Vol. 69, No. 4 October 1996

MATHEMATICS MAGAZINE

- The Baseball-Card Collector's Query
- Curves and Surfaces with Reflection Properties

An Official Publication of The MATHEMATICAL ASSOCIATION OF AMERICA

EDITORIAL POLICY

Mathematics Magazine aims to provide lively and appealing mathematical exposition. This is not a research journal and, in general, the terse style appropriate for such a journal (lemma-theorem-proof-corollary) is not appropriate for an article for the *Magazine*. Articles should include examples, applications, historical background, and illustrations, where appropriate. They should be attractive and accessible to undergraduates and would, ideally, be helpful in supplementing undergraduate courses or in stimulating student investigations. Manuscripts on history are especially welcome, as are those showing relationships between various branches of mathematics and between mathematics and other disciplines.

A more detailed statement of author guidelines appears in this *Magazine*, Vol. 69, pp. 78–79, and is available from the Editor. Manuscripts to be submitted should not be concurrently submitted to, accepted for publication by, or published by another journal or publisher.

Send new manuscripts to Paul Zorn, Editor, Department of Mathematics, St. Olaf College, 1520 St. Olaf Avenue, Northfield, MN 55057-1098. Manuscripts should be laser-printed, with wide line-spacing, and prepared in a style consistent with the format of *Mathematics Magazine*. Authors should submit three copies and keep one copy. In addition, authors should supply the full five-symbol Mathematics Subject Classification number, as described in *Mathematical Reviews*, 1980 and later. Copies of figures should be supplied on separate sheets, both with and without lettering added.

Cover illustration: Mathematical trading cards, by Greg Shultz, a senior art major at St. Olaf College, Northfield, Minnesota.

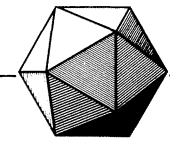
AUTHORS

James Sandefur *received his bachelor's degree* from Vanderbilt University in 1969, his M.A. from the University of Denver in 1971, and his Ph.D. from Tulane University in 1974. He is a professor at Georgetown University, where he has been since 1974. His research in differential equations led him to search for engaging dynamic applications of mathematics that were accessible to undergraduates; this culminated in his text *Discrete Dynamical Modeling*. Originally, the author thought that the baseball-card collector's query was such a simple dynamic application—until he awoke one night to the realization that there was apparently two different answers to the same question.

Daniel Drucker has an S.B. from MIT. He worked under Joseph A. Wolf at the University of California at Berkeley, receiving his Ph.D. in 1973 and joined the faculty of Wayne State University in 1975. Two years ago, he initiated a calculus sequence at WSU that uses progammable graphing calculators. His interests include differential geometry, Lie theory, linear algebra, and number theory, but he also plays the violin in a local orchestra. Thinking about curves with reflection properties on spheres led him to the approach used in the present article.

Phil Locke received his bachelor's degree from Bluffton (Ohio) College in 1959 and his doctorate from the University of New Hampshire in 1967. Since 1968 he has been at the University of Maine, where he is currently Assistant Chair. His mathematical interests run from the pedagogy of calculus reform to differental geometry. His favorite proofs are those that provide insight into *why* a theorem is true. His hobbies include playing traditional music on the fiddle.

Vol. 69, No. 4 October 1996



MATHEMATICS MAGAZINE

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, bimonthly except July/August.

The annual subscription price for the MATHEMATICS MAGAZINE to an individual member of the Association is $16 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $64. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% discount; and new members receive a 40% dues discount for the first two years of membership.) The nonmember/library subscription price is $68 per year.

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. 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.

Periodicals postage paid at Washington, D.C. and additional offices.

Postmaster: Send address changes to Mathematics Magazine, Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036-1385.

PRINTED IN THE UNITED STATES OF AMERICA

ARTICLES

The Baseball-Card Collector's Query

JAMES T. SANDEFUR
Georgetown University
Washington, DC 20057-0996

1. The Average Size of a Baseball-Card Collection

The following relates my attempt to solve a problem that was asked of a colleague by a baseball-card collector. In attempting to solve this problem, I used a number of techniques from undergraduate mathematics, including series, the exponential function, Newton's method, probability, statistics, simulation, and discrete dynamical systems. I was also reminded about the importance of carefully stating the problem.

Suppose that there exist baseball cards for n different baseball players. Assume for simplicity that each card is equally likely to be acquired each time a new card is purchased. One copy of each different card in a collection is put into the "original" pile. All duplicates, triplicates, etc. are put into the "duplicate" pile. The cost for obtaining the first "original" card is just the cost of that card. Once the collector has acquired a large number of "original" cards, the expected cost for obtaining one more "original" card will be relatively large, since most cards acquired will be duplicates.

There are many questions that could be asked relating to the cost of a collection of baseball cards. In this section, we will investigate one particular question, the baseball-card collector's question (BCQ):

> *What is the average number of cards, a_n, in the "original" pile when the two piles are equal for the first time?*

One reason for studying this question is that at this point, the collector has had to purchase two cards for each original obtained. We could just as well have asked at what point the "original" pile is half or a third of the size of the "duplicate" pile. Another reason for looking at this question is that a baseball-card collector actually asked a colleague this question.

It is clear that this question has an answer. When the collector acquires his or her first card, it is clearly put into the "original" pile, so the "original" pile is larger than the "duplicate" pile. But once $2n + 1$ cards have been acquired, the "duplicate" pile must be larger since the "original" pile can have at most n cards in it. At some point, the two piles must be the same size.

Let $p_n(j)$ be the probability that the two stacks are equal for the first time with j cards in each. Let's do a few simple calculations. Suppose there are only $n = 2$ distinct baseball cards. The first card collected goes into the "original" pile. There is a 50 percent chance the next card matches the first, so $p_2(1) = 0.5$. Since there are only 2 distinct cards, the other half of the time the two piles will be of equal size for the first time when there are two cards in each; that is, $p_2(2) = 0.5$. The average size of the "original" pile when the two piles are equal for the first time is then $a_2 = 1p_2(1) + 2p_2(2) = 1.5$.

Suppose $n = 3$. Simple calculations given that $p_3(1) = 1/3$, $p_3(2) = 8/27$, and $p_3(3) = 10/27$. Therefore, the average number of cards in the "original" stack when the piles are equal for the first time is

$$a_3 = 1p_3(1) + 2p_3(2) + 3p_3(3) = (1/3) + 2(8/27) + 3(10/27) = 55/27 = 2.037.$$

For $n = 4$, 5, and 6 the average number of cards in the "original" pile when they are equal for the first time is $a_4 = 2.611328$, $a_5 = 3.219725$, and $a_6 = 3.858451$, respectively. In general,

$$a_n = \sum_{j=1}^{n} jp_n(j).$$

To compute a_6, it is necessary to compute $p_6(1), \ldots, p_6(6)$. To find, say $p_6(4)$, it is necessary to find the probability of each of the ways that 4 originals can be obtained before getting 4 duplicates, with 6 different cards being possible. There are 5 ways of getting 4 originals and 4 duplicates:

$$oooodddd, \quad ooododdd, \quad oooddodd, \quad oodooddd, \quad \text{and} \quad oodododd$$

where o represents an original and d a duplicate. (Remember that $p_6(4)$ implies that the first time the number of originals equals the number of duplicates is when 4 originals have been obtained. So the probabilities of *ooddoodd* and other similar combinations do not need to be computed.) To compute the probability of one of these, say $p(oooddodd)$, we count the ways of getting cards in this order, then divide by 6^8;

$$p(oooddodd) = \frac{6 \cdot 5 \cdot 4 \cdot 3 \cdot 3 \cdot 3 \cdot 4 \cdot 4}{6^8}.$$

Another approach to studying BCQ is to look at the ratio, $r_n = a_n/n$, of the number of cards in the "original" pile of the number of possible cards, n, when the two piles are equal for the first time. In this case, the average ratio of the cards in the "original" pile to n when the two piles are first equal is $r_2 = 1.5/2 = 0.75$, $r_3 = 2.037/3 = 0.679$, $r_4 = 0.652832$, $r_5 = 0.643945$, and $r_6 = 0.643075$. One reason for looking at the ratio is that the answers to BCQ for different n-values can be compared.

2. The Expected Number of Originals in a Collection of Size k

In trying to solve BCQ, I made the mistake of trying to solve the general problem before working the special cases just discussed. My approach was to define $e_n(k)$ as *the expected number of cards in the "original" pile when k cards have been acquired, and there are n distinct cards.* Clearly, $e_n(0) = 0$ and $e_n(1) = 1$. Suppose that $e_n(k)$ has been computed. Then $e_n(k+1)$ is $e_n(k)$ plus the probability that the $k+1$th card is different from the previous k cards. Since $e_n(k)$ is the expected number of different cards currently in the collection, the probability the next card will be different from the first k cards is $(n - e_n(k))/n$, that is, the number of cards different from the ones owned divided by the number of different cards. This gives the first-order affine dynamical system

$$e_n(k+1) = e_n(k) + (n - e_n(k))/n = (1 - 1/n)e_n(k) + 1.$$

Since we know that $e_n(1) = 1$, this dynamical system gives that $e_n(2) = (1 - 1/n) + 1$, $e_n(3) = (1 - 1/n)^2 + (1 - 1/n) + 1$, and so forth. In general, $e_n(k)$ is given by the

finite geometric series

$$e_n(k) = (1-1/n)^{k-1} + (1-1/n)^{k-2} + \cdots + (1-1/n) + 1.$$

This finite geometric series can be rewritten as

$$e_n(k) = -n(1-1/n)^k + n.$$

Let $f_n(k)$ equal the expected number of cards in the "duplicate" stack when k cards have been obtained. Clearly, $f_n(k) = k - e_n(k) = k - n + n(1-1/n)^k$. The stacks are the same size when $e_n(k) = f_n(k)$.

When $n = 2$, the formula for $e_n(k)$ gives that $e_2(1) = 1$ and $f_2(1) = 0$; $e_2(2) = 1.5$ and $f_2(2) = 0.5$; $e_2(3) = 1.75$ and $f_2(3) = 1.25$; and $e_2(4) = 1.875$ and $f_2(4) = 2.125$. My second mistake was to assume the two piles would be the same size for a collection containing $k = k_n$ cards where k_n satisfies the equation

$$e_n(k) = f_n(k) \quad \text{or} \quad -2n(1-1/n)^2 + 2n - k = 0$$

after simplification. For $n = 2$, the "original" and "duplicate" stacks are never expected to be the same size, but once 4 cards have been obtained, the "duplicate" stack is expected to be bigger than the "original" stack. Thus the solution for $e_n(k)$ can be used to find the minimum collection size for which the "original" stack is expected to be at most the size of the "duplicate" stack; that is, the smallest integer k for which

$$-2n(1-1/n)^k + 2n - k \le 0.$$

The approximate solution, k_n, can be found using Newton's method. The actual solution is then $\lceil k_n \rceil$, the smallest integer greater than or equal to k_n. If, say, $n = 1000$, then

$$-2000(1-1/1000)^k + 2000 - k = 0$$

gives $k_{1000} = 1594.17$, and therefore, $\lceil k_n \rceil = 1595$.

Although I was attempting to solve BCQ, $\lceil k_n \rceil$ is actually the answer to an alternative question (AQ):

> *How many cards must I collect before the "original" pile is expected to be smaller than the "duplicate" pile?*

To help distinguish between BCQ and AQ, define $p_n(i,j)$ as the probability that the "original" pile has i cards and the "duplicate" pile has j cards when $k = i + j$ cards have been obtained. This gives the collection of probabilities

$$\begin{array}{cccccc} p_n(0,0) & p_n(0,1) & p_n(0,2) & p_n(0,3) & \cdots \\ p_n(1,0) & p_n(1,1) & p_n(1,2) & p_n(1,3) & \cdots \\ p_n(2,0) & p_n(2,1) & p_n(2,2) & p_n(2,3) & \cdots \\ p_n(3,0) & p_n(3,1) & p_n(3,2) & p_n(3,3) & \cdots \\ \vdots & \ddots & & & \end{array}$$

Clearly, $p_n(0,j) = 0$ for all j, and $p_n(i,j) = 0$ if $i > n$.

Consider the sum

$$\sum_{j=0}^{n} j p_n(j,j),$$

which gives the expected number of cards in the first pile when the two piles are equal. This resembles the solution to BCQ, but $p_n(j,j) \ge p_n(j)$ because the two piles

may be equal at j cards each while also having been equal with fewer cards each. On the other hand,

$$e_n(k) = \sum_{j=0}^{k} jp_n(j, k-j),$$

uses the probabilities along "reflected" diagonals, from upper right to lower left. Solving AQ means finding the first such diagonal that gives $e_n(k) \le k/2$. Differences between BCQ and AQ are reflected in the different probabilities used in answering each question.

The interested reader is encouraged to show that the dynamical system

$$e_n(k+1) = (1 - 1/n)e_n(k) + 1$$

is satisfied by $e_n(k) = \sum_{j=0}^{k} jp_n(j, k-j)$. This can be done using the relations

$$p_n(i,j) = \frac{(n-i+1)p_n(i-1,j) + ip_n(i,j-1)}{n} \text{ if } i > 1 \text{ and } j > 0,$$

$$p_n(1,j) = \frac{p_n(1,j-1)}{n}, \; p_n(i,0) = \frac{(n-i+1)p_n(i-1,0)}{n}, \; \sum_{j=1}^{k} p_n(j, n-j) = 1.$$

3. The Fraction of the Cards in Each Pile

For BCQ, we studied $r_n = a_n/n$, the fraction of the n distinct cards in the "original" pile when the two piles were, on average, equal for the first time. For AQ, the two piles are *expected* to be the same size when the collection contains k_n cards. In this case, there are $k_n/2$ cards in each pile. As we did for BCQ, let's study the fraction of the n distinct cards expected to be in the "original" pile when there are k_n cards in the collection. Denote this fraction by $x_n = k_n/2n$, where k_n is the solution to

$$-2n(1 - 1/n)^k + 2n - k = 0.$$

It follows that x_n solves the equation

$$1 - (1 - 1/n)^{2nx} - x = 0.$$

As n tends to infinity, $(1 - 1/n)^n$ tends to e^{-1}, so as the number of possible cards increases, the ratio $x_n = k_n/2n$ approaches the solution to the equation

$$1 - e^{-2x} - x = 0.$$

The approximate solution $X = 0.796\,812\,13$ to eight decimal places, can be found using Newton's method, a graphing calculator, or a computer algebra system.

This implies that for large n, $x_n = k_n/2n \approx X$, or $k_n \approx 2nX$. This means, in context, that the two piles are expected to have the same number of cards when the collection contains about $1.6n$ cards. At this point, moreover, the collection will contain about 80 percent of the possible cards.

To see how fast x_n converges to 0.796812 as n increases, I used Newton's method to find roots of $1 - (1 - 1/n)^{2nx} - x = 0$ for several values of n. In particular, I found the following pairs (n, x_n):

(2, 0.9225)	(3, 0.8834)	(4, 0.8627)
(10, 0.8237)	(50, 0.8023)	(100, 0.7995)
(1000, 0.7971)	(10 000, 0.796 84)	(100 000, 0.796 815)
(1 000 000, 0.796 812 40)		

4. Comparing the Answers to the Two Questions

How does the answer to AQ relate to the answer to BCQ? Some relatively simple calculations reveal that BCQ and AQ have quite different answers, at least when n is relatively small. For BCQ, *the average of the ratio* of the "original" stack to n when the two stacks are equal for the first time was seen to be $r_n = a_n/n = 0.75$, when $n = 2$. This contrasts with $x_2 = k_2/2(2) = 0.9225$. More comparisons are:

$n =$	2	3	4	5	6
$r_n =$	0.75	0.679	0.652 832	0.643 945	0.643 075
$x_n =$	0.9225	0.8834	0.8627	0.849 91	0.841 277

The second list of ratios was computed using the solution $x_n = k_n/2n$ and not $\lceil k_n \rceil/2n$. It should be clear that as n goes to infinity, $k_n/2n$ and $\lceil k_n \rceil/2n$ converge to the same limit.

5. Computer Simulations of Card Collecting

To learn more about BCQ I developed a computer program to simulate the random acquisition of cards. To check my program, I used $n = 5$ different cards, and simulated the collection of cards until both piles were the same for the first time. I repeated the simulation a total of 1000 times and obtained $r'_5 = 0.641$ as the average ratio of cards in the "original" pile on $n = 5$. This agreed with my previously computed answer of $r_5 = 0.644$. I then simulated the problem 1000 times using $n = 10$. This gave a ratio of $r'_{10} = 0.6777$. I then made 100 runs each for $n = 100$, 1000, 10 000, and 100 000, getting sample means of $r'_{100} = 0.7808$, $r'_{1000} = 0.7973$, $r'_{10\,000} = 0.797\,505$, $r'_{100\,000} = 0.796\,623$. Three runs using $n = 1\,000\,000$ gave a sample mean of $r_{1\,000\,000} = 0.796\,69$.

The results of these simulations lead me to believe that the answers to BCQ and AQ are related in that both r_n and x_n tend to $X \approx 0.796\,812$, the root of $1 - e^{-2x} - x = 0$, as n tends to infinity.

For simulations in which the number of possible originals, n, is relatively small, the following TI-calculator program works well. For large n or a large number of simulations, some computer system should be used. I used a Basic program, which I will be glad to send to interested readers.

```
:Disp "NUMBER DIFF CARDS":Input N :0 → K :0 → L :Lbl 1
:K/N → P :rand → A :If A > P :1 + K → K :If A ≤ P :1 + L → L :If
K > L :Goto 1: Disp K
```

6. Are the Answers to the Questions the Same in the Limit?

A heuristic argument that $r_n \to X$ goes as follows. Let n be large. The probability that the two piles are the same at 1 each is $p_n(1) = 1/n$. For the first few cards added to the collection, the "original" pile is growing faster than the "duplicate" pile, since we are more likely to get an "original" than a "duplicate." In fact $p_n(2) = 4/n^2 - 4/n^3$ and $p_n(3)$ is on the order of $1/n^3$. But when the "original" pile has more than $n/2$ cards in it, the "duplicates" start accumulating faster than the "originals." When the collection has $k_n \approx 2nX$ cards in it, the two stacks are expected to be the same size, $e_n(k) = f_n(k) = k_n/2 \approx 0.8n$. Thus, for collection sizes k close to k_n, the "duplicate" pile is growing much faster than the original. In fact, at this point the "original" pile has about 80 percent of the possible cards, so there is a 20 percent chance of a new

card and an 80 percent chance of a duplicate. Thus, $de_n(k)/dk \approx 0.2$ and $df_n(k)/dk \approx 0.8$ at $k \approx k_n$. (The reader should compute these derivatives to check this claim.) Thus, $p_n(a)$ will be its largest for a-values close to $k_n/2$ and these probabilities should be significantly larger than all of the other probabilities. This should result in an average size a_n for the "original" pile that is close to $k_n/2$ and r_n should then be close to X.

This argument helps explain why the answers to the problems are apparently the same for large n. It is not meant to be a proof that the two limits are identical. It also leaves me wondering whether the stronger result,

$$|a_n - k_n| \to 0$$

is true.

This argument led me to thinking about rates of change and derivatives. We might ask for instance: For what value of k does the number of "originals" most exceed the number of "duplicates"; that is, for what k is

$$e_n(k) - f_n(k) = -2n(1 - 1/n)^k + 2n - k$$

largest. The calculus solution is to take the derivative of the function $e_n(k) - f_n(k)$ with respect to k, set the derivative equal to zero and solve. This solution is

$$k' = \frac{\ln 2 + \ln n + \ln(\ln n - \ln(n-1))}{\ln n - \ln(n-1)},$$

which should be rounded to the nearest integer.

Another solution is the value of k for which

$$e_n(k) = -n(1 - 1/n)^k + n = n/2$$

since beyond this point we are more likely to get a "duplicate" card. The solution to this is

$$k'' = \frac{\ln 2}{\ln n - \ln(n-1)}.$$

It is interesting that the difference in these solutions,

$$k' - k'' = \frac{\ln n + \ln(\ln n - \ln(n-1))}{\ln n - \ln(n-1)}$$

is slightly less than one-half for all values of n greater than 2, so that the two answers are essentially the same.

7. Conclusion

After all of this work, an exact answer to BCQ for arbitrary n still eludes me, although I believe this answer can be derived by properly using the sum

$$\sum_{j=0}^{n} j p_n(j,j),$$

and the relationship

$$p_n(i,j) = \frac{(n-i+1)\,p_n(i-1,j) + i p_n(i,j-1)}{n}.$$

Acknowledgement. I would like to thank Ray Bobo for sharing this problem with me and for making many helpful comments in the preparation of this article. I would also like to thank the referees for their helpful suggestions.

A Natural Classification of Curves and Surfaces With Reflection Properties

DANIEL DRUCKER
Wayne State University
Detroit, MI 48202

PHIL LOCKE
University of Maine
Orono, ME 04469

Introduction

Parabolic mirrors and elliptic domes ("whispering galleries") are familiar examples of focally reflective surfaces. In this paper we show that only conic curves and their corresponding surfaces of revolution (and in degenerate cases, lines and planes) are focally reflective. In [1], curves and surfaces with reflection properties were classified by solving differential equations in polar and spherical coordinates. Here we use a coordinate-free method to achieve the same classification by relating reflection properties to the defining focal properties of conics. The well-known orthogonality property of confocal conics comes as a bonus.

Finite Points and Points at Infinity

We identify each point P in $\mathbb{R}^n$ (for us, $n = 2$ or $n = 3$) with the vector $\overrightarrow{OP}$ from the origin to P. This enables us to write $Q - P$ instead of $\overrightarrow{PQ}$ whenever we wish. Points of $\mathbb{R}^n$ will sometimes be called *finite*. By contrast, a *point at infinity* is a line through the origin, viewed as a new "point" not in $\mathbb{R}^n$. (The term "point at infinity" comes from imagining a point that is approached by moving infinitely far away from the origin in either direction along the line.) The point at infinity specified by a line l will be denoted l^*. The set consisting of $\mathbb{R}^n$, together with its points at infinity, is denoted $\mathbb{P}^n$ and called *projective n-space*.[1] If $P \in \mathbb{R}^n$, we define *the line joining P to* l^* to be the line l_P through P parallel to l. We say that l^* "lies on" l_P.

Curves with Reflection Properties

Let $\alpha\colon I \to \mathbb{R}^2$ be a smooth regular parametrized curve in $\mathbb{R}^2$ defined on an open interval I, and let F_1, F_2 be points in $\mathbb{P}^2 \setminus \alpha(I)$. ("Regular" means that $\alpha'(t) \neq 0$ for all $t \in I$.) Following [1], we say that α has a *reflection property* with foci F_1 and F_2 if, for each point $P \in \alpha(I)$, the following conditions hold:

(i) Any vector normal to the curve α at P lies in the span of the vectors $\overrightarrow{F_1P}$ and $\overrightarrow{F_2P}$.

