includes an outline of a one-credit-hour course on math anxiety and avoidance (as taught at Mesa Community College) and an extensive bibliography.

NEWS & LETTERS

ALLENDOERFER, FORD AND PÓLYA 1984 AWARDS

At the business meeting on August 13, 1985, in Laramie, Wyoming, the MAA honored eight authors for excellence in expository writing. The awards, in the amount of $200 each, were for articles published in 1984 in *Mathematics Magazine*, the *American Mathematical Monthly*, and the *College Mathematics Journal*.

Recipients of the Carl B. Allendoerfer award were:

Philip D. Straffin, Jr. and Bernard Grofman, "Parliamentary Coalitions: A Tour of Models", *Math. Magazine*, 57 (1984) 259-274.

Frederick S. Gass, "Constructive Ordinal Notation Systems", *Math. Magazine*, 57 (1984) 131-141.

Recipients of the Lester R. Ford award were:

Donald G. Saari and John B. Urenko, "Newton's Method, Circle Maps, and Chaotic Motion", *Amer. Math. Monthly*, 91 (1984) 3-17.

John D. Dixon, "Factorization and Primality Tests", *Amer. Math. Monthly*, 19 (1984) 333-352.

Recipients of the George Pólya award were:

Anthony Barcellos, "The Fractal Geometry of Mandelbrot", *CMJ*, 15 (1984) 98-114.

Kay W. Dundas, "To Build a Better Box", *CMJ*, 15 (1984) 30-36.

NCTM SEEKS AUTHORS

The Educational Materials Committee of the National Council of Teachers of Mathematics invites manuscripts for the 1988 NCTM Yearbook, "Algebraic Concepts in the Curriculum, K-12", edited by Arthur Coxford of the University of Michigan. The Yearbook Advisory Panel seeks papers addressing important issues in the teaching of algebra at all levels in the curriculum, as well as papers reporting proven classroom practices in teaching specific algebraic topics. Guidelines for the preparation of manuscripts are available from the General Editor, Albert P. Shulte, Oakland Schools, 2100 Pontiac Lake Road, Pontiac, MI 48054.

MORE ON "ONLY" CRITICAL POINT

Another reference to be added to the trio of notes concerning lonely critical points for functions of two variables (this MAGAZINE, May 1985, pp. 146-150) is B. Calvert and M.K. Vamanamurty, "Local and global extrema for functions of several variables", J. Australian Math. Soc., ser. A, 29(1980) 362-368.

-- Ira Rosenholtz
Univ. of Wyoming

SUMS OF INTEGER POWERS -- A GEOMETRIC APPROACH

C. Kelly ("An Algorithm for Sums of Integer Powers", this MAGAZINE, Nov. 1984, 296-297) develops a formula for the sum of k-th powers of the first n integers, $S_k(n)$:

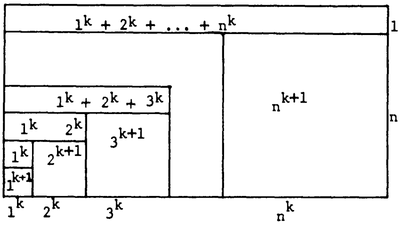
$$(k+1)S_k(n) = (n+1)^{k+1} - 1 - \sum_{i=0}^{k-1} \binom{k+1}{i} S_i(n).$$

A second approach to this problem is to partition rectangles in such a way that the constituent areas give rise to the desired sums. Desbrow, [1], used such a technique to evaluate $S_1(n)$, $S_2(n)$, and $S_3(n)$.

We would like to note the ingenious derivation of

$$(n+1)S_k(n) = S_{k+1}(n) + \sum_{p=1}^{n} S_k(p) \qquad (1)$$

by the Arab mathematician Al-Haitham (c. 965-1039). This derivation is given on p. 84 of [2], but seems not to be well known. Edwards sets this result in its historical context where Al-Haitham used these sums to extend some of Archimede's volume results.



From this diagram we can read off formula (1) which gives us the recursive formula:

$$s_{k+1}(n) = (n+1)s_k(n) - \sum_{p=1}^{n} s_k(p) \ .$$

References:

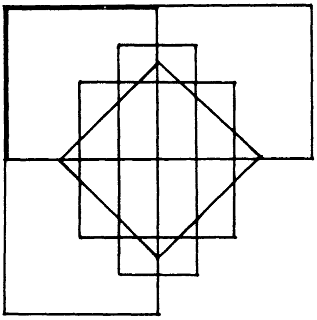
[1] D. Desbrow, Sums of Integer Powers, Mathematical Gazette, 66 (1982) 97-100.

[2] C. H. Edwards, Jr., The Historical Development of the Calculus, Springer-Verlag, 1979, 83-85.

-- Kent M. Neuerberg
Univ. of Calif., Davis

VENN DIAGRAM OF RECTANGLES

A.J. Schwenk's Venn diagram for five sets (this MAGAZINE, Nov. 1984, p. 297), while being artistically impressive, is difficult to reproduce. The diagram below consisting of five intersecting rectangles is easily drawn. (Heavy lines indicate intersection of edges of the rectangles.)