[1]Projective geometry is not used in this article. We refer interested readers to [6] or [5] for a gentle, non-axiomatic introduction to the subject. For a projective geometry interpretation of the focal properties of conics, see [7].

(ii) The line normal to α at P bisects one of the pairs of opposite angles formed by the intersection of the lines joining F_1 and F_2 to P.

If F_1 is a point at infinity, we take $\overrightarrow{F_1P}$ to be any nonzero vector parallel to the line joining P to F_1, and similarly for F_2. Condition (i) is vacuous unless F_1, F_2, and P lie on the same line; in this case, (i) says that α is orthogonal to that line at P. (This occurs, for example, when $F_1 = F_2$, but not when α is part of the line through a pair of distinct foci.)

Now let α be as above, and let F_1 and F_2 be *distinct* points in $\mathbb{P}^2 \setminus \alpha(I)$. At each point $P = \alpha(t)$ along α, we set $\mathbf{r}_k(t) = \overrightarrow{F_kP}$ and $\mathbf{u}_k(t) = \mathbf{r}_k(t)/\|\mathbf{r}_k(t)\|$, $k = 1, 2$. The vector-valued functions $\mathbf{r}_k$ and $\mathbf{u}_k$ describe, respectively, the position vectors from F_k to points along α and the corresponding unit vectors.

We say that α is a *positive* (resp. *negative*) *reflector* with foci F_1, F_2 if, for all $t \in I$, (i) holds and $\mathbf{u}_1(t) + \mathbf{u}_2(t)$ (resp. $\mathbf{u}_1(t) - \mathbf{u}_2(t)$) is normal to α. Figures 1 and 2 illustrate these reflection properties for finite foci. It follows from (ii) and the smoothness of our curves that every curve having a reflection property with distinct foci F_1, F_2 must be a positive or negative reflector with foci F_1, F_2. Conversely, every positive or negative reflector with distinct foci F_1 and F_2 has a reflection property with the same foci. (Condition (i) is used here to handle cases in which $\mathbf{u}_1 = \pm\mathbf{u}_2$.) Thus a curve α that satisfies (i) will also satisfy (ii) $\Leftrightarrow$ either $[\mathbf{u}_1(t) + \mathbf{u}_2(t)] \cdot \alpha'(t) = 0$ for all $t \in I$ or $[\mathbf{u}_1(t) - \mathbf{u}_2(t)] \cdot \alpha'(t) = 0$ for all $t \in I$.

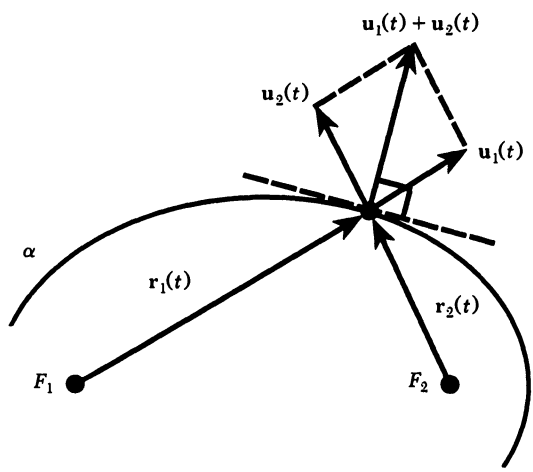


FIGURE 1
A positive reflector α with distinct finite foci F_1, F_2.

A well-known orthogonality property of conics now follows easily:

A positive and a negative reflector having the same (distinct) foci are orthogonal at all points of intersection.

To see why, let α_+ and α_- be two such reflectors and let P be any point of intersection. Then the unit position vectors $\mathbf{U}_1$ and $\mathbf{U}_2$ at P are the same for both curves. Now $\mathbf{U}_1 + \mathbf{U}_2$ is orthogonal to α_+ at P, $\mathbf{U}_1 - \mathbf{U}_2$ is orthogonal to α_- at P, and $\mathbf{U}_1 + \mathbf{U}_2$ is orthogonal to $\mathbf{U}_1 - \mathbf{U}_2$ since $(\mathbf{U}_1 + \mathbf{U}_2) \cdot (\mathbf{U}_1 - \mathbf{U}_2) = \|\mathbf{U}_1\|^2 - \|\mathbf{U}_2\|^2 = 1 - 1 = 0$. Thus α_+ and α_- are orthogonal at P, as claimed.

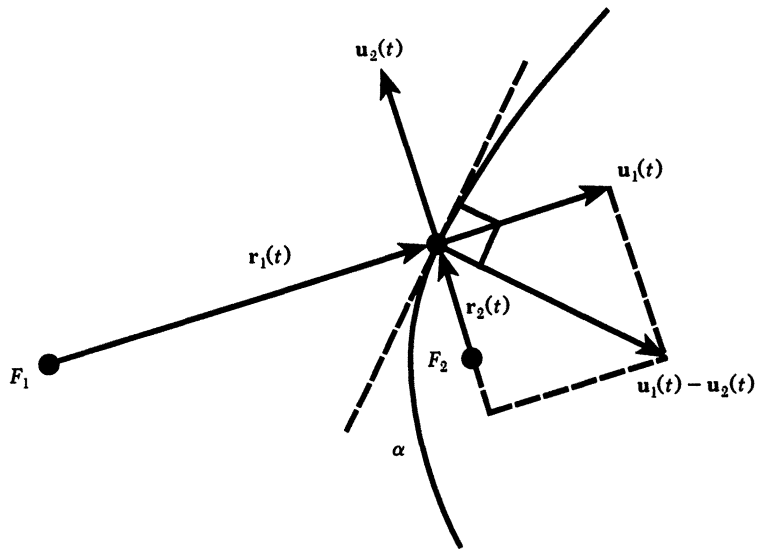


FIGURE 2
A negative reflector α with distinct finite foci F_1, F_2.

We now proceed, through several cases, to classify all curves with reflection properties.

Case 1: F_1 and F_2 are distinct and finite. Condition (ii) holds if and only if $[\mathbf{u}_1(t) \pm \mathbf{u}_2(t)] \cdot \alpha'(t) = 0$ for all $t \in I$; the sign is positive or negative according to whether α is a positive or negative reflector. Since $\mathbf{r}_k(t) = \overrightarrow{F_k P} = \alpha(t) - F_k$, we see that $\mathbf{r}_k'(t) = \alpha'(t)$, and

$$\|\mathbf{r}_k(t)\|' = \left[\sqrt{\mathbf{r}_k(t) \cdot \mathbf{r}_k(t)}\right]' = \frac{\mathbf{r}_k(t) \cdot \mathbf{r}_k'(t)}{\sqrt{\mathbf{r}_k(t) \cdot \mathbf{r}_k(t)}} = \frac{\mathbf{r}_k(t) \cdot \alpha'(t)}{\|\mathbf{r}_k(t)\|} = \mathbf{u}_k(t) \cdot \alpha'(t). \tag{1}$$

Therefore,

$$\begin{aligned} \text{(ii) holds} &\Leftrightarrow \left[\|\mathbf{r}_1(t)\| \pm \|\mathbf{r}_2(t)\|\right]' = 0 \text{ for all } t \in I \\ &\Leftrightarrow \|\mathbf{r}_1(t)\| \pm \|\mathbf{r}_2(t)\| = c \text{ for all } t \in I, \ c \text{ a constant} \\ &\Leftrightarrow d(P, F_1) \pm d(P, F_2) = c \text{ for all } P \in \alpha(I), \end{aligned} \tag{2}$$

where $d(P, F_k)$ denotes the distance from P to F_k.

With a "+" sign, (2) describes an ellipse if $c > d(F_1, F_2)$, the line segment $\overline{F_1 F_2}$ if $c = d(F_1, F_2)$, and the empty set otherwise. With a "−" sign, equation (2) describes one branch of a hyperbola if $0 < |c| < d(F_1, F_2)$, a straight line (the perpendicular bisector of $\overline{F_1 F_2}$) if $c = 0$, and a ray from one focus in the direction opposite the other focus if $|c| = d(F_1, F_2)$. Notice that condition (i) rules out both $\overline{F_1 F_2}$ and the ray, since all their points are collinear with F_1 and F_2. Condition (i) does not, however, rule out the perpendicular bisector. We conclude, therefore, that if α is a positive reflector, then it is part of an ellipse; if α is a negative reflector, it is part of a hyperbola or a straight line. By the italicized remark above, an ellipse is orthogonal to any hyperbola with the same foci, as well as to the perpendicular bisector of the line segment joining the foci.

Case 2: F_1 is finite and F_2 is a point at infinity. Let $F_2 = l^*$ and let $\mathbf{u}$ be a unit vector parallel to l. Then $\mathbf{u}_2(t) = \mathbf{u}$ for all t, and, using (1),

$$\begin{aligned}\text{(ii) holds} &\Leftrightarrow [\mathbf{u}_1(t) \pm \mathbf{u}] \cdot \alpha'(t) = 0 \quad \text{for all } t \in I \\ &\Leftrightarrow [\|\mathbf{r}_1(t)\| \pm \mathbf{r}_1(t) \cdot \mathbf{u}]' = 0 \quad \text{for all } t \in I \\ &\Leftrightarrow \|\mathbf{r}_1(t)\| \pm \mathbf{r}_1(t) \cdot \mathbf{u} = c \quad \text{for all } t \in I, \ c \text{ a constant} \\ &\Leftrightarrow d(P, F_1) + \varepsilon(P - F_1) \cdot \mathbf{u} = c \quad \text{for all } P \in \alpha(I) \text{ and } \varepsilon = \pm 1. \quad (3)\end{aligned}$$

Suppose that $\varepsilon = +1$ in (3). Since $|(P - F_1) \cdot \mathbf{u}| \le \|P - F_1\| = d(P, F_1)$, (3) has solutions P only when $c \ge 0$. If $c = 0$, then solutions P must satisfy $(P - F_1) \cdot \mathbf{u} = -d(P, F_1) < 0$, so they must lie on the open ray from F_1 in the direction of $-\mathbf{u}$. Condition (i) rules out this possibility, so we must have $c > 0$. Then solutions P satisfy

$$d(P, F_1) = ((F_1 + c\mathbf{u}) - P) \cdot \mathbf{u}. \qquad (4_+)$$

Let l_+ be the line through the point $F_1 + c\mathbf{u}$ orthogonal to $\mathbf{u}$. Since $F_1 \notin \alpha(I)$, the right-hand side of (4_+) must be positive, which means that solutions P must lie on the side of l_+ in the direction of $-\mathbf{u}$. (See Figure 3a.) Thus (4_+) says that $d(P, F_1) = d(P, l_+)$. This condition describes a parabola with focus F_1 and directrix l_+; the vertex is at $F_1 + \frac{1}{2}c\mathbf{u}$ and the parabola opens in the direction of $-\mathbf{u}$.

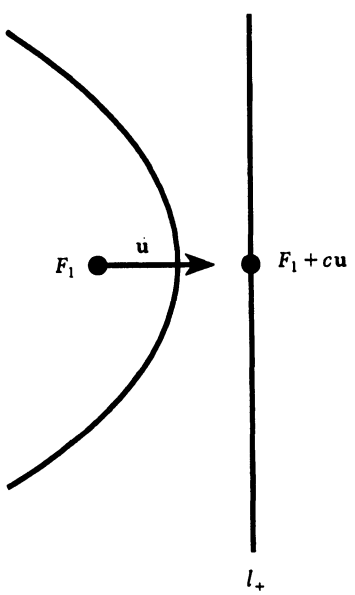


FIGURE 3a.

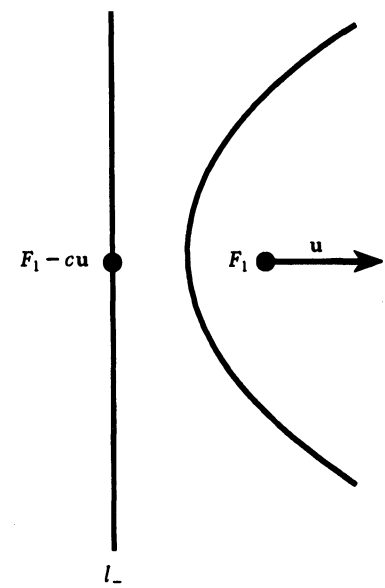


FIGURE 3b.

The analysis is similar when $\varepsilon = -1$; $c > 0$ and the solutions satisfy

$$d(P, F_1) = (P - (F_1 - c\mathbf{u})) \cdot \mathbf{u}. \qquad (4_-)$$

This condition describes the parabola with focus F_1 and directrix l_-, the line through $F_1 - c\mathbf{u}$ orthogonal to $\mathbf{u}$. The vertex is at $F_1 - \frac{1}{2}c\mathbf{u}$ and the parabola opens in the direction of $\mathbf{u}$. (See Figure 3b.)

As c ranges through positive values, we obtain two families of parabolas, one for each of $\varepsilon = \pm 1$. All these parabolas share the same foci, so any two parabolas, one from each family, intersect orthogonally. Note, however, that $\mathbf{u}$ was only required to be parallel to l; its direction was arbitrary. Choosing $-\mathbf{u}$ instead of $\mathbf{u}$ interchanges the

two families of parabolas. Thus, in this case, positive and negative reflectors are not qualitatively different.

Case 3: F_1 and F_2 are distinct points at infinity. Write $F_1 = l_1^*$ and $F_2 = l_2^*$ and let $\mathbf{u}_1$ and $\mathbf{u}_2$ be unit vectors parallel to l_1 and l_2 respectively. Here, condition (i) is vacuous. As in earlier cases, $\alpha'(t)\cdot(\mathbf{u}_1 + \varepsilon\mathbf{u}_2) = 0$ for all t, where $\varepsilon = \pm 1$. Thus $\alpha(t)\cdot(\mathbf{u}_1 + \varepsilon\mathbf{u}_2) = c$ for some constant c. Equivalently, $\{\alpha(t) - c(\mathbf{u}_1 + \varepsilon\mathbf{u}_2)/\|\mathbf{u}_1 + \varepsilon\mathbf{u}_2\|^2\}\cdot(\mathbf{u}_1 + \varepsilon\mathbf{u}_2) = 0$ for all t. It follows that α is part of the straight line that is orthogonal to $\mathbf{u}_1 + \varepsilon\mathbf{u}_2$ (i.e., parallel to $\mathbf{u}_1 - \varepsilon\mathbf{u}_2$) and passes through the point $c(\mathbf{u}_1 + \varepsilon\mathbf{u}_2)/\|\mathbf{u}_1 + \varepsilon\mathbf{u}_2\|^2$. As c ranges through real values, we obtain two families of straight lines: for $\varepsilon = 1$, a family of positive reflectors, orthogonal to $\mathbf{u}_1 + \mathbf{u}_2$; and for $\varepsilon = -1$, a family of negative reflectors, orthogonal to $\mathbf{u}_1 - \mathbf{u}_2$. As usual, the two families intersect orthogonally.

Case 4: F_1 and F_2 are finite and equal. When $F_1 = F_2$, condition (i) says that at each point $P = \alpha(t)$, α is normal to the line through P and F_1; i.e., $\alpha'(t)\cdot(\alpha(t) - F_1) = 0$ for all $t \in I$. But

$$\alpha'(t)\cdot(\alpha(t) - F_1) = 0 \text{ for all } t \Leftrightarrow [(\alpha(t) - F_1)\cdot(\alpha(t) - F_1)]' = 0 \text{ for all } t$$

$$\Leftrightarrow \|\alpha(t) - F_1\|^2 = c \text{ for all } t,$$

for a constant c. Since $\alpha(I)$ contains more than one point, c must be positive; thus α is part of the circle of radius $\sqrt{c}$ centered at F_1.

Case 5: F_1 and F_2 are the same point at infinity. Let $F_1 = F_2 = l^*$, and let $\mathbf{u}$ be a unit vector parallel to l. Then by (i), $\alpha'(t)\cdot\mathbf{u} = 0$ for all $t \in I$, so $\alpha(t)\cdot\mathbf{u} = c$ for all t, where c is a constant. Equivalently, $(\alpha(t) - c\mathbf{u})\cdot\mathbf{u} = 0$ for all t; i.e., α is part of the straight line through the point $c\mathbf{u}$, orthogonal to $\mathbf{u}$.

We have shown that a curve α with a reflection property must be part of an ellipse, hyperbola, parabola, circle, or straight line. Working backwards through the arguments shows that, conversely, each of these curves has a reflection property. We summarize our results in a theorem.

THEOREM. *A smooth connected plane curve has a reflection property if and only if it is part of an ellipse, hyperbola, parabola, circle, or straight line.*

A fixed pair of foci determines a family of curves with reflection properties. Positive and negative reflectors with the same distinct foci are orthogonal. The following table summarizes the classification:

Distinct foci	Both finite	One finite, one infinite	Both infinite
Positive reflectors:	confocal ellipses	confocal parabolas	parallel lines
Negative reflectors:	confocal hyperbolas and the perpendicular bisector of the line segment joining the foci	confocal parabolas	parallel lines

Equal foci	Finite		Infinite
	concentric circles		parallel lines

Figure 4 depicts a pair of distinct finite foci and the orthogonal families they determine; Figure 5 shows families of parabolas that arise from distinct foci when only one focus is finite.

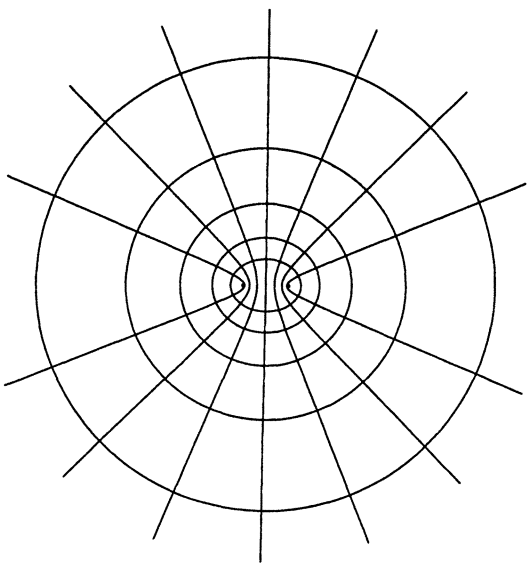


FIGURE 4
Confocal ellipses and hyperbolas

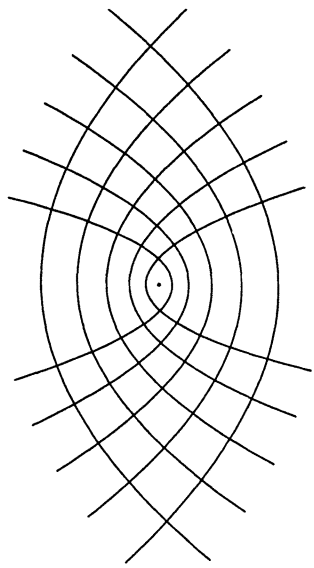


FIGURE 5
Confocal parabolas

Note. The heart of the classification is the observation that condition (ii) is equivalent to relation (2) in Case 1 and to relations (4_+) and (4_-) in Case 2. That part of the argument is essentially a bidirectional version of the calculations in [4], which used differentials. The classification argument is simpler than that in [1] because it uses the reflection property to obtain the defining focal properties of the conic sections, rather than to obtain their equations in a particular coordinate system.

Surfaces with Reflection Properties

The notion of a reflection property for curves extends in a natural way to surfaces in $\mathbb{R}^3$ (indeed, to hypersurfaces in $\mathbb{R}^n$, $n \geq 3$, though we leave that as an exercise for the interested reader).

A smooth connected surface $\mathcal{S}$ in $\mathbb{R}^3$ is said to have a *reflection property* if there are points F_1, F_2 (called *foci*) in $\mathbb{P}^3 \setminus \mathcal{S}$ such that, for each point P in $\mathcal{S}$,

(i) any vector normal to $\mathcal{S}$ at P lies in the span of the vectors $\overrightarrow{F_1P}$ and $\overrightarrow{F_2P}$; and
(ii) the normal line to $\mathcal{S}$ at P bisects one of the pairs of opposite angles formed by the intersection of the lines joining P to F_1 and F_2.

If F is a focus at infinity, then we take $\overrightarrow{FP}$ to be any nonzero vector parallel to the line joining P to F. If F_1, F_2, and P lie on a line, then by (i), that line is normal to $\mathcal{S}$ at P, and (ii) adds no more information. Otherwise (i) follows from (ii).

Since (i) and (ii) are local conditions, we lose no generality by assuming that $\mathcal{S}$ is a parametrized surface; i.e., that $\mathcal{S} = \sigma(\mathcal{U})$, where $\sigma\colon \mathcal{U} \to \mathbf{R}^3$ is a smooth regular map from an open subset $\mathcal{U}$ of $\mathbf{R}^2$ onto $\mathcal{S}$ in $\mathbf{R}^3$. "Regular" means that at each point $P = \sigma(u_0, v_0)$ of the surface, the tangent vectors $\partial_u\sigma(u_0, v_0) = (\partial\sigma/\partial u)(u_0, v_0)$ and $\partial_v\sigma(u_0, v_0) = (\partial\sigma/\partial v)(u_0, v_0)$ are linearly independent and thus span the tangent plane to $\mathcal{S}$ at P.

For brevity, we denote typical points of $\mathcal{U}$ and $\mathcal{S}$ by p and $P = \sigma(p)$. As before, we set $\mathbf{r}_k(p) = \overrightarrow{F_kP}$ and $\mathbf{u}_k(p) = \mathbf{r}_k(p)/\|\mathbf{r}_k(p)\|$ for $k = 1, 2$, using them to define positive and negative reflectors. If F_1 and F_2 are distinct, then a surface $\mathcal{S}$ has a reflection property with foci F_1 and F_2 if and only if it is a positive or negative reflector with those foci.

Suppose $\mathcal{S}$ has a reflection property with distinct foci. When both foci are finite, the analogue of (1) is the pair of relations

$$\partial_u\|\mathbf{r}_k(p)\| = \mathbf{u}_k(p)\cdot\partial_u\sigma(p), \qquad \partial_v\|\mathbf{r}_k(p)\| = \mathbf{u}_k(p)\cdot\partial_v\sigma(p). \tag{5}$$

Since $\mathcal{S}$ is a positive or negative reflector, these relations imply that

$$\partial_u[\|\mathbf{r}_1(p)\| \pm \|\mathbf{r}_2(p)\|] = [\mathbf{u}_1(p) \pm \mathbf{u}_2(p)]\cdot\partial_u\sigma(p) = 0$$

and

$$\partial_v[\|\mathbf{r}_1(p)\| \pm \|\mathbf{r}_2(p)\|] = [\mathbf{u}_1(p) \pm \mathbf{u}_2(p)]\cdot\partial_v\sigma(p) = 0.$$

(Choose the appropriate sign.) Hence $\|\mathbf{r}_1(p)\| \pm \|\mathbf{r}_2(p)\|$ is constant as a function of p; this means that every plane cross-section of $\mathcal{S}$ passing through the foci is the *same* curve having a reflection property with foci F_1, F_2. Thus $\mathcal{S}$ is part of an ellipsoid of revolution for a positive reflector; for a negative reflector, $\mathcal{S}$ is part of a hyperboloid of revolution or part of the plane that is the perpendicular bisector of the line segment joining the foci. Similar reasoning shows that with one finite focus and one focus at infinity, $\mathcal{S}$ is part of a paraboloid of revolution. With both foci at infinity, $\mathcal{S}$ is part of a plane.

When $\mathcal{S}$ has a reflection property with equal finite foci, (i) says that $\mathbf{r}_1(p)$ is normal to $\mathcal{S}$ for each $p \in \mathcal{U}$. This says that $\partial_u\mathbf{r}_1(p)\cdot\mathbf{r}_1(p) = \partial_v\mathbf{r}_1(p)\cdot\mathbf{r}_1(p) = 0$ for all p, so $\partial_u(\|\mathbf{r}_1(p)\|^2) = \partial_v(\|\mathbf{r}_1(p)\|^2) = 0$ for all p. It follows that $\|\mathbf{r}_1(p)\|$ is constant as a function of p, so $\mathcal{S}$ is part of a sphere centered at F_1.

Finally, if $\mathcal{S}$ has a reflection property with equal foci l^* at infinity, we let $\mathbf{u}$ be a unit vector parallel to l and, proceeding as in Case 5, find that $\mathcal{S}$ is part of a plane orthogonal to $\mathbf{u}$. To summarize:

THEOREM. *A smooth connected surface has a reflection property if and only if it is part of an ellipsoid of revolution, a hyperboloid of revolution, a paraboloid of revolution, a sphere, or a plane.*

A fixed pair of foci determines a family of surfaces with reflection properties. Positive and negative reflectors with the same distinct foci are orthogonal. The following table summarizes the classification.

Distinct foci	Both finite	One finite, one infinite	Both infinite
Positive reflectors:	confocal ellipsoids	confocal paraboloids	parallel planes
Negative reflectors:	confocal hyperboloids and the plane perpendicular bisector of the line segment joining the foci	confocal paraboloids	parallel planes

Equal foci	Finite		Infinite
	concentric spheres		parallel planes

The general result for hypersurfaces in $\mathbb{R}^n$ can be found in **[2]** (see also **[3]**) or **[8]**, but with different proofs.

Acknowledgement. The authors independently submitted different versions of this classification proof at about the same time. We wish to acknowledge that Harley Flanders sent yet another version of this proof to the first author only about a month later. When three mathematicians have the same idea at about the same time, it is not just a coincidence. Perhaps it means that this is the "right" way to do the classification.

REFERENCES

1. D. Drucker, Reflection properties of curves and surfaces, this MAGAZINE 65 (1992), 147–157.
2. D. Drucker, Euclidean hypersurfaces with reflection properties, *Geometriae Dedicata* 33 (1990), 325–329.
3. D. Drucker, Reflective Euclidean hypersurfaces, *Geometriae Dedicata* 39 (1991), 361–362.
4. H. Flanders, The optical property of the conics, *Amer. Math. Monthly* 75 (1968), 399.
5. L. E. Garner, *An Outline of Projective Geometry*, North Holland, New York, 1981.
6. K. Hanes, An Introduction to Analytic Projective Geometry and Its Applications (UMAP Module 710), *UMAP Modules: Tools for Teaching 1990*, Birkhauser, Boston, 1990, pp. 111–149.
7. O. Veblen and J. W. Young, *Projective Geometry*, Vol. II, Ginn and Company, Boston, 1918, pp. 189–196.
8. B. Wegner, Comment on 'Euclidean hypersurfaces with reflection properties,' *Geometriae Dedicata* 39 (1991), 357–359.

NOTES

A Parenthetical Note (to a Paper of Guy)

MARK KRUSEMEYER
Carleton College
Northfield, MN 55057

Guess my number $1, 2, 5, 14, \ldots$. What is the next number in this sequence? Of course such questions have no logical validity, even if they do persist on tests purporting to measure mathematical skill or aptitude. But in defense of such questions it can be argued that they often seize the eye and the imagination—and that when the terms of the sequence are generated by some specific procedure, the pattern-finding needed to predict future terms can be an important mathematical activity.

Back to $1, 2, 5, 14, \ldots$. To an alert middle school student, the next number is surely 41, since each term shown can be found from the previous one by multiplying by 3, then subtracting 1. Extending the sequence in this way gives us M1458 in Sloane and Plouffe's *Encyclopedia* **[5]**: the n-th term is $\frac{1}{2}(3^{n-1} + 1)$.

Sloane and Plouffe list thirteen other sequences that start either with the same four terms, or with an extra 1 at the beginning for lagniappe. One example is the stamp-folding sequence M1455: $1, 1, 2, 5, 14, 38, \ldots$. The n-th term of this sequence gives the number of ways to take a strip of n ungummed, blank stamps (so you can't tell left from right, top from bottom, or front from back) and fold it so that all n stamps end up on top of each other. Although the sixth term was cited as 39 in **[1]** on the authority of Table 4 of **[2]**, the rest of that table and Koehler's Theorem 3.4 show that 39 was a misprint.

A professional guess To professional mathematicians, the best-known sequence starting $1, 2, 5, 14, \ldots$ is M1459 in **[5]**, which consists of the *Catalan numbers*

$$c_n = \frac{1}{n+1}\binom{2n}{n}.$$

This sequence and several of its "start-alikes" figure prominently in Richard Guy's delightful article on the Second Strong Law (or, if you prefer, second delightful article on the Strong Law) of Small Numbers **[1]**. In particular, Guy mentions two ways of associating the Catalan numbers with parenthesization. The purpose of this note is to point out a direct connection between those two ways.

The first "manifestation" of c_n is as the number of possible interpretations of a non-associative product of $n + 1$ letters. Computing such a product involves carrying out n multiplications, so in principle n pairs of parentheses are needed, but in practice the outer pair is left off and only $n - 1$ pairs are displayed. For example, the product $abcde$ can be interpreted as $(ab)(c(de))$, as $(((ab)c)d)e$, or in any of twelve other ways, so $c_4 = 14$. When Catalan numbers are mentioned in textbooks on combinatorics, it is often in this context; see, e.g., **[3]**, Section 5.4.2. We will call a parenthesization of a product of $n + 1$ letters a *bracketing* and denote the set of all such bracketings by B_n. To try to avoid confusion with the other type of parenthesization (which is about to be discussed), we will use "floor" symbols rather than

parentheses when writing bracketings. For instance, we will write

$$B_3 = \{\lfloor a\lfloor bc\rfloor\rfloor d, \lfloor\lfloor ab\rfloor c\rfloor d, \lfloor ab\rfloor\lfloor cd\rfloor, a\lfloor b\lfloor cd\rfloor\rfloor, a\lfloor\lfloor bc\rfloor d\rfloor\}$$

for the set of bracketings whose elements are counted by $c_3 = 5$. Note that the actual choice of letters is immaterial; $\lfloor r\lfloor at\rfloor\rfloor s$ should be considered the same bracketing as $\lfloor a\lfloor bc\rfloor\rfloor d$.

The other "manifestation" of c_n is not as common in the literature. It can be found in [6], and it appears to have originated with Conway and Guy. They observed that c_n is the number of ways to arrange n pairs of *empty* parentheses, subject only to the restriction that no closing parenthesis can precede the corresponding opening one. For example, ( )) (is an *ineligible* way to arrange two pairs of parentheses, because the second closing parenthesis comes too early. We will call an arrangement of pairs of parentheses in which no closing parenthesis precedes the corresponding opening one a *CG-arrangement*; the set of all such arrangements of n pairs will be denoted by CG_n. When writing *CG*-arrangements, we will use "ceiling" symbols rather than parentheses. For instance,

$$CG_3 = \{\lceil\lceil\lceil\rceil\rceil, \lceil\lceil\rceil\lceil\rceil, \lceil\lceil\rceil\rceil\lceil, \lceil\rceil\lceil\lceil\rceil, \lceil\rceil\lceil\rceil\lceil\}.$$

(For a recent proof starting from *all* $\binom{2n}{n}$ ways to arrange n pairs of parentheses that CG_n has exactly c_n elements, see [4].)

Now for the connection It is not immediately clear that there is any connection between bracketings and *CG*-arrangements. If we simply omit the letters from a bracketing, we lose far too much information; for instance, the five bracketings in B_3 yield $\lfloor\lfloor\rfloor\rfloor$ (four times) and $\lfloor\rfloor\lfloor\rfloor$ (once). If we consider symmetry, we see that of the bracketings in B_3, only $\lfloor ab\rfloor\lfloor cd\rfloor$ is its own reflection, whereas of the arrangements in CG_3, $\lceil\lceil\lceil\rceil\rceil$, $\lceil\lceil\rceil\lceil\rceil$, and $\lceil\rceil\lceil\rceil\lceil$ are all symmetrical. Thus Guy suggested in [1] that one was "unlikely... [to] find a direct combinatorial comparison."

However, we will see that there is a natural, recursive way to define a 1–1 correspondence between *CG*-arrangements and bracketings. We will also see that this correspondence does, indeed, "break symmetry."

The following idea is at the heart of the recursive definition. (After the first draft of this note was written, I found the same basic idea in the discussion on p. 43 of [7], another source for the second manifestation of c_n.) Given a *CG*-arrangement α, we define the *break point* of α to be the *first* point *after* the beginning of α, reading from left to right, where every pair of parentheses that has been opened has been closed again. For example, in each case shown below, the break point occurs at the arrow:

$$\lceil\rceil_{\uparrow}\lceil\lceil\rceil\rceil$$

$$\lceil\lceil\rceil\lceil\rceil\rceil_{\uparrow}\lceil\rceil\lceil\rceil$$

Note that the break point may be at the very end of α. Since the break point is never at the beginning of α and since the parentheses to either side of the break point are balanced, there are *CG*-arrangements β and γ such that α is of the form $\lceil\beta\rceil\gamma$, with the break point immediately before γ. For instance, in the cases shown

above we have

$$\beta = \varnothing,\ \gamma = [[]] \quad \text{and}$$

$$\beta = [][],\ \gamma = [][], \text{ respectively.}$$

(There is one exception: If $\alpha = \varnothing$, the break point is undefined, and of course α is then *not* of the form $[\beta]\gamma$.)

Note that even without specifying the break point, β and γ are uniquely determined by $\alpha = [\beta]\gamma$, for if β were to extend through the break point, at that point a parenthesis in β would be closed prematurely.

Roughly speaking, the recursive way to map *CG*-arrangements to bracketings is to split each *CG*-arrangement at the break point, then to map each of the separate arrangements β and γ such that $\alpha = [\beta]\gamma$ to a bracketing, and finally to "reassemble" those bracketings. We will now make this precise by defining 1–1 correspondences F_n: $CG_n \to B_n$ for all $n \geq 0$.

For $n = 0$, there is no problem: To the unique (empty) arrangement of no pair of parentheses, we associate the unique (invisible) bracketing of a single letter. For $n > 0$ and $\alpha \in CG_n$, we have seen above that there is a unique way to write $\alpha = [\beta]\gamma$ with $\beta \in CG_k$, $\gamma \in CG_{n-k-1}$, $0 \leq k < n$. In defining $F_n(\alpha)$, we may assume by recursive hypothesis that $F_k(\beta)$ and $F_{n-k-1}(\gamma)$ are already defined. This allows us to form the bracketing

$$F_n(\alpha) = \lfloor F_k(\beta) \rfloor \lfloor F_{n-k-1}(\gamma) \rfloor$$

of $(k+1)+(n-k) = n+1$ letters. (If $k = 0$ or $k = n-1$, the "extra" floor symbols around $F_k(\beta)$ or $F_{n-k-1}(\gamma)$, respectively, should be omitted.)

As an example, let's find $F_5(\alpha)$, where α is the *CG*-arrangement $[[][]]_{\uparrow}[[]]$. The break point is indicated by the arrow, and we have $\alpha = [\beta]\gamma$ with $\beta = [][]$, $\gamma = [[]]$. These arrangements, in turn, break up as $\beta = [\beta_1]\beta_2$ with $\beta_1 = \varnothing$, $\beta_2 = []$ and $\gamma = [\gamma_1]\gamma_2$ with $\gamma_1 = []$, $\gamma_2 = \varnothing$. Finally, $\beta_2 = \gamma_1 = []$ breaks up as $[\delta]\varepsilon$ with $\delta = \varepsilon = \varnothing$. Therefore, we have

$$F_1(\beta_2) = F_1(\gamma_1) = F_0(\delta)F_0(\varepsilon) = ab;$$

$$F_2(\beta) = F_0(\beta_1)\lfloor F_1(\beta_2) \rfloor = a\lfloor bc \rfloor$$

(remember that the actual letters are immaterial to the bracketing!);

$$F_2(\gamma) = \lfloor F_1(\gamma_1) \rfloor F_0(\gamma_2) = \lfloor ab \rfloor c;$$

$$F_5(\alpha) = \lfloor F_2(\beta) \rfloor \lfloor F_2(\gamma) \rfloor = \lfloor a\lfloor bc \rfloor \rfloor \lfloor \lfloor de \rfloor f \rfloor.$$

So to the *CG*-arrangement $\alpha = [[][]][[]]$ corresponds the bracketing $\lfloor a\lfloor bc \rfloor \rfloor \lfloor \lfloor de \rfloor f \rfloor$. Note that this is a symmetrical bracketing, even though there is no apparent symmetry to α.

We can use induction on n to show that the F_n are 1–1 correspondences; here is an outline of the proof. The basis step, for $n = 0$, is clear, so we can assume that F_n is a 1–1 correspondence for each $n < N$. Now note that any bracketing $\Omega \in B_N$ can be written uniquely in one (and only one) of the following three forms: $a\lfloor \Phi \rfloor$, with $\Phi \in B_{N-1}$; $\lfloor \Phi \rfloor a$, with $\Phi \in B_{N-1}$; $\lfloor \Phi \rfloor \lfloor \Psi \rfloor$, with $\Phi \in B_k, \Psi \in B_{N-k-1}$ for some $k < N$. In the first case, since F_{N-1} is a 1–1 correspondence, there is a unique $\varphi \in CG_{N-1}$ with $F_{N-1}(\varphi) = \Phi$; we then have $F_N([]\varphi) = a\lfloor \Phi \rfloor$. Similarly, in the

second case we have $F_N(\lceil \varphi \rceil) = \lfloor \Phi \rfloor a$, where once again $F_{N-1}(\varphi) = \Phi$. Finally, in the third case there are unique $\varphi \in CG_k$, $\psi \in CG_{N-k-1}$ with $F_k(\varphi) = \Phi$, $F_{N-k-1}(\psi) = \Psi$, and we then have $F_N(\lceil \varphi \rceil \psi) = \lfloor \Phi \rfloor \lfloor \Psi \rfloor$. So F_N is onto, and from the uniqueness of φ, ψ in the various cases above we soon see that F_N is one-to-one as well.

The following table shows the correspondence F_4. That is, in each row the bracketing is the image under F_4 of the CG-arrangement.

CG-arrangement	Bracketing	CG-arrangement	Bracketing
$\lceil\lceil\lceil\lceil\rceil\rceil\rceil\rceil$	$\lfloor\lfloor\lfloor ab \rfloor c \rfloor d \rfloor e$	$\lceil\lceil\rceil\lceil\rceil\lceil\rceil$	$\lfloor a \lfloor bc \rfloor \rfloor \lfloor de \rfloor$
$\lceil\lceil\lceil\rceil\rceil\lceil\rceil$	$\lfloor\lfloor ab \rfloor c \rfloor \lfloor de \rfloor$	$\lceil\rceil\lceil\lceil\rceil\lceil\rceil$	$a \lfloor\lfloor b \lfloor cd \rfloor \rfloor e \rfloor$
$\lceil\rceil\lceil\lceil\lceil\rceil\rceil$	$a\lfloor\lfloor\lfloor bc \rfloor d \rfloor e \rfloor$	$\lceil\lceil\rceil\lceil\rceil\lceil\rceil$	$\lfloor ab \rfloor \lfloor c \lfloor de \rfloor \rfloor$
$\lceil\lceil\lceil\rceil\lceil\rceil\rceil$	$\lfloor\lfloor a \lfloor bc \rfloor \rfloor d \rfloor e$	$\lceil\rceil\lceil\lceil\rceil\lceil\rceil$	$a\lfloor\lfloor bc \rfloor \lfloor de \rfloor \rfloor$
$\lceil\lceil\lceil\rceil\rceil\lceil\rceil$	$\lfloor\lfloor ab \rfloor \lfloor cd \rfloor \rfloor e$	$\lceil\rceil\lceil\rceil\lceil\rceil$	$a \lfloor b \lfloor\lfloor cd \rfloor e \rfloor \rfloor$
$\lceil\lceil\rceil\lceil\lceil\rceil$	$\lfloor a\lfloor\lfloor bc \rfloor d \rfloor \rfloor e$	$\lceil\lceil\rceil\lceil\rceil$	$\lfloor a \lfloor b \lfloor cd \rfloor \rfloor \rfloor e$
$\lceil\lceil\rceil\lceil\rceil\lceil\rceil$	$\lfloor ab \rfloor \lfloor\lfloor cd \rfloor e \rfloor$	$\lceil\rceil\lceil\rceil\lceil\rceil$	$a\lfloor\, b \lfloor c \lfloor de \rfloor \rfloor \rfloor$

Acknowledgements. It is a pleasure to thank Richard Guy, and more generally the Department of Mathematics and Statistics of the University of Calgary, for hospitality during a sabbatical leave. More recently, Richard generously suggested several ways to improve the exposition of the present note, including the use of "floor" and "ceiling" symbols to distinguish between different types of parenthesization.

REFERENCES

1. Richard K. Guy, The second strong law of small numbers, this MAGAZINE 63 (1990), 3–20.
2. John E. Koehler, Folding a strip of stamps, *Journal of Combinatorial Theory* 5 (1968), 135–152.
3. Fred S. Roberts, *Applied Combinatorics*, Prentice-Hall, Englewood Cliffs, NJ, 1984.
4. Daniel Rubenstein, Catalan numbers revisited, *Journal of Combinatorial Theory* A 68 (1994), 486–490.
5. N. J. A. Sloane and Simon Plouffe, *The Encyclopedia of Integer Sequences*, Academic Press, San Diego, CA, 1995.
6. Dennis Stanton and Dennis White, *Constructive Combinatorics*, Springer-Verlag, New York, 1986.
7. Herbert S. Wilf, *generatingfunctionology*, 2nd edition, Academic Press, San Diego, CA, 1994.

Math Bite:
Recitation of Large Primes

What is the largest prime for which you can recite all digits? Is it the Mersenne prime 8191, or the last prime-numbered year of the twentieth century (1999), or the Fermat prime 65537? The largest known repunit prime, which is the fifth of its kind, represents a case for easy memorization. With virtually no effort, all 1031 of its digits (each of which is 1) can be recited.

—RICHARD L. FRANCIS
SOUTHEAST MISSOURI STATE UNIVERSITY
CAPE GIRARDEAU, MO 63701

On Systems of Linear Diophantine Equations

FELIX LAZEBNIK
University of Delaware
Newark, DE 19716

Introduction Something happened to me recently I would wager has happened to many who read this note. Teaching a new topic, you cannot understand one of the proofs. Your first attempt to fill the gap fails. You look through your books for an answer. Next, you ask colleagues, go to the library, maybe even use the interlibrary loan. All in vain. Then it strikes you that, in fact, you cannot answer an even more basic and seemingly more interesting question. You peruse the books again. They seem to have answers to thousands of strange questions, but not to yours (the most natural one!). At the same time you cannot believe that your question could have been overlooked by generations of mathematicians. Days pass; the agony continues.

Then one day, some way or other, you find the answer. In my case the answer was in a book I already owned. It followed from a theorem I had known for a long time, but I had never thought of this particular application. I must admit, indeed, that this theorem appeared in almost *every* book I had checked, but never with a pointer to this particular application, even as an exercise. Were the authors unaware of the application? Or did it seem too obvious to mention? In any case, here is the story.

In my graduate combinatorics course, a proof of the existence of a design was based on the following question: Given a system of linear equations $A\mathbf{x} = \mathbf{b}$, where $A = (a_{i,j})$ is an $m \times n$ matrix with integer entries, and $\mathbf{b}$ is an $m \times 1$ column vector with integer components, does the system have an *integer* solution, i.e. an $n \times 1$ solution vector $\mathbf{x}$ with integer components? The suggested method ([7], Th. 15.6.5) makes use of "a well-known theorem of van der Waerden":

THEOREM (van der Waerden). *An integer solution of the system exists if and only if, for every row vector* $\mathbf{v}$ *with rational components such that* $\mathbf{v}A$ *has integer components,* $\mathbf{vb}$ *is an integer.*

I had never seen this theorem, and I was surprised that such a criterion could be useful (which it was!). In trying to prove the theorem, I realized that I did not know *any* good method for resolving a more basic question:

> *How can one tell whether a system of linear diophantine equations has a solution? If solutions exist, how can one find any or all of them?* (*)

I could not find this question in any of at least 30 modern texts on abstract algebra or number theory. The place I found it at last was the classical text of van der Waerden [**14**, Exercise 12.3]. Not for the first time this book contained an answer that I could not find in more recent sources—why hadn't I started with it? (Interestingly, the book contains very few exercises, but this one was there.)

The theory behind the solution is closely related to the famous structure theorem for finitely generated abelian groups, or, more generally, for finitely generated modules over principal ideal domains. Various proofs can be found in many books on abstract algebra, e.g., see [**8**]. We present a matrix version of the theorem. Let $\mathbb{Z}$ denote the ring of integers, $M_{m,n}(\mathbb{Z})$, $1 \le m \le n$, the ring of all integer $m \times n$

matrices, $SL_k(\mathbb{Z})$ the set of all square $k \times k$ matrices with integer entries and determinant 1 or -1 (*unimodular* matrices). By $D = \text{diag}(d_1, d_2, \ldots, d_m) \in M_{m,n}(\mathbb{Z})$ we denote the diagonal matrix that has an integer d_i in the (i,i) entry, $i = 1, \ldots, m$, and zeros elsewhere. Then we have:

THEOREM 1. *Let $A \in M_{m,n}(\mathbb{Z})$. There exist $L \in SL_m(\mathbb{Z})$ and $R \in SL_n(\mathbb{Z})$ such that*

$$LAR = D = \text{diag}(d_1, d_2, \ldots, d_s, 0, \ldots, 0),$$

where $d_i > 0$, $i = 1, \ldots, s$, and $d_i | d_{i+1}$, $i = 1, \ldots, s-1$.

A proof can be found, e.g., in **[14]** or **[8]**. The idea is to use elementary operations of rows and columns of A. Matrices L and R correspond to compositions of these operations. Though matrices L and R in Theorem 1 may vary, the matrix D is uniquely defined by A and it is called the *Smith normal form* of A.