-- A.V. Boyd
Univ. of the Witwatersrand,
Johannesburg

RUBIK'S MAGAZINE

Readers will find more "impossible" art by our cover artist Tamás F. Farkas, in the eighth issue of a fascinating game magazine, *RUBIK'S logic & fantasy in space*, 4/83. Edited by mathematician-inventor Ernő Rubik, the quarterly publication is published in English, German, French, Russian and Hungarian. Mailing address is:

Editorial Office of RUBIK'S
H 1906 BUDAPEST, P.O.B. 223
HUNGARY .

MATH OLYMPIAD WINNERS

Eight U.S. students have earned Olympiad medals in the final round of a three-stage mathematics competition involving 380,000 high school students. The 64th USAMO competitors were the top performers in the American High School Mathematics Examination (AHSME) and the American Invitational Mathematics Examination (AIME) which were held in high schools throughout the United States and Canada in February and March 1985. The final round in this competition was the Fourteenth USA Mathematical Olympiad (USAMO), a challenging examination designed to test ingenuity as well as mathematical background.

The eight USAMO winners are:

*	Joseph G. Keane,	Pittsburgh, PA
*	Waldemar P. Horwat,	Hoffman Estates, IL
	John A. Overdeck,	Columbia, MD
	Yeh Ching-Tung,	Saratoga, CA
*	Bjorn M. Poonen,	Winchester, MA
	Zinkoo Han,	Brooklyn, NY
*	Jeremy A. Kahn,	New York, NY
	John P. Dalbec,	Youngstown, OH

The winners were honored on June 4 in Washington, D.C. at an awards ceremony which was held at the National Academy of Sciences and the U.S. Department of State. The Samuel Greitzer-Murray Klamkin award for attaining the highest grade on the USAMO was won by Joseph C. Keane.

Following the ceremony, the eight winners and sixteen other students who did well in the USAMO participated in an intensive three-week seminar at the U.S. Military Academy at West Point. From these participants, a U.S. team of 6 students was chosen to compete in the 1985 International Mathematical Olympiad (IMO), held in Helsinki, Finland in July. The four winners whose names are starred above, together with David Grabiner, of Claremont, CA, and David Moews, of Willimantic, CT, comprised the U.S. team.

The U.S. team placed second in the IMO. Top-ranked teams and their scores were:

Romania	201
U.S.A.	180
Hungary	168
Bulgaria	165
Vietnam	144
U.S.S.R.	140

All U.S. team members won prizes for individual scores; Horwat and Kahn won first prizes and all others second prizes.

Exam questions for both the USAMO and Canadian Mathematical Olympiad follow; match your wits with the high school student winners! (IMO problems will appear in our next issue.)

14TH U.S.A. MATH OLYMPIAD APRIL 23, 1985

1. Determine whether or not there are any positive integral solutions of the simultaneous Diophantine equations

$$x_1^2 + x_2^2 + \dots + x_{1985}^2 = y^3 ,$$

$$x_1^3 + x_2^3 + \dots + x_{1985}^3 = z^2 ,$$

such that $x_i \neq x_j$ for all $i \neq j$.

2. Determine each real root of

$x^4 - (2\cdot 10^{10}+1)x^2 - x + 10^{20} + 10^{10} - 1 = 0$

correct to four decimal places.

3. Let A, B, C and D denote any four points in space such that at most one of the distances AB, AC, AD, BC, BD and CD is greater than 1. Determine the maximum value of the sum of the six distances.

4. There are n people at a party. Prove that there are two people such that, of the remaining n-2 people, there are at least $\lfloor n/2 \rfloor$-1 of them, each of whom either knows both or else knows neither of the two. Assume that "knowing" is a symmetric relation, and that $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x.

5. Let $a_1, a_2, a_3, \dots$ be a non-decreasing sequence of positive integers. For $m \geq 1$, define $b_m = \min\{n: a_n \geq m\}$, that is, b_m is the minimum value of n such that $a_n \geq m$. If $a_{19} = 85$, determine the maximum value of

$a_1 + a_2 + \dots + a_{19} + b_1 + b_2 + \dots + b_{85}$.

17TH CANADIAN MATH OLYMPIAD MAY 1, 1985

1. The lengths of the sides of a triangle are 6, 8 and 10 units. Prove that there is exactly one straight line which simultaneously bisects the area and perimeter of the triangle.

2. Prove or disprove that there exists an integer which is doubled when the initial digit is transferred to the end.

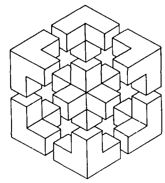
3. Let P_1 and P_2 be regular polygons of 1985 sides and perimeters x and y respectively. Each side of P_1 is tangent to a given circle of circumference c and this circle passes through each vertex of P_2. Prove $x + y \geq 2c$. (You may assume that $\tan\theta \geq \theta$ for $0 \leq \theta < \frac{\pi}{2}$).