Let us note immediately that Theorem 1 can be used to answer question (*). Given $A\mathbf{x} = \mathbf{b}$, rewrite it as $D\mathbf{y} = \mathbf{c}$ with $R\mathbf{y} = \mathbf{x}$, $LAR = D$ and $\mathbf{c} = L\mathbf{b}$. But the solution to the diagonal system $D\mathbf{y} = \mathbf{c}$ is easy. More details and a numerical example are given in the Applications section of this paper.

The question of finding an efficient algorithm for computing the Smith normal form of an integer matrix is not trivial. It is not clear that the direct application of elementary operations of rows and columns leads to a polynomial-time algorithm: it is conceivable that the integers get too large. For more details, see **[11]** and **[3]**.

Some history Theorem 1 has an interesting history: Question (*) seems not to have been asked, in full generality, until the mid-19th century. Its particular cases appeared in 1849–1850 in some number-theoretical studies of Hermite **[10**, p. 164, p. 265**]**. In 1858, Heger **[9]** formulated conditions for the solvability of $A\mathbf{x} = \mathbf{b}$ in the case where A has full rank (i.e., m) over $\mathbb{Z}$. In 1861, the problem was solved in full generality by H. J. S. Smith **[12]**. Theorem 1 appeared in a form close to the one above in an 1868 treatise by Frobenius **[5]** who generalized Heger's theorem **[5**, pp. 171–173**]**, and emphasized the unimodularity of the transformations **[5**, pp. 194–196**]**.

By then many important results on abelian groups had been discovered. Introduced by Gauss, the concept of an abelian group was developed both in number-theoretical studies of Gauss, Schering, Kronecker, and Dirichlet, and in the studies of elliptic functions and abelian integrals of Gauss, Abel, and Jacobi. Not until 1879 did Frobenius and Stickelberger **[6]** discover and use explicitly the connection between the theory of finitely generated abelian groups and Smith's theorem. In the same year, Frobenius showed that Smith's theory (extended to matrices over polynomial rings) could be used to classify square matrices over fields, up to similarity. (For further history, see **[4]** and the Historical Notes in **[2]**.) The story reminds us, in particular, that many basic notions and facts of linear algebra (including module theory) were developed within the context of number theory.

Applications Our first application is related to question (*). It also contains a proof of the aforementioned theorem of van der Waerden. Let $\mathbb{Q}$ denote the field of rational numbers.

PROPOSITION 2. *Let A, L, R, D be as in Theorem 1, $\mathbf{b} \in \mathbb{Z}^n$ and $\mathbf{c} = L\mathbf{b}$. Then the following four statements are equivalent:*

(1) *The system of linear equations $A\mathbf{x} = \mathbf{b}$ has an integer solution*
(2) *The system of linear equations $D\mathbf{y} = \mathbf{c}$ has an integer solution*

(3) *For every rational vector* $\mathbf{u}$ *such that* $\mathbf{u}A$ *is an integer vector, the number* $\mathbf{ub}$ *is an integer*

(4) *For every rational vector* $\mathbf{v}$ *such that* $\mathbf{v}D$ *is an integer vector, the number* $\mathbf{vc}$ *is an integer.*

Proof. (1) $\Leftrightarrow$ (2): Indeed, $A\mathbf{x} = \mathbf{b} \Leftrightarrow (L^{-1}DR^{-1})\mathbf{x} = \mathbf{b} \Leftrightarrow D(R^{-1}\mathbf{x}) = \mathbf{c} \Leftrightarrow D\mathbf{y} = \mathbf{c}$, where $\mathbf{y} = R^{-1}\mathbf{x}$. Since $R \in SL_m(\mathbb{Z})$, then $R^{-1} \in SL_m(\mathbb{Z})$. Therefore $\mathbf{x} \in \mathbb{Z}^n \Leftrightarrow \mathbf{y} = R^{-1}\mathbf{x} \in \mathbb{Z}^n$.

(3) $\Leftrightarrow$ (4): Indeed, $\mathbf{v}D \in \mathbb{Z}^n \Leftrightarrow \mathbf{v}(LAR) \in \mathbb{Z}^n \Leftrightarrow (\mathbf{v}L)AR \in \mathbb{Z}^n \Leftrightarrow (\mathbf{v}L)A \in \mathbb{Z}^nR^{-1} = \mathbb{Z}^n \Leftrightarrow \mathbf{u}A \in \mathbb{Z}^n$, where $\mathbf{u} = \mathbf{v}L$. $L \in SL_n(\mathbb{Z})$, then $\mathbf{u} \in \mathbb{Q}^m \Leftrightarrow \mathbf{v} \in \mathbb{Q}^m$, and, by (3), $\mathbf{ub} \in \mathbb{Z}$. But $\mathbf{ub} \in \mathbb{Z} \Leftrightarrow (\mathbf{v}L)(L^{-1}\mathbf{c}) \in \mathbb{Z} \Leftrightarrow \mathbf{vc} \in \mathbb{Z}$. Therefore (3) implies (4). Reversing the order of the argument, we get $\mathbf{u}A \in \mathbb{Z}^n \Leftrightarrow \mathbf{v}D \in \mathbb{Z}^n$ and $\mathbf{vc} \in \mathbb{Z} \Leftrightarrow \mathbf{ub} \in \mathbb{Z}$. Therefore (4) implies (3).

(2) $\Leftrightarrow$ (4): $D\mathbf{y} = \mathbf{c}$ implies $\mathbf{v}(D\mathbf{y}) = \mathbf{vc}$ for every $\mathbf{v} \in \mathbb{Q}^m$, hence $(\mathbf{v}D)\mathbf{y} = \mathbf{vc}$. If $\mathbf{v}D \in \mathbb{Z}^n$, then $\mathbf{vc} \in \mathbb{Z}$. Thus (2) implies (4). In order to prove that (4) implies (2), first we observe that $\mathbf{c} = (c_1, \ldots, c_s, 0, \ldots, 0)$. For suppose $c_j \neq 0$, $j > s$. Consider $\mathbf{v} = (0, \ldots, 0, 1/(2c_j), 0, \ldots, 0)$ where $1/(2c_j)$ appears in the j-th position. Since $\mathbf{v}D = \mathbf{0} \in \mathbb{Z}^n$, then by (4) $\mathbf{vc} = 1/2 \in \mathbb{Z}$, and we arrive at a contradiction. Thus $c_j = 0$ for $j > s$. Next, for $i = 1, \ldots, s$, we consider vectors $\mathbf{v_i} = (0, \ldots, 0, 1/d_i, 0, \ldots, 0)$. Since $\mathbf{v_i}D \in \mathbb{Z}^n$, then by (4), $\mathbf{v_i c} \in \mathbb{Z}$ and hence $c_i/d_i \in \mathbb{Z}$. Let $\mathbf{y} = (y_1, \ldots, y_s, 0, \ldots, 0)$, where $y_i = c_i/d_i$, $i = 1, \ldots, s$. Then $\mathbf{y} \in \mathbb{Z}^n$, and $D\mathbf{y} = \mathbf{c}$. ∎

With notations as in Proposition 2, one can reduce the solution of the system $A\mathbf{x} = \mathbf{b}$ to a solution of $D\mathbf{y} = \mathbf{c}$ by performing elementary transformations (over $\mathbb{Z}$) of rows and columns of matrix A augmented by vector $\mathbf{b}$. Matrices L and R can be constructed by multiplying matrices corresponding to these transformations. System $D\mathbf{y} = \mathbf{c}$ has a solution if and only if $c_{s+1} = \cdots = c_m = 0$, and $d_i | c_i$ for $i = 1, \ldots, s$. A general solution of $D\mathbf{y} = \mathbf{c}$ can be given in the form $\mathbf{y} = (y_1, \ldots, y_s, t_1 \ldots, t_{m-s})$, where $y_i = c_i/d_i$, $i = 1, \ldots, s$, and $t_1, \ldots, t_{m-s}$ are free integer parameters. Then the general solution of $A\mathbf{x} = \mathbf{b}$ is just $R\mathbf{y}$. Clearly, we may assume that each equation is reduced by the greatest common divisor of the coefficients of the variables.

EXAMPLE. Solve the system of diophantine equations $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{pmatrix} 2 & 1 & 4 \\ -5 & 2 & 6 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad \text{and} \quad \mathbf{b} = \begin{pmatrix} 17 \\ -13 \end{pmatrix}.$$

Solution. Consider a sequence of elementary transformations of rows and columns of A. It is well known that they can be achieved by multiplying A by unimodular matrices. Let us represent the transformation of rows by 2×2 matrices L_i and the ones of columns by 3×3 matrices R_j, where the lower indices reflect the order of multiplications. We consider the following transformations (matrices):

$$R_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad R_2 = \begin{pmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad R_3 = \begin{pmatrix} 1 & 0 & -4 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$L_4 = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}, \quad R_5 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -5 & 1 \end{pmatrix}, \quad R_6 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}.$$

Let $L = L_4$ and $R = R_1R_2R_3R_5R_6$. Then

$$D = LAR = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}\begin{pmatrix} 2 & 1 & 4 \\ -5 & 2 & 6 \end{pmatrix}\begin{pmatrix} 0 & 1 & 2 \\ 1 & 18 & 32 \\ 0 & -5 & -9 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$

and

$$\mathbf{c} = L\mathbf{b} = \begin{pmatrix} 17 \\ -47 \end{pmatrix}.$$

Solving $D\mathbf{y} = \mathbf{c}$, and taking $\mathbf{x} = R\mathbf{y}$, we get

$$\mathbf{x} = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 18 & 32 \\ 0 & -5 & -9 \end{pmatrix}\begin{pmatrix} 17 \\ -47 \\ t_1 \end{pmatrix} = \begin{pmatrix} -47 + 2t_1 \\ -829 + 32t_1 \\ 235 - 9t_1 \end{pmatrix}, \quad t_1 \in \mathbb{Z},$$

and the problem is solved. ∎

Another application is concerned with a special instance of the following fundamental question in number theory. Let $\mathbb{Z}[x_1, \ldots, x_t]$ denote the ring of polynomials in t variables with integral coefficients, and let $F(\mathbf{x}) \in \mathbb{Z}[x_1, \ldots, x_t]$. It is clear that if the equation $F(\mathbf{x}) = 0$ has an integer solution, then for any integer $n \geq 1$, the congruence $F(\mathbf{x}) \equiv 0 \pmod{n}$ has a solution. The converse, in general, is false, even for the case of one variable. A simple counterexample is provided by $F(x) = (2x + 1)(3x + 1)$. To show that $(2x + 1)(3x + 1) \equiv 0 \pmod{n}$ has a solution, write n in the form $n = 2^a3^bm$, where $\gcd(m, 2) = \gcd(m, 3) = 1$, and a and b are nonnegative integers. Then use the Chinese Remainder Theorem. For more on the relation between congruences and equations see, e.g., [1]. Nevertheless the following is valid.

PROPOSITION 3. *Let $A \in M_{m,n}(\mathbb{Z})$, and $\mathbf{b} \in \mathbb{Z}^n$. Then the system of linear equations $A\mathbf{x} = \mathbf{b}$ has an integer solution if and only if the corresponding system of congruences $A\mathbf{x} \equiv \mathbf{b} \pmod{n}$ has a solution for every positive integer n.*

Proof. Obviously, the first statement implies the second. Suppose the system of congruences has a solution for every positive integer n. Let L, R, D, $\mathbf{y}$ and $\mathbf{c}$ be as in Proposition 2, and let $N \in \mathbb{Z}$ be such that the transition from $A\mathbf{x} = \mathbf{b}$ to $D\mathbf{y} = \mathbf{c}$ uses integers with absolute values smaller than N. Then for every $n \geq N$, $A\mathbf{x} \equiv \mathbf{b} \pmod{n}$ $\Leftrightarrow D\mathbf{y} \equiv \mathbf{c} \pmod{n} \Leftrightarrow d_i y_i \equiv c_i \pmod{n}$, $i = 1, \ldots, s$. The latter system of congruences is solvable in particular when n is a multiple of d_s. Since $d_i | d_s$ for every i, $1 \leq i \leq s$, this implies $d_i | (d_i y_i - c_i)$, hence $d_i | c_i$ for all $i = 1, \ldots, s$. Therefore $D\mathbf{y} = \mathbf{c}$ has an integer solution, and so does $A\mathbf{x} = \mathbf{b}$. ∎

The following statement allows one easily to compute the index of a subgroup of the additive group $\mathbb{Z}^n$, when the index is finite.

PROPOSITION 4. *Let f: $\mathbb{Z}^n \to \mathbb{Z}^n$ be a $\mathbb{Z}$-linear map and $A \in M_{n,n}(\mathbb{Z})$ be its matrix with respect to some choice of bases. Suppose A has rank n. Then the index of $f(\mathbb{Z}^n)$ in $\mathbb{Z}^n$ is equal to* $|\det A|$.

Proof. By Theorem 1 we can find two unimodular matrices L and R such that $LAR = D = \operatorname{diag}(d_1, d_2, \ldots, d_n)$. Since A is of rank n, all $d_i \neq 0$. Therefore the abelian group $f(\mathbb{Z}^n) \cong d_1\mathbb{Z} \oplus d_2\mathbb{Z} \oplus \cdots \oplus d_n\mathbb{Z}$, and the order of $\mathbb{Z}^n/f(\mathbb{Z}^n)$ is $|d_1d_2 \ldots d_n| = |\det D|$. Since L and R are unimodular, $|\det D| = |\det A|$. ∎

EXAMPLE. Let f:$\mathbb{Z}^2 \to \mathbb{Z}^2$ be defined by $f((x, y)) = (28x + 38y, 12x + 16y)$. Choosing both bases to be the standard basis of $\mathbb{Z}^2$, we get $A = \begin{pmatrix} 28 & 12 \\ 38 & 16 \end{pmatrix}$. Therefore

the index $[\mathbb{Z}^2: f(\mathbb{Z}^2)]$ is equal to $|\det A| = 8$. The Smith normal form of A is $D = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}$, hence $f(\mathbb{Z}^2) \cong 2\mathbb{Z} \oplus 4\mathbb{Z}$.

Our next application is related to Proposition 4. It deals with some basic notions of the geometric number theory. Let $\mathbf{R}$ denote the field of real numbers, and $S = \{\mathbf{s}_1, \ldots, \mathbf{s}_m\}$ be a linearly independent set of vectors in $\mathbf{R}^n$. The additive subgroup $L = \langle S \rangle$ of $\mathbf{R}^n$ generated by S is called the *lattice* generated by S. A *fundamental domain* $T = T(S)$ of the lattice L is defined as

$$T = \left\{ \sum_{1 \le i \le m} x_i \mathbf{s}_i : 0 \le x_i < 1,\ x_i \in \mathbf{R} \right\}.$$

The *volume* $v(T)$ of T is defined in the usual way, as the square root of the absolute value of the determinant of an $m \times m$ matrix whose i-th row is the coordinate vector of $\mathbf{s}_i$ in the standard basis. Though T itself depends on a particular set of generators of L, the volume of T does not!

PROPOSITION 5. *Let* $S = \{\mathbf{s}_1, \ldots, \mathbf{s}_m\}$ *and* $U = \{\mathbf{u}_1, \ldots, \mathbf{u}_t\}$ *be two sets of linearly independent vectors which generate the same lattice* L. *Then* $m = t$ *and* $v(T(S)) = v(T(U))$.

Proof. We leave it to the reader. In case of difficulties, look through [**13**, pp. 30–33 and pp. 168–169]. ■

If one considers A with entries from a field, then by elementary operations of rows and columns, A can be brought to a diagonal form. It is a trivial exercise to check that an elementary row (column) operation preserves the dimensions of both row and column spaces of A. Therefore matrices LAR and A have equal dimensions of their row spaces and equal dimensions of their column spaces. Since the dimensions of row space and column space for a diagonal matrix are equal, we have a proof of the following fundamental result.

PROPOSITION 6. *The dimension of the row space of a matrix with entries from a field is equal to the dimension of its column space.* ■

Acknowledgement. References [3], [11], and remarks concerning the algorithmic aspects of finding the Smith normal form of an integer matrix were kindly suggested to the author by an anonymous referee. I am also very grateful to Gary Ebert, Todd Powers, David Saunders, Andrew Woldar, the editor, and referees, whose numerous comments substantially improved the original version of this paper.

REFERENCES

1. Z. I. Borevich and I. R. Sharfarevitch, *Number Theory*, Academic Press, 1966.
2. N. Bourbaki, *Elements of Mathematics: Algebra I, Chapters 1–3*, Hermann, Paris, 1974;
   N. Bourbaki, *Elements of Mathematics: Algebra II, Chapters 4–7*, Springer-Verlag, Berlin and New York, 1989.
3. T.-W. J. Chou and G. E. Collins, Algorithms for the solution of systems of linear Diophantine equations, *SIAM J. Computing* 11 (1982), 687–708.
4. L. E. Dickson, *History of the Theory of Numbers, Volume 2*, G. E. Stechert & Co., New York, 1934.
5. G. Frobenius, Theorie der linearen Formen mit ganzen Coefficienten, *Jour. für Math.*, 86 (1878), 146–208.
6. G. Frobenius und L. Stickelberger, Über Gruppen von Vertauschbaren Elementen, *J. de Crelle* LXXXVI, (1879), 217.
7. M. Hall, Jr., *Combinatorial Theory*, Second Ed., John Wiley & Sons, New York, 1986.
8. N. Jacobson, *Basic Algebra I*, W. H. Freeman and Co., San Francisco, 1974.

9. I. Heger, Denkschriften Acad. Wiss. Wien (Math. Nat.), 14 II (1858), 1–122.
10. Ch. Hermite, *Œuvres, t. I*, Gauthier-Villars, Paris, 1905.
11. R. Kannan and A. Bachem, Polynomial time algorithms to compute Hermite and Smith normal forms of an integer matrix, *SIAM J. Computing*, 8 (1979), 499–507.
12. H. J. S. Smith, On systems of linear indeterminate equations and congruences, p. 367, in *Collected Mathematical Papers*, vol. I, 367–409, Oxford, 1894. (= *Phil. Trans. London*, 151 (1861), 293–326).
13. I. N. Stewart and D. O. Tall, *Algebraic Number Theory*, Second Edition, Chapman and Hall, New York, 1987.
14. B. L. van der Waerden, *Algebra, Volume 2*, Frederick Ungar Publishing Co., New York, 1970.

The Golden Ratio is Less Than $\pi^2/6$

JAMES D. HARPER
Central Washington University
Ellensburg, WA 98926

As a mathematics teacher, I am pleased when an example turns out particularly neat and tidy. Occasionally, as a bonus, the example reveals an unexpected relationship. The relationship in the title of this note is not as unexpected or striking as, say, the implications of the Riemann hypothesis. Indeed, a hand-held calculator will convince anyone the statement is true. What is unexpected, beyond the serendipity of discovering this inequality as I worked out an example for a graduate analysis class, is that the proof centers on the Cauchy-Schwarz inequality.

Algebraically, the golden ratio, ϕ, is the larger root of the equation $x^2 - x = 1$; numerically, $\phi = (1 + \sqrt{5})/2 \approx 1.6180$. The other number in the title is, as Euler discovered, the sum of the squares of the harmonic sequence: $1/1^2 + 1/2^2 + 1/3^2 + \ldots$.

Recall that the space of all square-summable real sequences is an inner product space with the usual "dot product":

$$(x_1, x_2, x_3, \ldots) \cdot (y_1, y_2, y_3, \ldots) = x_1 y_1 + x_2 y_2 + x_3 y_3 + \ldots .$$

The Cauchy-Schwarz inequality guarantees that this inner product exists: For all vectors X and Y, $(X \cdot Y)^2 \le \|X\|^2 \|Y\|^2$; equality occurs if and only if one vector is a scalar multiple of the other.

My example begins with the harmonic sequence $X = (1, 1/2, 1/3, \ldots)$ and its cousin $Y = (1/2, 1/3, 1/4, \ldots)$. Both sequences are square-summable, with respective sums $S = \pi^2/6$ and $S - 1$. Now, by the Cauchy-Schwarz inequality,

$$\left(\sum_{n=1}^{\infty} \frac{1}{n(n+1)} \right)^2 = |X \cdot Y|^2 < \|X\|^2 \|Y\|^2 = S(S-1).$$

The series on the left is the classic telescoping example, with sum 1. Therefore, $1^2 < S(S-1)$, and completing the square gives: $5/4 < (S - 1/2)^2$. The desired inequality: $(1 + \sqrt{5})/2 < \pi^2/6$ now follows immediately.

Another surprise is how close these two numbers are to each other; to four decimals, $\phi = 1.6180 < 1.6449 = \pi^2/6$. Although our sequence vectors are not equal, they are "almost" equal in that the limit of the ratio of their terms is 1.

9. I. Heger, Denkschriften Acad. Wiss. Wien (Math. Nat.), 14 II (1858), 1–122.
10. Ch. Hermite, *Œuvres, t. I*, Gauthier-Villars, Paris, 1905.
11. R. Kannan and A. Bachem, Polynomial time algorithms to compute Hermite and Smith normal forms of an integer matrix, *SIAM J. Computing*, 8 (1979), 499–507.
12. H. J. S. Smith, On systems of linear indeterminate equations and congruences, p. 367, in *Collected Mathematical Papers*, vol. I, 367–409, Oxford, 1894. (= *Phil. Trans. London*, 151 (1861), 293–326).
13. I. N. Stewart and D. O. Tall, *Algebraic Number Theory*, Second Edition, Chapman and Hall, New York, 1987.
14. B. L. van der Waerden, *Algebra, Volume 2*, Frederick Ungar Publishing Co., New York, 1970.

The Golden Ratio is Less Than $\pi^2/6$

JAMES D. HARPER
Central Washington University
Ellensburg, WA 98926

As a mathematics teacher, I am pleased when an example turns out particularly neat and tidy. Occasionally, as a bonus, the example reveals an unexpected relationship. The relationship in the title of this note is not as unexpected or striking as, say, the implications of the Riemann hypothesis. Indeed, a hand-held calculator will convince anyone the statement is true. What is unexpected, beyond the serendipity of discovering this inequality as I worked out an example for a graduate analysis class, is that the proof centers on the Cauchy-Schwarz inequality.

Algebraically, the golden ratio, ϕ, is the larger root of the equation $x^2 - x = 1$; numerically, $\phi = (1 + \sqrt{5})/2 \approx 1.6180$. The other number in the title is, as Euler discovered, the sum of the squares of the harmonic sequence: $1/1^2 + 1/2^2 + 1/3^2 + \ldots$.

Recall that the space of all square-summable real sequences is an inner product space with the usual "dot product":

$$(x_1, x_2, x_3, \ldots) \cdot (y_1, y_2, y_3, \ldots) = x_1 y_1 + x_2 y_2 + x_3 y_3 + \ldots .$$

The Cauchy-Schwarz inequality guarantees that this inner product exists: For all vectors X and Y, $(X \cdot Y)^2 \le \|X\|^2 \|Y\|^2$; equality occurs if and only if one vector is a scalar multiple of the other.

My example begins with the harmonic sequence $X = (1, 1/2, 1/3, \ldots)$ and its cousin $Y = (1/2, 1/3, 1/4, \ldots)$. Both sequences are square-summable, with respective sums $S = \pi^2/6$ and $S - 1$. Now, by the Cauchy-Schwarz inequality,

$$\left(\sum_{n=1}^{\infty} \frac{1}{n(n+1)} \right)^2 = |X \cdot Y|^2 < \|X\|^2 \|Y\|^2 = S(S-1).$$

The series on the left is the classic telescoping example, with sum 1. Therefore, $1^2 < S(S-1)$, and completing the square gives: $5/4 < (S - 1/2)^2$. The desired inequality: $(1 + \sqrt{5})/2 < \pi^2/6$ now follows immediately.

Another surprise is how close these two numbers are to each other; to four decimals, $\phi = 1.6180 < 1.6449 = \pi^2/6$. Although our sequence vectors are not equal, they are "almost" equal in that the limit of the ratio of their terms is 1.

A Proof in the Spirit of Zeilberger of an Amazing Identity of Ramanujan

M. D. HIRSCHHORN
University of New South Wales
Sydney 2052, NSW, Australia

1. Introduction In a recent paper [1], I discussed the following statement, to be found in Ramanujan's lost notebook [2]:

If

$$\sum_{n\geq 0} a_n x^n = \frac{1+53x+9x^2}{1-82x-82x^2+x^3},$$

$$\sum_{n\geq 0} b_n x^n = \frac{2-26x-12x^2}{1-82x-82x^2+x^3},$$

and

$$\sum_{n\geq 0} c_n x^n = \frac{2+8x-10x^2}{1-82x-82x^2+x^3},$$

then

$$a_n^3 + b_n^3 = c_n^3 + (-1)^n. \tag{C}$$

I proved this statement and showed how Ramanujan may have discovered it. In proving the statement I found explicit expressions for a_n, b_n and c_n, and verified the conclusion (C). In this paper I show that in order to prove Ramanujan's statement it is sufficient to check just the first seven cases, and then I do so. This proof is in the spirit of Zeilberger [3].

2. Checking the first seven cases is sufficient Each of $\{a_n\}$, $\{b_n\}$ and $\{c_n\}$ is generated by

$$\frac{N(x)}{D(x)}$$

where

$$\begin{aligned} D(x) &= 1-82x-82x^2+x^3 \\ &= (1-83x+x^2)(1+x) \\ &= (1-\alpha x)(1-\beta x)(1-\gamma x) \end{aligned}$$

with

$$\gamma = -1, \quad \alpha+\beta = 83, \quad \alpha\beta = 1$$

(note that α, β, and γ are distinct) and where $N(x)$ is a quadratic in x which depends on the sequence under consideration.

It follows by the method of partial fractions that each of a_n, b_n and c_n can be written as a linear combination of α^n, β^n, and γ^n. (This was done explicitly in [1].)

So, each of a_n^3, b_n^3 and c_n^3 can be written as a linear combination of the seven quantities

$$\alpha^{3n},\ \beta^{3n},\ (\alpha^2\gamma)^n = (-\alpha^2)^n,\ (\beta^2\gamma)^n = (-\beta^2)^n$$
$$(\alpha^2\beta)^n = (\alpha\gamma^2)^n = \alpha^n,\ (\alpha\beta^2)^n = (\beta\gamma^2)^n = \beta^n,$$
$$\text{and } \gamma^{3n} = (\alpha\beta\gamma)^n = (-1)^n.$$

Thus, each of $\{a_n^3\}$, $\{b_n^3\}$, and $\{c_n^3\}$ is generated by $N(x)/D(x)$ where $D(x)$ is the polynomial of degree seven,

$$D(x) = (1-\alpha^3x)(1-\beta^3x)(1+\alpha^2x)(1+\beta^2x)(1-\alpha x)(1-\beta x)(1+x)$$
$$= 1+\cdots+x^7,$$

and where $N(x)$ is a polynomial of degree at most six, and depends on the sequence under consideration.

It follows that

$$\sum_{n\geq 0}\left(a_n^3+b_n^3-c_n^3-(-1)^n\right)x^n = \frac{N(x)}{D(x)}$$

where $N(x)$ is a polynomial of degree at most six, and where $D(x)$ is as above.

Let $q_n = a_n^3+b_n^3-c_n^3-(-1)^n$, and suppose that $N(x) = d_0+d_1x+\cdots+d_6x^6$. Then

$$\sum_{n\geq 0} q_n x^n = \frac{d_0+d_1x+\cdots+d_6x^6}{D(x)}.$$

Note that if $q_0 = 0$ then $d_0 = 0$; if $q_0 = 0$ and $q_1 = 0$ then $d_0 = 0$ and $d_1 = 0$, and so on, and if $q_0,\ldots,q_6$ are all 0, then $d_0,\ldots,d_6$ are all 0, $N(x)$ is the zero polynomial, and $q_n = 0$ for all n.

In other words, if (C) is true for $n = 0,1,\ldots,6$, then (C) is true for all n.(!)

3. Checking the first seven cases We have

$$a_0^3+b_0^3-c_0^3-1 = 1^3+2^3-2^3-1 = 1+8-8-1 = 0$$

$$a_1^3+b_1^3-c_1^3+1 = 135^3+138^3-172^3+1 = 2460375+2628072-5088448+1 = 0$$

$$a_2^3+b_2^3-c_2^3-1 = 11161^3+11468^3-14258^3-1$$
$$= 1390302566281+1508214295232-2898516861512-1 = 0$$

$$a_3^3+b_3^3-c_3^3+1 = 926271^3+951690^3-1183258^3+1$$
$$= 794720108027000511+861958819711809000$$
$$-1656678927738809512+1 = 0$$

$$a_4^3+b_4^3-c_4^3-1 = 76869289^3+78978818^3-98196140^3-1$$
$$= 4542119879291901383845 69+4926425157409745091594 32$$
$$-9468545036701646475440 00-1 = 0$$

$$a_5^3+b_5^3-c_5^3+1 = 6379224759^3+6554290188^3-8149096378^3+1$$
$$= 2595994163433662399084126 77479$$
$$+2815639161232358998768839 24672$$
$$-5411633324666021397852966 02152+1 = 0$$

and

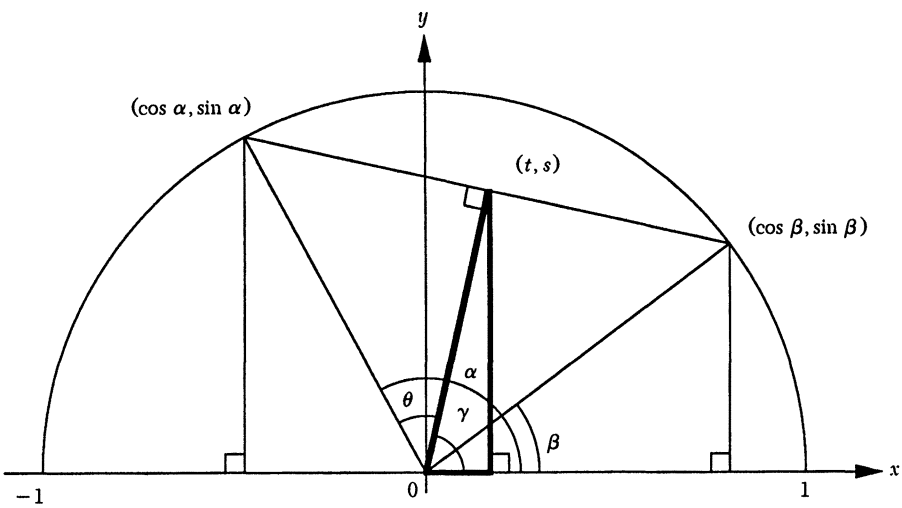
$$\begin{aligned} a_6^3 + b_6^3 - c_6^3 - 1 &= 529398785665^3 + 543927106802^3 - 676276803218^3 - 1 \\ &= 1483709311818771712048818272589546 25 \\ &\quad + 1609244775067813934836090651947216 08 \\ &\quad - 3092954086886585646884908924536762 32 - 1 = 0. \end{aligned}$$

Thus Ramanujan's statement is proved.

REFERENCES

1. M. D. Hirschhorn, An amazing identity of Ramanujan, this MAGAZINE 68 (1995), 199–201.
2. S. Ramanujan, The Lost Notebook and Other Unpublished Papers, New Delhi, Narosa, 1988.
3. D. Zeilberger, "=", Invited address, 896th meeting of the AMS, University of Richmond, Richmond, Virginia, 11th–13th November 1994.
4. D. Zeilberger, The joy of brute force, http://www.math.temple.edu/~zeilberg (click on articles), 1995.

Proof Without Words: The Sum-Product Identities



$$\theta = \frac{\alpha - \beta}{2} \qquad \gamma = \frac{\alpha + \beta}{2}$$

$$\frac{\sin\alpha + \sin\beta}{2} = s = \cos\frac{\alpha - \beta}{2}\sin\frac{\alpha + \beta}{2}$$

$$\frac{\cos\alpha + \cos\beta}{2} = t = \cos\frac{\alpha - \beta}{2}\cos\frac{\alpha + \beta}{2}$$

—SIDNEY H. KUNG
JACKSONVILLE UNIVERSITY
JACKSONVILLE, FL 32211

and

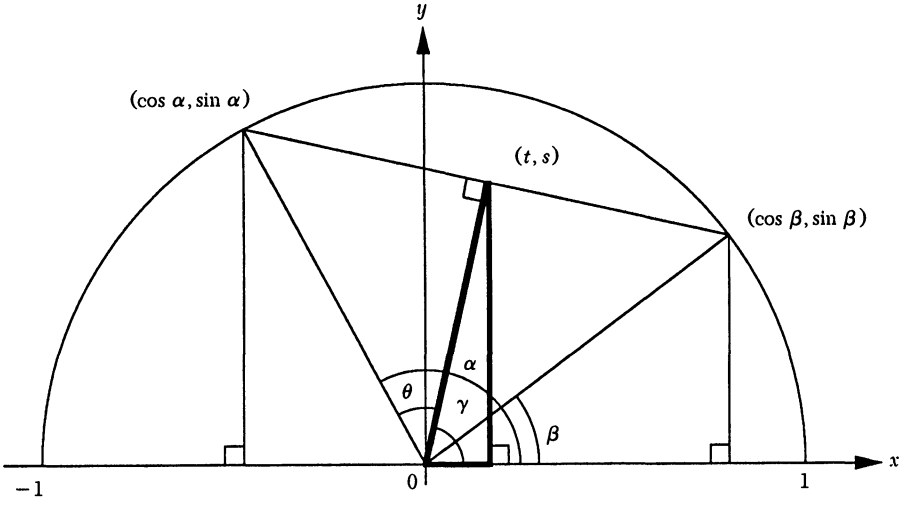
$$\begin{aligned} a_6^3 + b_6^3 - c_6^3 - 1 &= 529398785665^3 + 543927106802^3 - 676276803218^3 - 1 \\ &= 1483709311818771712048818272589546 25 \\ &\quad + 160924477506781393483609065194721608 \\ &\quad - 309295408688658564688490892453676232 - 1 = 0. \end{aligned}$$

Thus Ramanujan's statement is proved.

REFERENCES

1. M. D. Hirschhorn, An amazing identity of Ramanujan, this MAGAZINE 68 (1995), 199–201.
2. S. Ramanujan, The Lost Notebook and Other Unpublished Papers, New Delhi, Narosa, 1988.
3. D. Zeilberger, "=", Invited address, 896th meeting of the AMS, University of Richmond, Richmond, Virginia, 11th–13th November 1994.
4. D. Zeilberger, The joy of brute force, http://www.math.temple.edu/~zeilberg (click on articles), 1995.

Proof Without Words: The Sum-Product Identities



$$\theta = \frac{\alpha - \beta}{2} \qquad \gamma = \frac{\alpha + \beta}{2}$$

$$\frac{\sin\alpha + \sin\beta}{2} = s = \cos\frac{\alpha - \beta}{2}\sin\frac{\alpha + \beta}{2}$$

$$\frac{\cos\alpha + \cos\beta}{2} = t = \cos\frac{\alpha - \beta}{2}\cos\frac{\alpha + \beta}{2}$$

—SIDNEY H. KUNG
JACKSONVILLE UNIVERSITY
JACKSONVILLE, FL 32211

Maximizing the Product of Summands; Minimizing the Sum of Factors

EUGENE F. KRAUSE
University of Michigan
Ann Arbor, MI 48109-1003

Introduction The spirit of the times urges all of us, at whatever level we teach, to promote mathematical investigation and exploration by our students. That being the case, it is particularly important for future teachers to have the experience of participating in mathematical research in at least one of their college courses. This paper outlines an extended investigation that was carried out, as a group project, by two classes of prospective secondary teachers. A good deal of instructor guidance was necessary, and many of the details of proof had to be skimmed over, but enough of the work was left to the students so that they had a sense of participating in the creation of new mathematical knowledge and of behaving like mathematicians.

The first problem The problem that launched the project was brought to my attention by Professor Nic Heideman of Rhodes University, Grahamstown, South Africa. It is intended for 12-year-olds.

> Given a positive integer k, find positive integers $x_1, x_2, \ldots, x_n$ that sum to k and have maximal product.

For example, here are two failed choices for the x_i in the case $k = 14$.

$$14 = 2 + 3 + 4 + 5 \rightarrow 2 \cdot 3 \cdot 4 \cdot 5 = 120$$
$$14 = 3 + 3 + 2 + 2 + 2 + 2 \rightarrow 3^2 \cdot 2^4 = 144$$

Further investigation of the case $k = 14$ leads to the formulation of some general principles: "Never use a 1." "Never use a number greater than 4." "A 4 can always be replaced by two 2s." "Two 3s are better than three 2s." And these principles, in turn, suggest a loosely framed algorithm: "Use as many threes as possible and make up the difference with twos." For school children the original problem can be viewed as solved at this point, but for college students there is still work to do. The whole matter of recasting the problem and its solution into mathematical language lies ahead.

DEFINITION 1. Given a positive integer k. A *partition* of k is a collection of (not necessarily distinct) positive integers $\{x_1, x_2, \ldots, x_n\}$ such that $\sum_{i=1}^{n} x_i = k$. A partition $\{x_1, x_2, \ldots, x_n\}$ of k is called a *winning partition* of k if $\prod_{i=1}^{n} x_i \geq \prod_{i=1}^{m} y_i$ for any other partition $\{y_1, y_2, \ldots, y_m\}$ of k.

In view of Definition 1, the original problem is simply this:

Problem 1. Given a positive integer k, find a winning partition of k.

The algorithmic solution we arrived at earlier can now be recast in the form of a theorem.

THEOREM 1. *Every positive integer $k > 1$ has a winning partition. If we agree to replace any 4 by two 2s, then each k has a unique winning partition. The winning partitions are as follows:*

- *If $k \equiv 0 \pmod 3$, then the winning partition of k consists of all 3s. The associated maximal product is $3^{k/3}$.*

- *If* $k \equiv 1 \pmod 3$, *then the winning partition of* k *consists of two* 2*s and the rest* 3*s. The maximal product is* $2^2 \cdot 3^{(k-4)/3}$.
- *If* $k \equiv 2 \pmod 3$, *then the winning partition of* k *consists of one* 2 *and the rest* 3*s. The maximal product is* $2 \cdot 3^{(k-2)/3}$.

Proof. The four general principles enunciated earlier constitute a proof. First, no winning partition can include the integer 1. Second, no winning partition can include an integer expressible in the form $a+b$ where $a > 2$ and $b \geq 2$, because replacing $a+b$ by a and b in the partition produces a larger associated product, since $a+b < a \cdot b$:

$$ab - b = (a-1)b \geq (a-1) \cdot 2 = a + (a-2) > a$$

Third, replacing 4 by two 2s in any partition leaves the associated product unchanged. Fourth, $2^3 < 3^2$.

The second problem There are at least two reasons why it is difficult to be content with Theorem 1 as the endpoint of this journey of exploration. First, the three-case nature of its conclusion is aesthetically somewhat unsatisfying. Second, upon re-examination, the restrictions in the original problem, that the number k and all of the summand/factors be positive *integers*, begin to appear almost "arbitrary and capricious." Would it not be more natural to relax those conditions and allow any positive *real numbers*?

DEFINITION 2. Given a positive real number k. A *real-partition* of k is a collection of (not necessarily distinct) positive real numbers $\{x_1, x_2, \ldots, x_n\}$ such that $\sum_{i=1}^{n} x_i = k$. A real-partition $\{x_1, x_2, \ldots, x_n\}$ is called a *winning real-partition* of k if $\prod_{i=1}^{n} x_i \geq \prod_{i=1}^{m} y_i$ for any other real-partition $\{y_1, y_2, \ldots, y_m\}$ of k.

The second problem, then, is simply this:

Problem 2. Given a positive real number k, find a winning real-partition of k.

Our strategy will be to look at all of the 2-number real-partitions of k and find the one with maximal product (the "winner" in the 2-number category), then find the winner in the 3-number category, then the winner in the 4-number category, etc. Finally we will look for the winner among winners. Lemma 1 describes the category winners.

LEMMA 1. *Given a positive real number* k *and a (fixed) positive integer* n. *Among all of the real-partitions of* k *that consist of* n *numbers, the winning real-partition (the one that yields maximal product) consists of* n *copies of* k/n.

Proof. The result follows from two facts. The first is that the (continuous) real-valued product function p defined on the (compact) set

$$D = \left\{ (x_1, x_2, \ldots, x_n) \,\middle|\, \sum_{i=1}^{n} x_i = k,\ x_i \geq 0 \text{ for all } i \right\}$$

by $p(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} x_i$ attains a maximum at some point of D. The second is that if $x_i \neq x_j$ for some i and j, then $(x_1, x_2, \ldots, x_i, \ldots, x_j, \ldots, x_n)$ does *not* maximize p: in view of the arithmetic/geometric mean inequality, replacing both x_i and x_j by their average produces an n-tuple that yields a greater value for p.

Example. According to Lemma 1, these are the winners in the 2-, 3-, 4-, and 5-number categories for $k = 10$:

$$\{10/2, 10/2\}, \{10/3, 10/3, 10/3\}, \{10/4, 10/4, 10/4, 10/4\},$$
$$\{10/5, 10/5, 10/5, 10/5, 10/5\}$$

The associated products are

$$(10/2)^2 = 25, \quad (10/3)^3 = 37.\overline{037}, \quad (10/4)^4 = 39.0625, \quad (10/5)^5 = 32$$

It appears that the winner among winners is $\{10/4, 10/4, 10/4, 10/4\}$.

What we need to do now is determine, for arbitrary k, which value of n yields the winner among the category winners

$$\{k/1\}\{k/2, k/2\}\{k/3, k/3, k/3\} \ldots \{k/n, k/n, \ldots, k/n\} \ldots$$

We begin by compiling a partial table, Table 1, of values of k and approximate values of the products $(k/n)^n$ associated with the first few category winners for k. For each k the largest product among winning products is circled. Patterns in the table suggest two key facts, which we formalize as Lemma 2 and Lemma 3.

TABLE 1

k	$(k/1)^1$	$(k/2)^2$	$(k/3)^3$	$(k/4)^4$	...
1	①	.25	.037	.003...	
2	②	1	.296	.0625	
3	③	2.25	1	.316...	
$2^2/1^1 = 4$	④	④	2.37	1	
5	5	(6.25)	4.63	2.44	
6	6	⑨	8	5.06	
$3^3/2^2 = 6.75$	6.75	(11.39...)	(11.39...)	8.10...	
7	7	12.25	(12.70...)	9.37...	
8	8	16	(18.96...)	16	
9	9	20.25	(27)	25.62...	
$4^4/3^3 = 9.\overline{481}$	$9.\overline{481}$	22.47...	(31.56...)	(31.56...)	
10	10	25	37.03...	(39.06...)	
⋮					

LEMMA 2. *The sequence* $2^2/1^1, 3^3/2^2, 4^4/3^3, 5^5/4^4, \ldots$ *is strictly increasing and has no upper bound.*

Proof. The sequence is unbounded because $n^n/(n-1)^{n-1} = [n/(n-1)]^{n-1} \cdot n$, which is greater than n. It is strictly increasing because the function of a real variable, $f(x) = x^x/(x-1)^{x-1}$, has positive derivative for $x > 1$.

Lemma 3 generalizes the observation that if, for example, k is between $3^3/2^2$ and $4^4/3^3$, then the winning product for k is found in column 3 of Table 1.

LEMMA 3. *For $k \geq 4$, if n is the positive integer such that*

$$n^n/(n-1)^{n-1} \leq k < (n+1)^{n+1}/n^n \tag{1}$$

then

$$(k/1)^1 < (k/2)^2 < (k/3)^3 < \cdots < [k/(n-1)]^{n-1} \leq (k/n)^n \tag{2}$$

and

$$(k/n)^n > [k/(n+1)]^{n+1} > [k/(n+2)]^{n+2} > \ldots \tag{3}$$

Furthermore,

$$(k/n)^n = [k/(n-1)]^{n-1} \text{ if and only if } n^n/(n-1)^{n-1} = k.$$

Proof. First note that, by Lemma 2, there is a unique positive integer n such that (1) holds. To establish (2), begin with

$$n^n/(n-1)^{n-1} \leq k \tag{4}$$

from (1) and apply Lemma 2 to deduce that $(i+1)^{i+1}/i^i < k$ for all $i = 1, 2, \ldots, n-2$. Replacing k by k^{i+1}/k^i in this inequality yields $(k/i)^i < [k/(i+1)]^{i+1}$ for all $i = 1, 2, \ldots, n-2$; that is

$$(k/1)^1 < (k/2)^2 < (k/3)^3 < \cdots < [k/(n-1)]^{n-1}$$

which is nearly (2). To establish the final inequality in (2), replace k by k^n/k^{n-1} in (4).

To establish (3), begin with the inequality $k < (n+1)^{n+1}/n^n$ from (1) and apply Lemma 2 to deduce that $k < (i+1)^{i+1}/i^i$ for all $i = n, n+1, n+2, \ldots$. Replacing k by k^{i+1}/k^i in this inequality yields $[k/(i+1)]^{i+1} < (k/i)^i$ for all $i = n, n+1, \ldots$, that is

$$(k/n)^n > [k/(n+1)]^{n+1} > [k/(n+2)]^{n+2} > \ldots$$

which is exactly (3). The "Furthermore" statement in Lemma 3 is a trivial exercise in algebra.

We now have only to summarize our work in a theorem that provides a complete answer to the second problem.

THEOREM 2. *Every positive real number k has a winning real-partition consisting of n copies of k/n where n is the least positive integer for which*

$$k < \frac{(n+1)^{n+1}}{n^n}$$

The associated maximal product is $(k/n)^n$. If $k > n^n/(n-1)^{n-1}$, then this winning real-partition is unique. If $k = n^n/(n-1)^{n-1}$, then there is a second winning real-partition: $(n-1)$ copies of $k/(n-1)$.