4. Prove that 2^{n-1} divides $n!$ if and only if $n = 2^{k-1}$ for some positive integer k.

5. Let $1 < x_1 < 2$ and, for $n = 1,2,\dots,$ define

$$x_{n+1} = 1 + x_n - \frac{1}{2}x_n^2$$

Prove that, for $n \geq 3$, $|x_n - \sqrt{2}| < 2^{-n}$.



The Mathematical Association of America

CORPORATE MEMBERS

The Mathematical Association of America gratefully acknowledges the interest and support of the following:

CORPORATE MEMBERS

Addison-Wesley Publishing Company, Inc.
American Mathematical Society, Inc.
Birkhäuser Boston, Inc.
Digital Equipment Corporation
International Business Machines Corporation
Macmillan Publishing Company
Marcel Dekker, Inc.
The Mitre Corporation
Pergamon Press, Inc.
Prindle Weber and Schmidt, Inc.
Saunders College Publishing, Inc.
Scott Foresman & Company, Inc.
Springer-Verlag New York, Inc.
John Wiley & Sons, Inc.

CONTRIBUTING CORPORATE MEMBER

Bell Laboratories, Inc.

MAA STUDIES IN MATHEMATICS

Studies in Numerical Analysis

MAA Studies in Mathematics #24
Gene H. Golub, Editor
415 pp. Hardbound.
List: $42.00 MAA Member: $31.00

This volume is a collection of papers describing the wide range of research activity in numerical analysis. The articles describe solutions to a variety of problems using many different kinds of computational tools. Some of the computations require nothing more than a hand held calculator: others require the most modern computer. While the papers do not cover all of the problems that arise in numerical analysis, they do offer an enticing and informative sample.

Numerical analysis has a long tradition within mathematics and science, beginning with the work of the early astronomers who needed numerical procedures to help them solve complex problems. The subject has grown and developed many branches, but it has not become compartmentalized. Solving problems using numerical techniques often requires an understanding of several of the branches. This fact is reflected in the papers in this collection.

Computational devices have expanded tremendously over the years, and the papers in this volume present the different techniques needed for and made possible by several of these computational devices.

Table of Contents

Order From:
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

Carus Mathematical Monograph #21

From Error Correcting Codes through Sphere Packings to Simple Groups,

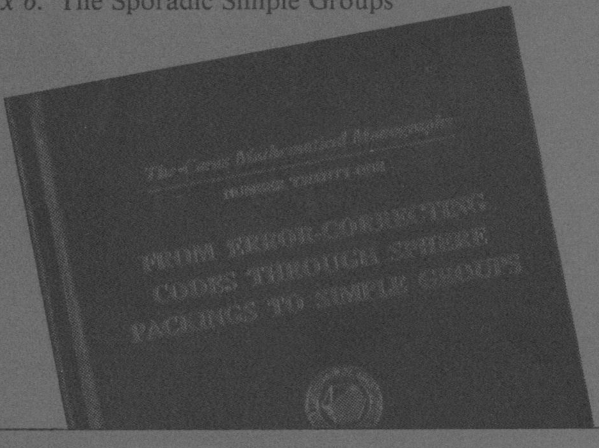
by Thomas M. Thompson
224 pp. Hardbound
List: $24.00   MAA Member: $18.50

Two of the most fascinating problems to challenge mathematicians in recent years concern the construction of data transmission codes that can correct errors introduced by static and the search for efficient ways to pack ping-pong balls into a box. Can one design the best error-correcting codes? Can one find the most efficient sphere packing?

Therein lies a fascinating story which is told with great skill and clarity in this important addition to the Carus Mathematical Monograph series. The author has packed (sic) into 175 pages all of the basic mathematical ideas of this saga woven into a gripping historical account of the journey from error-correcting codes to sphere packings to simple groups.

Table of Contents

Chapter 1. The Origin of Error-Correcting Codes
Chapter 2. From Coding to Sphere Packing
Chapter 3. From Sphere Packing to New Simple Groups
Appendix 1. Densest Known Sphere Packings
Appendix 2. Further Properties of the (12,24) Golay Code and the Related Steiner System S (5,8,24)
Appendix 3. A Calculation of the Number of Spheres with Centers in A_2 adjacent to one, two, three, and four adjacent spheres with centers in A_2.
Appendix 4. The Mathieu Group M_{24} and the order of M_{22}
Appendix 5. The Proof of Lemma 3.3
Appendix 6. The Sporadic Simple Groups


The Carus Mathematical Monographs
NUMBER TWENTY-ONE
FROM ERROR-CORRECTING CODES THROUGH SPHERE PACKINGS TO SIMPLE GROUPS

Order From:
The Mathematical Association of America
1529 Eighteenth St, NW
Washington, DC 20036

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, DC 20036

MATHEMATICS MAGAZINE VOL. 58, NO. 4, SEPTEMBER 1985