Example. Find the winning real-partition for $k = 50$. According to Theorem 2, we need to find the least positive integer n such that

$$50 < (n+1)^{n+1}/n^n \tag{5}$$

and then the winning real-partition will consist of n copies of $50/n$. Solving (5) looks like it will require a guess-and-check process, but if we analyze the expression in n a bit, we can make a good first guess. Notice that

$$(n+1)^{n+1}/n^n = [(n+1)/n]^n(n+1) = [1+1/n]^n(n+1)$$

and recall that $[1+(1/n)]^n$ approaches e as n gets large. Thus inequality (5) is roughly equivalent to $50 < e(n+1)$, so a reasonable first guess is $n \approx 50/e \approx 18$. Checking, we find that $19^{19}/18^{18} \approx 50.28$ and $18^{18}/17^{17} \approx 47.56$. Thus $n = 18$ is the least positive integer solution to (5). The winning real-partition consists of 18 copies of $50/18$, and the associated maximal product is $(50/18)^{18} \approx 96{,}951{,}601$. Compare this with the winning integer partition of 50, sixteen 3s and one 2, and its associated maximal product, $3^{16} \cdot 2^1 = 86{,}093{,}442$.

The third problem The example just completed provides a fascinating clue that we are getting close to something fundamental. The ubiquitous number e seems to play a key role. For the optimal real-partition of k, $\{k/n, k/n, \ldots, k/n\}$, the value of n is approximately k/e, and thus the repeated summand, k/n, is approximately e.

To decide how to formulate a third problem that will reveal the role of e, we look back at what we were doing in the final stages of solving Problem 2, but now we write repeated addition as multiplication and repeated multiplication as exponentiation.

> We were looking for a positive integer n and a positive real number r such that $r \cdot n = k$ and r^n is maximized.

The only restriction left to loosen is the restriction that n be an integer. That is, we are about to allow ourselves to think, for example, of 10 as "the sum of three-and-one-half $2\frac{6}{7}$s," and to evaluate "the product of three-and-one-half $2\frac{6}{7}$s."

DEFINITION 3. Given a positive real number k. A *pseudo-partition* of k is an ordered pair of positive real numbers (x, y) such that $x \cdot y = k$. A pseudo-partition (x, y) of k is a *winning pseudo-partition* of k if $x^y \geq u^v$ for any other pseudo-partition (u, v) of k.

For example, $(50/18, 18)$ is a pseudo-partition of 50, but it is not a winning pseudo-partition because the pseudo-partition $(e, 50/e)$ has a greater associated power, $e^{50/e} \approx 97{,}364{,}484$.

Problem 3. Given a positive real number k, find a winning pseudo-partition of k.

THEOREM 3. *Every positive real number k has a unique winning pseudo-partition, namely* $(e, k/e)$. *The associated maximal power is* $e^{k/e}$.

Proof. The task is to maximize the function $f(x, y) = x^y$ subject to the constraints $x \cdot y = k$, $x > 0$, $y > 0$. Substituting k/x for y into the formula for f yields a function of one variable, $g(x) = x^{k/x}$. The function g is maximized where its (easily calculated) derivative is zero, namely at $x = e$. Thus $x = e$ and $y = k/e$ maximize f.

Intermission When we compare Theorem 3 to Theorem 2 and Theorem 1 it appears that we have reached a satisfactory stopping point. The statement of Theorem 3 is a model of simplicity. There are no special cases, no technical inequalities, no exceptions to uniqueness. The proof of Theorem 3 is short and direct. And Theorem 3, while including neither Theorem 2 nor Theorem 1 as a special case, certainly illuminates them both. To anthropomorphize: The numbers in the winning integer

partition of 50 were all trying to be e, but were prevented by the two restrictions that (1) they had to be integers, and (2) they had to be integral in number. So most settled for 3 and some for 2. The numbers in the winning real-partition of 50 were all trying to be e, but were prevented by the single restriction that they had to be integral in number. So they all settled for $2.\overline{7}$. Once we allowed a "real number of real summands," all those "summands" could become e and the "associated product," now a power with real exponent, could achieve its highest value.

With a result as conclusive as Theorem 3 in hand, it is tempting to consider the project finished. But a glance back at the sum-product questions that we have answered immediately suggests a related set of "dual" questions in which the roles of sum and product are interchanged and the goal is to minimize rather than maximize. Our hope is that the new theorems answering these new questions will closely resemble old Theorems 1, 2, and 3.

The new first problem The dual of the original problem for 12-year olds is this: Given a positive integer k, find positive integers whose product is k and whose sum is minimal. To solve this problem we should follow a path parallel to the one that we traveled before. We begin with a definition of the multiplicative analog of (integer) partition.

DEFINITION 1′. Given a positive integer k. An *integer factor set* of k is a collection of (not necessarily distinct) positive integers $\{x_1, x_2, \ldots, x_n\}$ such that $\prod_{i=1}^{n} x_i = k$. An integer factor set $\{x_1, x_2, \ldots, x_n\}$ of k is called a *winning integer factor set* of k if $\sum_{i=1}^{n} x_i \le \sum_{i=1}^{m} y_i$ for any other integer factor set $\{y_1, y_2, \ldots, y_m\}$ of k.

Example. $\{2, 2, 2, 3, 3\}$ is a winning integer factor set of 72; $\{3, 4, 6\}$ is not.

Problem 1′. Given a positive integer k, find a winning integer factor set of k.

THEOREM 1′. *Every positive integer $k > 1$ has a winning integer factor set. If we agree to replace any 4 by two 2s, then each k has a unique winning integer factor set, namely the set of all prime factors (with repetitions) of k.*

Proof. Three readily proved observations constitute a proof. First, no winning integer factor set can include the integer 1. Second, no winning integer factor set can include a composite number $a \cdot b$ where $a > 2$ and $b \ge 2$. (Use the same argument as in the proof of Theorem 1.) Third, replacing 4 by two 2s in any integer factor set leaves the associated sum unchanged.

Notice that this new Theorem 1′ is *nothing at all* like old Theorem 1. Our hoped for duality has not yet materialized.

The new second problem Just as we did before, we now remove the restriction that the numbers in a factor set be integers.

DEFINITION 2′. Given a positive real number k. A *real factor set* of k is a collection of (not necessarily distinct) positive real numbers $\{x_1, x_2, \ldots, x_n\}$ such that $\prod_{i=1}^{n} x_i = k$. A real factor set $\{x_1, x_2, \ldots, x_n\}$ of k is called a *winning real factor set* of k if $\sum_{i=1}^{n} x_i \le \sum_{i=1}^{m} y_i$ for any other real factor set $\{y_1, y_2, \ldots, y_m\}$ of k.

Problem 2′. Given a positive real number k, find a winning real factor set of k.

As before, our strategy is to first determine the winner among all 2-number real factor sets, the winner among all 3-number real factor sets, the winner among all 4-number real factor sets, and so forth. Then we go on to find the winner among winners. As before, a simple lemma gives us the winners in each weight class.

LEMMA 1′. *Given a positive real number k and a (fixed) positive integer n. Among all of the real factor sets of k that consist of n numbers, the winning real factor set (the one that yields the minimal sum) consists of n copies of $k^{1/n}$.*

Proof. The lemma follows from two facts. The first is that no real factor set that contains two (or more) *distinct* numbers can have minimal sum. Again this is an immediate consequence of the inequality between geometric and arithmetic means. The second fact is that the continuous function $s(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} x_i$ does attain a minimum at some point of the set

$$D = \left\{ (x_1, x_2, \ldots, x_n) \,\middle|\, \prod_{i=1}^{n} x_i = k, \text{ all } x_i > 0 \right\}.$$

To prove this second fact requires some care, since the set D is not bounded. We omit the rather technical argument to save space.

Example. The winning real factor sets in the first four weight classes for $k = 10$ are these:

$\{10\}$ with associated sum 10

$\{10^{1/2}, 10^{1/2}\}$ with associated sum $2 \cdot 10^{1/2} \approx 6.32$

$\{10^{1/3}, 10^{1/3}, 10^{1/3}\}$ with associated sum $3 \cdot 10^{1/3} \approx 6.46$

$\{10^{1/4}, 10^{1/4}, 10^{1/4}, 10^{1/4}\}$ with associated sum $4 \cdot 10^{1/4} \approx 7.11$

It appears that the winner among winners is $\{10^{1/2}, 10^{1/2}\}$.

To seek out a pattern in the winning real factor sets for various values of k we proceed, as before, to compile a table (Table 2) of the sums associated with the different weight classes for small integral values of k. Winning sums are circled. It appears from the table that the winning real factor sets jump to the right as k increases. Of particular interest are the values of k at which these jumps (ties for winner) occur. The first is at $k = 4$. The second is at the value of k where $2k^{1/2} = 3k^{1/3}$, that is, at $k = (3/2)^{3 \cdot 2} \approx 11.39$. The third is where $3k^{1/3} = 4k^{1/4}$, that is, at $k = (4/3)^{4 \cdot 3} \approx 31.57$. Extending this pattern suggests strongly a theorem that solves Problem 2′.

TABLE 2

k	$1k^{1/1}$	$2k^{1/2}$	$3k^{1/3}$	$4k^{1/4}$	...
2	②	2.83	3.78	4.76	
3	③	3.46	4.33	5.26	
4	④	④	4.76	5.66	
5	5	(4.47)	5.13	5.98	
6	6	(4.90)	5.45	6.26	
⋮	⋮	⋮	⋮	⋮	
10	10	(6.32)	6.46	7.11	
11	11	(6.63)	6.67	7.28	
12	12	6.93	(6.87)	7.44	
⋮	⋮	⋮	⋮	⋮	

THEOREM 2′. *Every positive real number k has a winning real factor set consisting of n copies of $k^{1/n}$ where n is the least positive integer for which*

$$k < \left(\frac{n+1}{n}\right)^{(n+1)n}$$

The associated minimal sum is $nk^{1/n}$. If $k > (n/(n-1))^{n(n-1)}$, then this winning real factor set is unique. If $k = (n/(n-1))^{n(n-1)}$, then there is a second winning real factor set: $n-1$ copies of $k^{1/(n-1)}$.

The proof Theorem 2′ depends on two lemmas, Lemmas 2′ and 3′ below, that are analogous in statement to Lemmas 2 and 3. As was the case with Lemma 1′, the proofs of Lemmas 2′ and 3′ are trickier than were the proofs of Lemmas 2 and 3, and again, for reasons of space, we omit them.

LEMMA 2′. *The sequence $(2/1)^{2\cdot1}, (3/2)^{3\cdot2}, (4/3)^{4\cdot3}, \ldots$ is strictly increasing and has no upper bound.*

LEMMA 3′. *For $k \geq 4$, if n is the positive integer such that*

$$(n/(n-1))^{n(n-1)} \leq k < ((n+1)/n)^{(n+1)n}$$

then

$$k > 2k^{1/2} > 3k^{1/3} > \cdots > (n-1)k^{1/(n-1)} \geq nk^{1/n}$$

and

$$nk^{1/n} < (n+1)k^{1/(n+1)} < (n+2)k^{1/(n+2)} < \cdots$$

Furthermore

$$nk^{1/n} = (n-1)k^{1/(n-1)} \text{ if and only if } \left(\frac{n}{n-1}\right)^{n(n-1)} = k.$$

As we did with Theorem 2, we now apply Theorem 2′ to the case of a large k, say $k = 1000$. Only a few guesses are needed to find the least integer n for which $1000 < ((n+1)/n)^{(n+1)n}$, namely $n = 7$. Thus, by Theorem 2′, the winning real factor set consists of 7 copies of $1000^{1/7}$ and the associated sum is $7 \times 1000^{1/7} \approx 7 \times 2.68270 \approx 18.77887$. We notice two things about this example. First, the sum associated with the winning real factor set is less than the sum, 21, associated with the winning integer factor set, $\{2, 2, 2, 5, 5, 5\}$, as it must be since every integer factor set is also a real factor set. Second, the repeating number, $1000^{1/7}$, in the winning real factor set is very close to e.

The New Third Problem Guided by our approach to the old third problem, we reconsider the fact that the winning real factor set for k consists of n copies of $k^{1/n}$, and we ask what the situation would be if we relaxed the condition that n be a counting number. (In the definition below, y assumes the role of n.)

DEFINITION 3′. Given a positive real number k. A *pseudo factor set* of k is an ordered pair of positive real numbers (x, y) such that $x^y = k$. A pseudo factor set (x, y) of k is a *winning pseudo factor set* of k if $xy \leq uv$ for any other pseudo factor set (u, v) of k.

Example. $(1000^{1/7}, 7)$ is a pseudo factor set of 1000, but it is not a winning pseudo factor set because the pseudo factor set $(e, \ln 1000)$ has a smaller associated "sum," 18.77723 (approximately).

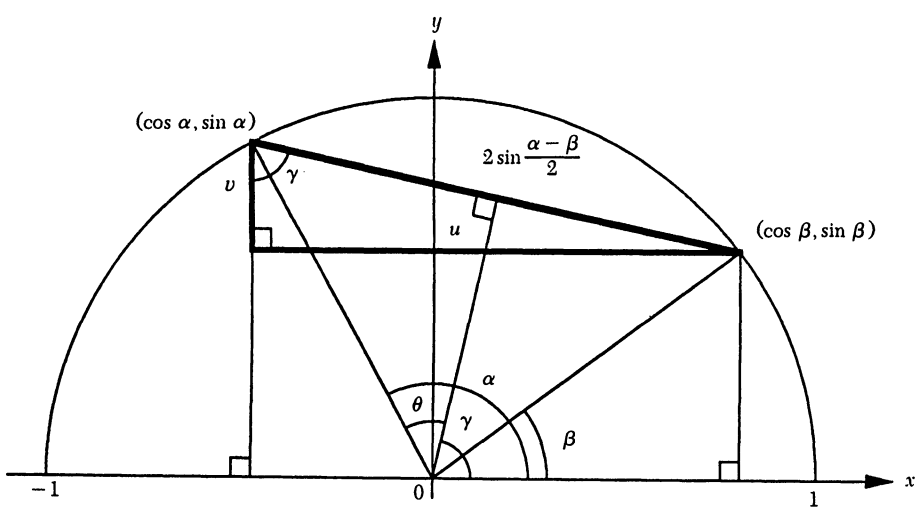
Problem 3′. Given a positive real number k, find a winning pseudo factor set of k.

THEOREM 3′. *Every positive real number k has a unique winning pseudo factor set, namely $(e, \ln k)$. The associated minimal sum is $e \ln k$.*

Proof. The task is to minimize the function $f(x, y) = xy$ subject to the constraints, $x^y = k$, $x > 0$, $y > 0$. Solving the constraint equation $x^y = k$ for y yields $y = \ln k / \ln x$. Substituting that value for y into the formula for f yields a function of one variable, $g(x) = x \ln k / \ln x$. The function g is minimized where its (easily calculated) derivative is zero, namely at $x = e$. Thus f is minimized when $x = e$ and $y = \ln k$.

Conclusion In the first half of this paper we began with a problem intended for children, generalized it twice, and solved all three problems. In the second half we formulated three dual problems and, paralleling our strategies from the first half, solved the three new problems. For the first problems in the two families, the "integer-integer" problems, the results were a bit disappointing: (new) Theorem 1′ bore little resemblance to (old) Theorem 1. For the second problems, the "real-integer" problems, (new) Theorem 2′ paralleled (old) Theorem 2 closely. Some of the details of proof, however, were different and more difficult. For the third problems, the "real-real" problems, (new) Theorem 3′ and (old) Theorem 3 turned out to be almost identical, and even their proofs were "dual."

Proof Without Words: The Difference-Product Identities



$$\theta = \frac{\alpha - \beta}{2} \qquad \gamma = \frac{\alpha + \beta}{2}$$

$$\sin \alpha - \sin \beta = \upsilon = 2 \sin \frac{\alpha - \beta}{2} \cos \frac{\alpha + \beta}{2}$$

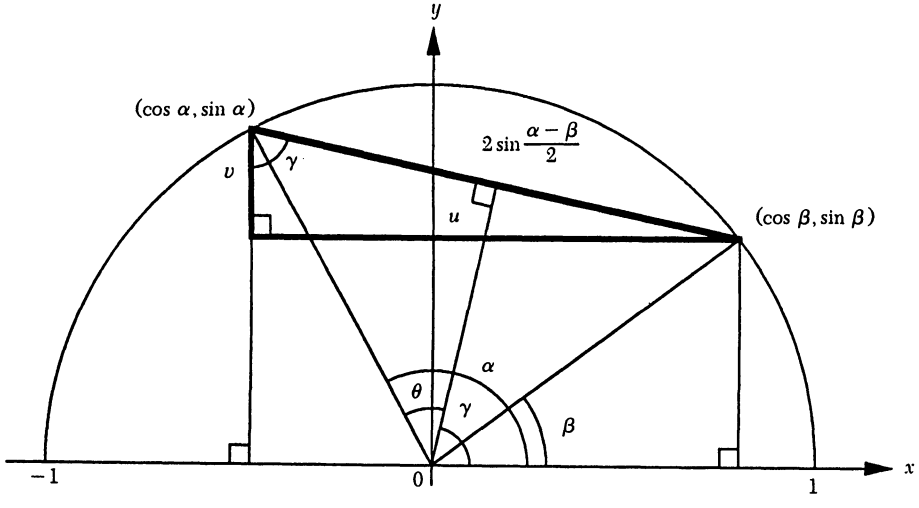
$$\cos \beta - \cos \alpha = u = 2 \sin \frac{\alpha - \beta}{2} \sin \frac{\alpha + \beta}{2}$$

—Sidney H. Kung
Jacksonville University
Jacksonville, FL 32211

Proof. The task is to minimize the function $f(x, y) = xy$ subject to the constraints, $x^y = k$, $x > 0$, $y > 0$. Solving the constraint equation $x^y = k$ for y yields $y = \ln k / \ln x$. Substituting that value for y into the formula for f yields a function of one variable, $g(x) = x \ln k / \ln x$. The function g is minimized where its (easily calculated) derivative is zero, namely at $x = e$. Thus f is minimized when $x = e$ and $y = \ln k$.

Conclusion In the first half of this paper we began with a problem intended for children, generalized it twice, and solved all three problems. In the second half we formulated three dual problems and, paralleling our strategies from the first half, solved the three new problems. For the first problems in the two families, the "integer-integer" problems, the results were a bit disappointing: (new) Theorem 1′ bore little resemblance to (old) Theorem 1. For the second problems, the "real-integer" problems, (new) Theorem 2′ paralleled (old) Theorem 2 closely. Some of the details of proof, however, were different and more difficult. For the third problems, the "real-real" problems, (new) Theorem 3′ and (old) Theorem 3 turned out to be almost identical, and even their proofs were "dual."

Proof Without Words: The Difference-Product Identities



$$\theta = \frac{\alpha - \beta}{2} \qquad \gamma = \frac{\alpha + \beta}{2}$$

$$\sin \alpha - \sin \beta = v = 2 \sin \frac{\alpha - \beta}{2} \cos \frac{\alpha + \beta}{2}$$

$$\cos \beta - \cos \alpha = u = 2 \sin \frac{\alpha - \beta}{2} \sin \frac{\alpha + \beta}{2}$$

—Sidney H. Kung
Jacksonville University
Jacksonville, FL 32211

A Markov Chain Analysis of the Game of Jai Alai

PHILIP J. BYRNE
College of St. Benedict
St. Joseph, MN 56374

ROBERT HESSE
University of Minnesota
Minneapolis, MN 55455

Introduction Jai Alai, a game resembling racquetball, evolved in Spain during the seventeenth century. In twentieth-century America, Connecticut, Florida, and Rhode Island operate Jai Alai frontons, where fans can watch the action and bet on the outcomes of games. While most fans find the play itself exciting, the real mathematical interest is the manner in which a winner is determined.

Before play begins, the eight players (or two-player teams) are placed in a queue with assigned post positions 1 to 8. A game consists of a sequence of short matches between two players; the first match pits player 1 against player 2. The winner of a match faces the next player in the queue, while the loser of a match returns to the back of the queue. The first seven matches are worth one point each; succeeding matches are worth two points. The winner is the first player to reach or exceed seven points.

Experienced Jai Alai bettors realize, intuitively, that players in low-numbered post positions have an advantage over players near the back of the queue. Informal analysis supports this point: for example, player 1 could win the game by winning the first seven matches, for one point each. Even if he loses an early match, player 1 will likely have a second opportunity to play. Player 6, on the other hand, could win by surviving his first five matches (the last three are worth two points each). But if player 6 loses any of these matches, another player may well reach seven points before player 6 returns to the front of the queue.

We will show how a Jai Alai game can be modelled as a Markov chain, and thus show how each player's winning depends on his post position. We will assume for convenience that all eight players have equal skill, but other assumptions about relative skills can be readily incorporated into the Markov chain analysis. The same approach can also be applied to other kinds of bets, such as trifectas and quinielas.

The model To model Jai Alai using Markov chains, we must first define an appropriate notion of a state. To describe the game at any time requires two data for each player: (1) his current position in the queue; and (2) his current score. Thus we assign to the ith position in the queue an ordered pair (a_i, b_i), with a_i the number of the player in position i, and b_i this player's current score. In particular, a_1 and a_2 are the players who will meet in the next match. We now define a state to be an ordered set $\{(a_1, b_1), (a_2, b_2), (a_3, b_3), (a_4, b_4), (a_5, b_5), (a_6, b_6), (a_7, b_7), (a_8, b_8)\}$. The initial state, for instance, is always $\{(1,0), (2,0), (3,0), (4,0), (5,0), (6,0), (7,0), (8,0)\}$. The second state must then be either $\{(1,1), (3,0), (4,0), (5,0), (6,0), (7,0), (8,0), (2,0)\}$ or $\{(2,1), (3,0), (4,0), (5,0), (6,0), (7,0), (8,0), (1,0)\}$.

We now assume that the outcomes of different matches are probabilistically independent of one another, and that the probability of player m winning a match against player n remains constant throughout the game. Then the probability of the game moving from one state to another depends only on the states, not on the previous history of the game. With these assumptions, a Jai Alai game becomes a Markov chain.

We'll need some basic notation and terminology and a fundamental result. First we order the states in some convenient manner, and assign them labels $1, 2, 3, \ldots$. If the game is in state i after r matches, the transition probability of moving to state j on the next match is denoted by p_{ij}. For most states in our Jai Alai game there are only two other states to which they can move. For every state i of this type, $p_{ij} = 0$ for all but two j's, so the ith row of the transition matrix $P = (p_{ij})$ has only two nonzero entries. The only other possible states are those in which some player has amassed seven or more points, winning the game. For all such states i we will take $p_{ii} = 1$ and $p_{ij} = 0$ if $i \neq j$. These states are called absorbing.

For any Markov chain with a finite number of states, we can label the absorbing states with integers $1, 2, \ldots, s$, and the nonabsorbing with integers $s+1$, $s+2, \ldots, s+t$. Then the transition matrix P has the form

$$P = \begin{pmatrix} I_s & 0 \\ R & Q \end{pmatrix}$$

where I_s is the $s \times s$ identity matrix, 0 is an $s \times t$ matrix of zeros, and R and Q are $t \times s$ and $t \times t$ matrices, respectively. In particular, R gives transition probabilities from nonabsorbing to absorbing states, and Q gives transition probabilities from nonabsorbing to nonabsorbing states. It's a general fact that the (i, j)-entry of the $t \times s$ matrix $(I_t - Q)^{-1}R$ gives the probability that the Markov chain ends up in absorbing state j given that the initial state was $s + i$ (see the Appendix for the sketch of a proof). For our Jai Alai application, the initial state is always that in which the queue has the players in numerical order, and each player has zero points.

A three-player example To illustrate these ideas, consider first a simplified Jai Alai game, with only three players and two points needed for a win. The first two matches are worth one point each; a third match, if needed, is worth two points. By analogy with the eight-player game, a state is an ordered set of three ordered pairs $\{(a_1, b_1), (a_2, b_2), (a_3, b_3)\}$, where a_i and b_i denote the player number and current score for the player in the ith position in the queue. There are, in all, eleven possible states, which we label as follows:

Label	State
7	{(1, 0), (2, 0), (3, 0)}
8	{(1, 1), (3, 0), (2, 0)}
1	{(1, 2), (2, 0), (3, 0)}
9	{(3, 1), (2, 0), (1, 1)}
2	{(3, 3), (1, 1), (2, 0)}
3	{(2, 2), (1, 1), (3, 1)}
10	{(2, 1), (3, 0), (1, 0)}
4	{(2, 2), (1, 0), (3, 0)}
11	{(3, 1), (1, 0), (2, 1)}
5	{(3, 3), (2, 1), (1, 0)}
6	{(1, 2), (2, 1), (3, 1)}

If the three players have equal ability, then the associated transition matrix for this three-player game is

state	state 1	2	3	4	5	6	7	8	9	10	11
1	1	0	0	0	0	0	0	0	0	0	0
2	0	1	0	0	0	0	0	0	0	0	0
3	0	0	1	0	0	0	0	0	0	0	0
4	0	0	0	1	0	0	0	0	0	0	0
5	0	0	0	0	1	0	0	0	0	0	0
6	0	0	0	0	0	1	0	0	0	0	0
7	0	0	0	0	0	0	0	.5	0	.5	0
8	.5	0	0	0	0	0	0	0	.5	0	0
9	0	.5	.5	0	0	0	0	0	0	0	0
10	0	0	0	.5	0	0	0	0	0	0	.5
11	0	0	0	0	.5	.5	0	0	0	0	0

From this we can compute directly the 5×6 matrix $(I_5 - Q)^{-1}R$. Its first row, (.25, .125, .125, .25, .125, .125), displays the probabilities of the Jai Alai game ending in the absorbing states 1, 2, 3, 4, 5, and 6, respectively, given the initial state 7. Since player 1 is the winner in absorbing states 1 and 6, player 2 in states 3 and 4, and player 3 in states 2 and 5, their respective probabilities of winning are .375, .375, and .25.

The eight-player game Analyzing the eight-player game is similar, but the transition matrix P is much larger. To find the number of states, we wrote two computer programs to count all the vertices in an appropriate tree diagram. The root vertex of this tree corresponds to the initial state of the game; two branches connect the root with two vertices representing the two states that can occur next. Since each match in a Jai Alai game has two possible outcomes, most vertices have two branches emanating from them. Any vertex corresponding to a winning state has no outgoing branches. Our first program counted 844,767 vertices in this tree diagram. In the eight-player game (unlike the three-player game) some states can be reached by more than one sequence of match outcomes. Our second program eliminated these duplications. We found, in the end, a total of 134,215 distinct states in the eight-player game.

Since we wish to compute $(I_t - Q)^{-1}R$, the size of the transition matrix P might seem to create computational problems. However, two useful observations come to our rescue. First, the matrix P is sparse: no row contains more than two nonzero entries. Sparse matrices often admit special, efficient algorithms for such operations as multiplication and inversion (see, e.g., [1]). Second, since the Jai Alai game has only one possible initial state, which we number $s + 1$, we need only compute the first row of $(I_t - Q)^{-1}R$ to determine the players' winning probabilities.

This can be done by finding the first row of $(I_t - Q)^{-1}$, which we denote by $(x_1, x_2, \ldots, x_t)$, and then multiplying by the matrix R. We can find $(x_1, x_2, \ldots, x_t)$ by comparing the first rows on both sides of the equation $(I_t - Q)^{-1}(I_t - Q) = I_t$, which gives us $(x_1, x_2, \ldots, x_t)(I_t - Q) = (1, 0, 0, \ldots, 0)$. Taking transposes yields the linear system $Ax = b$, where $A = (I_t - Q)^t$, $x = (x_1, x_2, \ldots, x_t)^t$, and $b = (1, 0, 0, \ldots, 0)^t$. The nonabsorbing states can be labelled in such a way that $I_t - Q$ is nearly upper triangular (in the three player game, $I_5 - Q$ *was* upper triangular). In this case, A is nearly lower triangular. Thus, after relatively few row operations, back-substitution can be performed, starting with x_1, to successively find values for $x_1, x_2, x_3, \ldots, x_t$.

The winning probabilities given below, which assume the players to be of equal ability, agree with those found by Moser [3] in her computer search through all possible games.

Player	Probability of Winning
1	.1631
2	.1631
3	.1386
4	.1240
5	.1020
6	.1026
7	.0888
8	.1177

The table has several interesting features. First, since players 1 and 2 begin the game at the front of the queue and play each other in the first match, symmetry of their situations naturally results in equal probabilities of winning. Note also that player 8 has a higher winning probability than players 5, 6, and 7. This reflects the fact that only player 8 can win the game by winning as few as four matches on his first turn to play. This more than compensates for player 8's smaller probability of getting a second chance to play after a loss. The table also supports the general intuition of Jai Alai bettors that players in low-numbered post positions have an advantage. Note, however, the small advantage of player 6 over player 5. A possible explanation is that player 5 needs a string of six wins, while player 6 would need to win only five matches.

Many related problems could be studied with the approach presented here. For example, Moser [3] used her computer search through the Jai Alai game tree to determine the probabilities for place, show, and exacta bets when all players have equal abilities, and also the probabilities of each player winning under certain combinations of unequally-skilled players. All of these situations could be handled using Markov chains. For place, show, and exacta bets, we would need to expand the number of possible states to account for the way ties are broken for place and show in Jai Alai. If the players have unequal skill, then for each ordered pair of players, (m, n), we would assign a probability, θ_{mn}, of player m winning a match against player n. The implication for the transition matrix P is straightforward. If the transition from state i to state j involves player m winning a match against player n, then $p_{ij} = \theta_{mn}$. This change from the earlier case affects only the nonzero entries of P, so P will again be a sparse matrix and the Markov chain analysis will remain computationally feasible.

Appendix To find the probability that a Markov chain ends up in a certain absorbing state given that it started in a particular nonabsorbing state, we note first that the (i, j) entry of the matrix R gives the probability of moving from nonabsorbing state $s + i$ to absorbing state j in one step. The (i, j) entry of the matrix QR gives the probability of moving from nonabsorbing state $s + i$ to some other nonabsorbing state in one step, and then to absorbing state j on the next step. That is, the entries of QR are the probabilities of moving from nonabsorbing states to absorbing states in two steps. Similarly, the entries of Q^2R give the probabilities of moving from nonabsorbing states to absorbing states in three steps, and so on. Therefore, the probability that the Markov chain eventually ends up in absorbing state j given that the initial state was $s + i$ is determined by the (i, j) entry of the matrix

$$R + QR + Q^2R + Q^3R + \ldots = (I_t + Q + Q^2 + Q^3 + \ldots)R.$$

It can be shown that all the entries of Q^n approach zero as n tends to infinity (see [**2**, pp. 43–45]). This condition yields the following matrix generalization of the familiar formula for the sum of a geometric series:

$$(I_t - Q)^{-1} = I + Q + Q^2 + Q^3 + \ldots$$

(see [**2**, p. 22]). Thus we see that the (i, j)-entry of the $t \times s$ matrix $(I_t - Q)^{-1}R$ gives the probability that the Markov chain ends up in absorbing state j given that the initial state was $s + i$.

REFERENCES

1. G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd edition, Johns Hopkins University Press, Baltimore, 1989.
2. J. G. Kemeny and J. L. Snell, *Finite Markov Chains*, Springer-Verlag, New York, 1976.
3. L. E. Moser, A mathematical analysis of the game of Jai Alai, *Amer. Math. Monthly* 89 (1982), pp. 292–300.

Poker With Wild Cards—A Paradox?

STEVE GADBOIS
Rhodes College
Memphis, TN 38112

I participate in a sporadic poker game whose organizer detests any use of wild cards. (A *wild card* can be called anything its holder wishes.) I'd always attributed this aversion to some personality quirk. Then I discovered a reason to share his concern.

After a recent class in which I tossed out an unsubstantiated claim about wild cards sometimes altering the accepted hierarchy of poker hands, I decided I'd better actually do the calculations before my students did. I wasn't surprised to substantiate my claim, but I was surprised to discover that unresolvable inconsistencies can arise when wild cards are used. This note shows how, in one common situation, *no matter what hierarchy is established, the resulting probabilities are incompatible with it.* So perhaps my friend (who happens to be a political scientist, as well as the frequent victor in our always-friendly games) has more innate mathematical talent than either of us realized.

The usual hierarchy of poker hands (when played without wild cards) is, from best to worst, royal flush, straight flush, four-of-a-kind, full house, flush, straight, three-of-a-kind, two pair, one pair, and junk.[1] Without wild cards, this hierarchy is consistent

[1]Some of these terms may not be self-explanatory. A *royal flush* consists of an ace, king, queen, jack, and ten, all in one suit. A *straight flush* comprises five in a row, all in one suit (but not a royal flush). A *full house* includes three-of-a-kind and one pair. A *flush* consists of five cards in one suit (but not a royal flush or a straight flush). A *straight* has five in a row (but not a royal flush or a straight flush). Any other hand is *junk*.

$$R + QR + Q^2R + Q^3R + \ldots = (I_t + Q + Q^2 + Q^3 + \ldots)R.$$

It can be shown that all the entries of Q^n approach zero as n tends to infinity (see [**2**, pp. 43–45]). This condition yields the following matrix generalization of the familiar formula for the sum of a geometric series:

$$(I_t - Q)^{-1} = I + Q + Q^2 + Q^3 + \ldots$$

(see [**2**, p. 22]). Thus we see that the (i, j)-entry of the $t \times s$ matrix $(I_t - Q)^{-1}R$ gives the probability that the Markov chain ends up in absorbing state j given that the initial state was $s + i$.

REFERENCES

1. G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd edition, Johns Hopkins University Press, Baltimore, 1989.
2. J. G. Kemeny and J. L. Snell, *Finite Markov Chains*, Springer-Verlag, New York, 1976.
3. L. E. Moser, A mathematical analysis of the game of Jai Alai, *Amer. Math. Monthly* 89 (1982), pp. 292–300.

Poker With Wild Cards—A Paradox?

STEVE GADBOIS
Rhodes College
Memphis, TN 38112

I participate in a sporadic poker game whose organizer detests any use of wild cards. (A *wild card* can be called anything its holder wishes.) I'd always attributed this aversion to some personality quirk. Then I discovered a reason to share his concern.

After a recent class in which I tossed out an unsubstantiated claim about wild cards sometimes altering the accepted hierarchy of poker hands, I decided I'd better actually do the calculations before my students did. I wasn't surprised to substantiate my claim, but I was surprised to discover that unresolvable inconsistencies can arise when wild cards are used. This note shows how, in one common situation, *no matter what hierarchy is established, the resulting probabilities are incompatible with it.* So perhaps my friend (who happens to be a political scientist, as well as the frequent victor in our always-friendly games) has more innate mathematical talent than either of us realized.

The usual hierarchy of poker hands (when played without wild cards) is, from best to worst, royal flush, straight flush, four-of-a-kind, full house, flush, straight, three-of-a-kind, two pair, one pair, and junk.[1] Without wild cards, this hierarchy is consistent

[1]Some of these terms may not be self-explanatory. A *royal flush* consists of an ace, king, queen, jack, and ten, all in one suit. A *straight flush* comprises five in a row, all in one suit (but not a royal flush). A *full house* includes three-of-a-kind and one pair. A *flush* consists of five cards in one suit (but not a royal flush or a straight flush). A *straight* has five in a row (but not a royal flush or a straight flush). Any other hand is *junk*.

with the relative frequency of the hands. (For the calculations for 5 card poker without wild cards, see [2]. Packel allows an ace to be either high or low in a straight (instead of just high). This does not affect the hierarchy itself.)

If the two jokers are added as wild cards, one more hand is possible: five-of-a-kind. The first table gives frequencies and probabilities for all possible hands. The verifications of these frequencies are nice exercises in combinatorics. Here's one sample of the systematic (if somewhat pedantic) reasoning involved. For four-of-a-kind, there are three distinct ways to specify the hand without redundancy.

(a) Select no joker of the two; select one denomination from the thirteen; select all four of that denomination; select one denomination from the remaining twelve; and select one of the four of that denomination.
(b) Select one joker of the two; select one denomination from the thirteen; select three of the four of that denomination; select one denomination from the remaining twelve; and select one of the four of that denomination.
(c) Select two jokers of the two; select one denomination from the thirteen; select two of the four of that denomination; select one denomination from the remaining twelve; and select one of the four of that denomination.

Thus the number of ways to get four-of-a-kind is

$$\binom{2}{0}\binom{13}{1}\binom{4}{4}\binom{12}{1}\binom{4}{1} + \binom{2}{1}\binom{13}{1}\binom{4}{3}\binom{12}{1}\binom{4}{1} + \binom{2}{2}\binom{13}{1}\binom{4}{2}\binom{12}{1}\binom{4}{1}$$
$$= 624 + 4992 + 3744 = 9360.$$

TABLE 1 Wild card poker frequencies and probabilities, based on the usual hierarchy

Rank	Type	Frequency	Probability
1	FIVE-OF-A-KIND	78	0.000025
2	ROYAL FLUSH	84	0.000027
3	STRAIGHT FLUSH	480	0.000152
4	FOUR-OF-A-KIND	9360	0.002960
5	FULL HOUSE	9360	0.002960
6	FLUSH	11448	0.003620
7	STRAIGHT	30540	0.009657
8	THREE-OF-A-KIND	233584	0.073860
9	TWO PAIR	123552	0.039068
10	ONE PAIR	1440464	0.455481
11	JUNK	1303560	0.412192
	(TOTAL)	$\binom{54}{5} = 3162510$	1

Observe one anomaly in the first table: Three-of-a-kind and two pair are in the wrong order. But if their positions are reversed, many hands that would have been three-of-a-kind are now best called two pair. For example, {A♣, 8♥, 4♠, JOKER, JOKER} can be called "three aces" (if three-of-a-kind beats two pair) or two aces and two eights (if two pair beats three-of-a-kind). So the numbers of these two types of hands change, as shown in the second table: Two pair and three-of-a-kind are in the wrong order again! (In fact, the situation is relatively worse than before the reversal.)

A look at the first table reveals that the same phenomenon occurs with one pair and junk. For example, {A♣, 8♥, 4♠, 2♦, JOKER} can be called "two aces" (if one pair beats junk) or "junk" (if junk beats one pair, calling the JOKER a king, say). (For other situations in which junk beats one pair or even two pair, see [1].)

TABLE 2 Wild card poker frequencies and probabilities, based on a revised hierarchy

Rank	Type	Frequency	Probability
1	FIVE-OF-A-KIND	78	0.000025
2	ROYAL FLUSH	84	0.000027
3	STRAIGHT FLUSH	480	0.000152
4	FOUR-OF-A-KIND	9360	0.002960
5	FULL HOUSE	9360	0.002960
6	FLUSH	11448	0.003620
7	STRAIGHT	30540	0.009657
8	TWO PAIR	302224	0.095565
9	THREE-OF-A-KIND	54912	0.017363
10	JUNK	1645784	0.520404
11	ONE PAIR	1098240	0.347268

The more one looks, the worse it gets. In the original hierarchy, there were 9360 four-of-a-kind hands and 9360 full house hands. So one could arbitrarily decide to rank a full house above four-of-a-kind. But this would really be disastrous, for then there would turn out to be 18096 full houses and 624 four-of-a-kind! With two added jokers as wild cards, there is *no* hierarchy of hands that is consistent with the frequency of the hands.

REFERENCES

1. Y. L. Cheung, Why poker is played with five cards, *The Mathematical Gazette* 73 (1989), 313–315.
2. Edward W. Packel, *The Mathematics of Games and Gambling*, Mathematical Association of America, 1981.

Counting Squares in $\mathbb{Z}_n$

WALTER D. STANGL
Biola University
LaMirada, CA 90639

An elementary number theory problem is to determine the possible forms of squares among the positive integers. For instance, it is easy to see that any square must be of the form $3k$ or $3k + 1$. (Since every positive integer can be written as either $3q$, $3q + 1$, or $3q + 2$, simply square these numbers and simplify.) Restated, this assertion is that 0 and 1 are the squares in $\mathbb{Z}_3$, the ring of equivalence classes of integers modulo 3. In general, a square has the form $nk + r$ if, and only if, r is a square in the ring $\mathbb{Z}_n$. How many squares are there in $\mathbb{Z}_n$?

Fundamental notions An element a in $\mathbb{Z}_n$ is a *square* in $\mathbb{Z}_n$ if and only if $x^2 = a$ has a solution in $\mathbb{Z}_n$. The *units* of $\mathbb{Z}_n$ are the elements that are relatively prime to n. The units that are squares are commonly called *quadratic residues* (or, more precisely, the quadratic residues mod n in a reduced residue system) [**1**, p. 84]. The quadratic residues have been completely characterized [**2**, p. 201], and the standard results will be utilized in what follows.

TABLE 2 Wild card poker frequencies and probabilities, based on a revised hierarchy

Rank	Type	Frequency	Probability
1	FIVE-OF-A-KIND	78	0.000025
2	ROYAL FLUSH	84	0.000027
3	STRAIGHT FLUSH	480	0.000152
4	FOUR-OF-A-KIND	9360	0.002960
5	FULL HOUSE	9360	0.002960
6	FLUSH	11448	0.003620
7	STRAIGHT	30540	0.009657
8	TWO PAIR	302224	0.095565
9	THREE-OF-A-KIND	54912	0.017363
10	JUNK	1645784	0.520404
11	ONE PAIR	1098240	0.347268

The more one looks, the worse it gets. In the original hierarchy, there were 9360 four-of-a-kind hands and 9360 full house hands. So one could arbitrarily decide to rank a full house above four-of-a-kind. But this would really be disastrous, for then there would turn out to be 18096 full houses and 624 four-of-a-kind! With two added jokers as wild cards, there is *no* hierarchy of hands that is consistent with the frequency of the hands.

REFERENCES

1. Y. L. Cheung, Why poker is played with five cards, *The Mathematical Gazette* 73 (1989), 313–315.
2. Edward W. Packel, *The Mathematics of Games and Gambling*, Mathematical Association of America, 1981.

Counting Squares in $\mathbb{Z}_n$

WALTER D. STANGL
Biola University
LaMirada, CA 90639

An elementary number theory problem is to determine the possible forms of squares among the positive integers. For instance, it is easy to see that any square must be of the form $3k$ or $3k + 1$. (Since every positive integer can be written as either $3q$, $3q + 1$, or $3q + 2$, simply square these numbers and simplify.) Restated, this assertion is that 0 and 1 are the squares in $\mathbb{Z}_3$, the ring of equivalence classes of integers modulo 3. In general, a square has the form $nk + r$ if, and only if, r is a square in the ring $\mathbb{Z}_n$. How many squares are there in $\mathbb{Z}_n$?

Fundamental notions An element a in $\mathbb{Z}_n$ is a *square* in $\mathbb{Z}_n$ if and only if $x^2 = a$ has a solution in $\mathbb{Z}_n$. The *units* of $\mathbb{Z}_n$ are the elements that are relatively prime to n. The units that are squares are commonly called *quadratic residues* (or, more precisely, the quadratic residues mod n in a reduced residue system) [**1**, p. 84]. The quadratic residues have been completely characterized [**2**, p. 201], and the standard results will be utilized in what follows.

We will adopt the following notation: $q(n)$ = the number of quadratic residues in $\mathbb{Z}_n$, and $s(n)$ = the number of squares in $\mathbb{Z}_n$. For example, $q(8) = 1$ since $x^2 = 1$ has a solution in $\mathbb{Z}_8$ (as a matter of fact, all four units, namely 1, 3, 5, and 7, are solutions), and $x^2 = 3$, $x^2 = 5$, and $x^2 = 7$ do not have any solutions in $\mathbb{Z}_8$. Also, $s(8) = 3$ since $x^2 = 0$ and $x^2 = 4$ also have solutions in $\mathbb{Z}_8$, but $x^2 = 2$ and $x^2 = 6$ do not.

A number-theoretic function $f(n)$ is *multiplicative* if $\gcd(m, n) = 1$ implies $f(mn) = f(m) \cdot f(n)$. Typical number-theoretic functions that are multiplicative include the number of positive divisors of n and the sum of these divisors [1, p. 109]. A number-theoretic function that is multiplicative is completely characterized by its values on powers of primes. Both $q(n)$ and $s(n)$ are multiplicative; we derive both recursive and closed-form formulas for these functions on the powers of primes. This will allow us to compute $s(n)$ and $q(n)$ for any n, based on the prime factorization of n.

Suppose $\gcd(m, n) = 1$. Then $\mathbb{Z}_{mn}$ is isomorphic to $\mathbb{Z}_m \times \mathbb{Z}_n$ under the ring isomorphism h: $\mathbb{Z}_{mn} \to \mathbb{Z}_m \times \mathbb{Z}_n$ defined by $h(z) = (z \bmod m, z \bmod n)$ [3, p. 80]. Suppose a is a square in $\mathbb{Z}_{mn}$. Then there is a b in $\mathbb{Z}_{mn}$ such that $b^2 = a$. Since h is a function from $\mathbb{Z}_{mn}$ onto $\mathbb{Z}_m \times \mathbb{Z}_n$, there exists $(x, y) \in \mathbb{Z}_m \times \mathbb{Z}_n$ such that $h(b) = (x, y)$. Then $h(a) = h(b^2) = [h(b)]^2 = (x, y)^2 = (x^2, y^2)$, so $h(a)$ is a square in $\mathbb{Z}_m \times \mathbb{Z}_n$. Hence $s(mn) \leq s(m) \cdot s(n)$.

On the other hand, if u in $\mathbb{Z}_m$ and v in $\mathbb{Z}_n$ are squares, then there exist x in $\mathbb{Z}_m$ and y in $\mathbb{Z}_n$ such that $(x^2, y^2) = (u, v)$ in $\mathbb{Z}_m \times \mathbb{Z}_n$. Thus $h^{-1}(u, v) = h^{-1}[(x, y)^2] = [h^{-1}(x, y)]^2$, so $h^{-1}(u, v)$ is a square in $\mathbb{Z}_{mn}$. Thus $s(mn) \geq s(m) \cdot s(n)$.

Combining these results yields the desired equality, showing that $s(n)$ is a multiplicative function. To extend the proof to $q(n)$ requires merely the observation that for any integer b, $\gcd(b, mn) = 1$ if, and only if, $\gcd(b, m) = 1$ and $\gcd(b, n) = 1$.

Recursion formula Our next goal is to prove a general recursion formula for the number of squares in $\mathbb{Z}_{p^n}$, where p is a prime greater than 2. Once this is achieved, formulas in closed form for the various components will complete our counting procedure. We begin with the observation that the squares in $\mathbb{Z}_{p^n}$ that are not quadratic residues are generated by the squares in $\mathbb{Z}_{p^{n-2}}$, i.e., b is a square in $\mathbb{Z}_{p^{n-2}}$ if and only if bp^2 is a square in $\mathbb{Z}_{p^n}$.

First, suppose there is c in $\mathbb{Z}_{p^{n-2}}$ such that $c^2 = kp^{n-2} + b$ in $\mathbb{Z}$. Then $c^2p^2 = kp^n + bp^2$. Now $cp < p^n$, so $(cp)^2 = bp^2$ is a square in $\mathbb{Z}_{p^n}$. Conversely, suppose there is y in $\mathbb{Z}_{p^n}$ such that $y^2 = mp^n + sp^2$ in $\mathbb{Z}$. Then p^2 divides y^2, so p divides y. Thus there is c such that $y = cp$. Then $c^2 = mp^{n-2} + s$ and s is a square in $\mathbb{Z}_{p^{n-2}}$.

Now we wish to count all the squares in $\mathbb{Z}_{p^n}$. We begin by observing that the squares are of two types. Since $q(p^n)$ counts the squares in $\mathbb{Z}_{p^n}$ that are units, we must merely count the squares that are non-units, i.e., multiples of p. Suppose kp is a square in $\mathbb{Z}_{p^n}$. Then there is a b such that $b^2 = cp^n + kp$. Then p divides b^2, and hence b. Thus p^2 divides b^2, and hence kp, so p divides k. Hence the multiples of p that are squares are multiples of p^2. But by the preceding result, the number of these will be given by $s(p^{n-2})$.

Thus we have proven the following recursion formula.

THEOREM. *For $n \geq 3$, $s(p^n) = q(p^n) + s(p^{n-2})$.*

Powers of odd primes In order to obtain explicit formulas for the functions $q(p^n)$ and $s(p^n)$, it is useful to deal with the case $p = 2$ separately. The argument for powers of an odd prime p depends on the existence of a primitive root for p^n for each n. In algebraic language, this says that the units of $\mathbb{Z}_{p^n}$ form a cyclic group with

respect to multiplication and hence have a generator [**1**, p. 62]. Since this is not true for powers of 2 greater than or equal to 3, our approach and results will need to be altered for that situation.

If p is an odd prime, the Euler phi-function yields the numbers of units of $\mathbb{Z}_{p^n}$, namely $p^n - p^{n-1}$. There is a primitive root of p^n. The even powers of this primitive root are clearly distinct quadratic residues, and the following formula is proven.

THEOREM. *If p is an odd prime, then $q(p^n) = (p^n - p^{n-1})/2$, for all $n \geq 1$.*

In order to count all of the squares in $\mathbb{Z}_{p^n}$, it is useful to look at the first two cases separately. Since 0 is the only non-unit in $\mathbb{Z}_p$, clearly $s(p) = q(p) + 1 = (p+1)/2$. In $\mathbb{Z}_{p^2}$, the non-units are multiples of p, and have squares equal to 0. So $s(p^2) = q(p^2) + 1 = (p^2 - p + 2)/2$.

Now suppose $n \geq 3$ and n is even. By repeated applications of the recursion formula, we obtain

$$\begin{aligned} s(p^n) &= \frac{p^n - p^{n-1}}{2} + \frac{p^{n-2} - p^{n-1}}{2} + \cdots \frac{p^4 - p^3}{2} + \frac{p^2 - p + 2}{2} \\ &= \frac{p^{n+1} - p^n + p^n - p^{n-1} + p^{n-1} - \cdots + p^3 - p^2 + 2p + p^2 - p + 2}{2(p+1)} \\ &= \frac{p^{n+1} + p + 2}{2(p+1)}. \end{aligned}$$

If n is odd, we obtain

$$\begin{aligned} s(p^n) &= \frac{p^n - p^{n-1}}{2} + \frac{p^{n-2} - p^{n-1}}{2} + \cdots \frac{p^3 - p^2}{2} + \frac{p+1}{2} \\ &= \frac{p^{n+1} - p^n + p^n - p^{n-1} - \cdots + p^2 + 2p + 1}{2(p+1)} \\ &= \frac{p^{n+1} + 2p + 1}{2(p+1)}. \end{aligned}$$

Our results are summarized in the following theorem.

THEOREM. *Suppose p is an odd prime. Then*

$$s(p) = \frac{p+1}{2} \quad \text{and} \quad s(p^2) = \frac{p^2 - p + 2}{2}.$$

If $n \geq 3$, then

$$s(p^n) = \begin{cases} \dfrac{p^{n+1} + p + 2}{2(p+1)} & n \text{ even} \\ \dfrac{p^{n+1} + 2p + 1}{2(p+1)} & n \text{ odd}. \end{cases}$$

Powers of two Now we proceed to the remaining case: powers of 2. We need a preliminary result before moving to our main goal.

Suppose $n \geq 3$, and $\gcd(a, 2^n) = 1$. Consider the equation $x^2 = a$ in $\mathbb{Z}_{2^n}$. Suppose b is a solution. Then, clearly, $-b$ is also a solution. Also $b \neq -b$, since otherwise $2b = 0$ which implies $\gcd(b, 2^n) \neq 1$ while we know $\gcd(b^2, 2^n) = 1$. Another pair of solutions is easily verified to be given by $2^{n-1} \pm b$. These values are also seen to be distinct by the above argument.

To show these four solutions are the only solutions, suppose $\gcd(c, 2^n) = 1$ and c is a solution in addition to b. Then $b^2 = a = c^2$ in $\mathbb{Z}_{2^n}$ implies $b^2 - c^2 = 0$ or $(b-c)(b+c) = 0$ in $\mathbb{Z}_{2^n}$. Since b and c are both odd, either $(b-c)$ or $(b+c)$ must be of the form $4m + 2 = 2(2m+1)$. So the other factor is a multiple of 2^{n-1} or 0. Hence $c = 2^{n-1} \pm b$ or $c = \pm b$.

Thus we conclude that if $x^2 = a$ has a solution in $\mathbb{Z}_{2^n}$, then the equation has exactly 4 distinct solutions in $\mathbb{Z}_{2^n}$.

We observe that the only quadratic residue in either $\mathbb{Z}_2$ or $\mathbb{Z}_4$ is 1. It follows that $q(2) = q(4) = 1$.

For $n \geq 3$, there are 2^{n-1} units in $\mathbb{Z}_{2^n}$, namely the odd numbers. Consider two units equivalent if their squares are equal. Then the units can be divided into equivalence classes of 4 units each; hence there will be $2^{-2}2^{n-1} = 2^{n-3}$ quadratic residues in $\mathbb{Z}_{2^n}$. Thus for $n \geq 3$, $q(2^n) = 2^{n-3}$.

We are now ready to prove our final formulas. Here's the result.

THEOREM.

$$s(2^n) = \begin{cases} \dfrac{2^{n-1}+4}{3} & n \text{ even} \\ \dfrac{2^{n-1}+5}{3} & n \text{ odd } n \geq 3. \end{cases}$$

Proof. The argument is by induction. Starting with $n = 2$, it is clear that $s(2^2) = 2$. Now assume that the formula holds for $n \leq k$. There are two cases.

Case I. $k+1$ is even. Then

$$s(2^{k+1}) = q(2^{k+1}) + s(2^{k-1}) = 2^{(k+1)-3} + \frac{2^{(k-1)-1}+4}{3}$$
$$= 2^{k-2} + \frac{2^{k-2}+4}{3} = \frac{4\cdot 2^{k-2}+4}{3} = \frac{2^{(k+1)-1}+4}{3}.$$

Case II. $k+1$ is odd. Then

$$s(2^{k+1}) = q(2^{k+1}) + s(2^{k-1}) = 2^{(k+1)-3} + \frac{2^{(k-1)-1}+5}{3}$$
$$= 2^{k-2} + \frac{2^{k-2}+5}{3} = \frac{4\cdot 2^{k-2}+5}{3} = \frac{2^{(k+1)-1}+5}{3}.$$

The preceding formulas are derivable directly from the recursion formula. For instance, if n is odd, repeated applications yield

$$s(2^n) = q(2^n) + q(2^{n-2}) + \cdots + q(2^3) + s(2^1)$$
$$= 2^{n-3} + 2^{n-5} + \cdots + 1 + 2.$$

So we need a formula for the sum of the even powers of 2. Letting $x_n = 1 + 2^2 + \cdots + 2^{2n}$, we have

$$x_n = (2^2)^0 + (2^2)^1 + \cdots + (2^2)^n$$
$$= \frac{(2^2)^{n+1} - 1}{2^n - 1}.$$

So $x_n = \dfrac{2^{2n+2}-1}{3}$, and

$$s(2^n) = x_{(n-3)/2} + 2$$
$$= \frac{2^{n-1}-1}{3} + 2 = \frac{2^{n-1}+5}{3}.$$

A formula for the sum of the odd powers of 2 is obtained from x_n by factoring, and then $s(2^n)$ is easily computed.

REFERENCES

1. Ivan Niven and Herbert Zuckerman, *An Introduction to the Theory of Numbers*, 4th edition, John Wiley & Sons, Inc., New York, 1980.
2. David M. Burton, *Elementary Number Theory*, 3rd edition, Wm. C. Brown, Dubuque, IA, 1994.
3. John Fraleigh, *A First Course in Abstract Algebra*, 3rd edition, Addison-Wesley, Reading, MA, 1982.

Magic Squares of Squares

JOHN P. ROBERTSON
560 Bair Road
Berwyn, PA 19312

A problem in the second edition of Guy's *Unsolved Problems in Number Theory* [**1**] is to prove or disprove that a three-by-three magic square can be constructed from nine distinct integer squares (Problem D15). There are relationships between magic squares, arithmetic progressions, Pythagorean right triangles, congruent numbers, and elliptic curves. This note will follow this chain and show that the following three problems are equivalent to the original problem:

P1. Prove or disprove that there are three arithmetic progressions such that each has three terms, each has the same difference between terms as the other two, the terms are all perfect squares, and the middle terms of the three arithmetic progressions themselves form an arithmetic progression.

P2. Prove or disprove that there are three rational right triangles with the same area, such that the squares of the hypotenuses are in arithmetic progression.

P3. Prove or disprove that there is an elliptic curve, $y^2 = x^3 - n^2x$, where n is a congruent number, with three rational points on the curve, (x_1, y_1), (x_2, y_2), and (x_3, y_3), such that each point is "double" another rational point on the elliptic curve ("double" in the sense of the group structure for points on an elliptic curve), and x_1, x_2, and x_3 are in arithmetic progression.

The original problem is due to LaBar [**2**]. Guy [**1**] notes that the problem requires finding x, y, and z so that the nine quantities x^2, y^2, z^2, $y^2+z^2-x^2$, $z^2+x^2-y^2$, $x^2+y^2-z^2$, $2x^2-y^2$, $2x^2-z^2$, and $3x^2-y^2-z^2$, are distinct perfect squares.

Magic squares and arithmetic progressions For any three-by-three magic square made up of distinct positive integers, there are three positive integers a, u, and v,

So $x_n = \dfrac{2^{2n+2}-1}{3}$, and

$$s(2^n) = x_{(n-3)/2} + 2$$
$$= \frac{2^{n-1}-1}{3} + 2 = \frac{2^{n-1}+5}{3}.$$

A formula for the sum of the odd powers of 2 is obtained from x_n by factoring, and then $s(2^n)$ is easily computed.

REFERENCES

1. Ivan Niven and Herbert Zuckerman, *An Introduction to the Theory of Numbers*, 4th edition, John Wiley & Sons, Inc., New York, 1980.
2. David M. Burton, *Elementary Number Theory*, 3rd edition, Wm. C. Brown, Dubuque, IA, 1994.
3. John Fraleigh, *A First Course in Abstract Algebra*, 3rd edition, Addison-Wesley, Reading, MA, 1982.

Magic Squares of Squares

JOHN P. ROBERTSON
560 Bair Road
Berwyn, PA 19312

A problem in the second edition of Guy's *Unsolved Problems in Number Theory* [**1**] is to prove or disprove that a three-by-three magic square can be constructed from nine distinct integer squares (Problem D15). There are relationships between magic squares, arithmetic progressions, Pythagorean right triangles, congruent numbers, and elliptic curves. This note will follow this chain and show that the following three problems are equivalent to the original problem:

P1. Prove or disprove that there are three arithmetic progressions such that each has three terms, each has the same difference between terms as the other two, the terms are all perfect squares, and the middle terms of the three arithmetic progressions themselves form an arithmetic progression.

P2. Prove or disprove that there are three rational right triangles with the same area, such that the squares of the hypotenuses are in arithmetic progression.

P3. Prove or disprove that there is an elliptic curve, $y^2 = x^3 - n^2x$, where n is a congruent number, with three rational points on the curve, (x_1, y_1), (x_2, y_2), and (x_3, y_3), such that each point is "double" another rational point on the elliptic curve ("double" in the sense of the group structure for points on an elliptic curve), and x_1, x_2, and x_3 are in arithmetic progression.

The original problem is due to LaBar [**2**]. Guy [**1**] notes that the problem requires finding x, y, and z so that the nine quantities x^2, y^2, z^2, $y^2+z^2-x^2$, $z^2+x^2-y^2$, $x^2+y^2-z^2$, $2x^2-y^2$, $2x^2-z^2$, and $3x^2-y^2-z^2$, are distinct perfect squares.

Magic squares and arithmetic progressions For any three-by-three magic square made up of distinct positive integers, there are three positive integers a, u, and v,

such that the magic square can be expressed (possibly after rotation or reflection) as:

$$\begin{array}{ccc} a+u+2v & a & a+2u+v \\ a+2u & a+u+v & a+2v \\ a+v & a+2u+2v & a+u. \end{array}$$

(See Martin Gardner [3].) Note that any such magic square can be decomposed into three arithmetic progressions:

$$\begin{gathered} a, a+u, a+2u; \\ a+v, a+u+v, a+2u+v; \\ a+2v, a+u+2v, a+2u+2v. \end{gathered}$$

Each of these three sequences has the same difference, u, between terms. Note also that corresponding terms of the three sequences are in arithmetic progression, with common difference v. Conversely, any set of three arithmetic progressions of length three with a common difference, and corresponding terms in arithmetic progression, can be rearranged into a three-by-three magic square.

For example, if $a = 1$, $u = 1$, and $v = 3$, we get the familiar magic square:

$$\begin{array}{ccc} 8 & 1 & 6 \\ 3 & 5 & 7 \\ 4 & 9 & 2 \end{array}$$

The first equivalent formulation, **P1**, of the original problem should now be clear.

Squares in arithmetic progression It is well known that it is possible to have three squares in arithmetic progression, but not four (Dickson [4, pp. 435–440]). For any increasing three-term arithmetic progression of pairwise relatively prime squares, r^2, s^2, t^2, there are positive integers p and q such that

$$\begin{aligned} r &= |p^2 - 2pq - q^2|, \\ s &= p^2 + q^2, \\ t &= p^2 + 2pq - q^2, \end{aligned} \qquad (*)$$

p and q are relatively prime, and one of them is even (Dickson [4, pp. 437–438]). For example, if $r = 1$, $s = 5$, and $t = 7$, then $p = 2$ and $q = 1$.

If r^2, s^2, t^2 are in increasing arithmetic progression, but are not relatively prime, then there are k, p, and q, with k a positive integer, p and q as above, and

$$\begin{aligned} r &= k|p^2 - 2pq - q^2|, \\ s &= k(p^2 + q^2), \text{ and} \\ t &= k(p^2 + 2pq - q^2). \end{aligned}$$

For the r^2, s^2, t^2 just above, the difference between terms is

$$s^2 - r^2 = t^2 - s^2 = 4k^2(p^3q - pq^3).$$

Thus the original problem can be stated as find k_1, p_1, q_1, k_2, p_2, q_2, k_3, p_3, and q_3 so that

$$k_1^2(p_1^3q_1 - p_1q_1^3) = k_2^2(p_2^3q_2 - p_2q_2^3) = k_3^2(p_3^3q_3 - p_3q_3^3) > 0,$$

and

$$k_1^2(p_1^2+q_1^2)^2,\ k_2^2(p_2^2+q_2^2)^2,\ \text{and}\ k_3^2(p_3^2+q_3^2)^2$$

are distinct and in arithmetic progression. (Note that the fact that one cannot have four squares in arithmetic progression makes unnecessary any further restrictions on the "horizontal" and "vertical" differences between terms.)

It is easy to generate any number of three-term arithmetic progressions of squares, all with the same difference between terms, as we now show. Let u^2, v^2, and w^2 be in arithmetic progression. Let $p=v^2$ and $q=v^2-u^2$. Then for the three-term arithmetic progression generated by p and q using $(*)$, the difference between terms is $4u^2v^2w^2(v^2-u^2)$, which is a perfect square times v^2-u^2. Multiplying each term of the sequence u^2, v^2, w^2 by $4u^2v^2w^2$ gives a sequence with the same difference as the sequence generated by p and q. This process of generating a new sequence from a previous one (including the step of multiplying all previous sequences by the appropriate constant so that all sequences have the same difference between terms) can be continued indefinitely. If the new sequence is always derived from the last sequence generated, then all the sequences will be different. This is not difficult to prove, but we do not do that here.

As an example, start with the sequence generated from $p=5$ and $q=2$ using $(*)$. These give the sequence $1^2, 29^2, 41^2$, with difference of terms 840. Next let $p=29^2=841$ and $q=840=29^2-1^2$. These give the sequence $1411199^2, 1412881^2, 1414561^2$, with difference of terms 840×2378^2. Not all sequences with difference a square times 840 are generated in this way. For example, the sequences generated by $p=6$ and $q=1$, and by $p=8$ and $q=7$ (and sequences generated from these two sequences) have differences between terms that are a square times 840, but are not included in the set of sequences generated from $p=5$ and $q=2$.

Pythagorean triples There are simple relationships between three-term arithmetic progressions of squares and Pythagorean triples. The latter are related to congruent numbers and rational points on elliptic curves, so these relationships will be of use to us.

Every three-term arithmetic progression of squares, r^2, s^2, t^2, can be associated with a Pythagorean triple, X, Y, Z, with $X^2+Y^2=Z^2$, by taking $X=(r+t)/2$, $Y=(t-r)/2$, and $Z=s$. Conversely, each Pythagorean triple generates a three-term arithmetic progression of squares by taking $r=X-Y$, $s=Z$, and $t=X+Y$. *Two three-term arithmetic progressions of squares have the same difference of terms if, and only if, the corresponding Pythagorean right triangles have the same area.* The second equivalent formulation, **P2**, of the original problem should now be clear.

Congruent numbers The square-free part of $XY/2$ (the result of dividing $XY/2$ by the largest possible integer square), where X, Y, Z is a Pythagorean triple, is (by definition) a *congruent number*. This is clearly also the square-free part of the difference between terms of the associated three-term arithmetic progression of squares.

It is more convenient to work with right triangles with square-free area. Note that if k is the largest integer such that k^2 divides $XY/2$, then the area of the triangle with sides X/k, Y/k, and Z/k is a square-free integer. In general, X/k, Y/k, and Z/k will not be integers.

Elliptic curves If n is a congruent number, there is a well-known mapping from rational right triangles with area n to rational points on the elliptic curve $y^2=x^3-n^2x$

given by

$$x = (Z/2)^2,\ y = (X^2 - Y^2)Z/8.$$

Koblitz [5] shows that for each such point, $P = (x, y)$, there is another rational point, Q, on the elliptic curve such that $2Q = P$ in the sense of the group structure (briefly described below) for points on elliptic curves. Conversely, each rational point on the elliptic curve that is the double of another point (except the point at infinity) corresponds to a rational right triangle with area n. See Koblitz [5] for further details on the correspondences between points on such elliptic curves and Pythagorean triples. The third equivalent formulation, **P3**, of the original problem should now be clear.

A group structure on an elliptic curve is described as follows. An elliptic curve consists of the points (x, y) that satisfy the defining equation, plus a *point at infinity*, which can be thought of as lying an infinite distance above the point $(0, 0)$. The inverse, or negative, of a point $P = (x, y)$ on the elliptic curve is the point $-P = (x, -y)$. The point at infinity is its own negative and is also the identity element for the group operation. Every vertical line intersects the point at infinity, and these are the only lines that intersect the point at infinity. If a line is tangent to the curve at some point, consider the line to intersect the curve twice there, unless the line is tangent to the curve at a point of inflection, in which case consider the line to intersect the curve three times at that point. With these conventions, if a line intersects the curve twice then the line intersects the curve exactly three times. This fact can be used to define a group operation, $\oplus$, by taking $P \oplus Q \oplus R = 0$ if P, Q, and R lie on the same straight line. That is, $P \oplus Q = -R$ if P, Q, and R are collinear. To determine $P \oplus P (= 2P)$ for a point other than the point at infinity, take the tangent through P, find the other point of intersection with the curve, and take the negative of this last point. If P and Q have rational coordinates, then $P \oplus Q$ will have rational coordinates. It is easy to see that $\oplus$ is commutative, that each group element has an inverse, and that the identity behaves as it should. That $\oplus$ is associative is more difficult. See Koblitz [5], or other references on elliptic curves for more details. The operation $\oplus$, as we have defined it, is not the only way to define a group structure on the elliptic curve (see Cassels [6]).

There is a relationship between the doubling of points on elliptic curves and the method given above to generate a new three-term arithmetic progression of squares from a given one. Namely, if the point P corresponds to the three-term arithmetic progression u^2, v^2, w^2, then $2P$ corresponds to the three-term arithmetic progression generated by $(*)$ with $p = v^2$ and $q = v^2 - u^2$.

One potential usefulness of the elliptic curve formulation is that, for a given congruent number n, the group structure of rational points on elliptic curves shows there are infinitely many candidates for terms in the needed arithmetic progression. Thus, one can list as many candidates as one wants. Ideally, one "solves" the elliptic curve, finding points that generate all rational points on the curve. Failing this, one can often at least find some integral or rational points on the elliptic curve, and use these to generate others. My experience has been that there usually are several integral points with x values between $-n$ and 0, from which other points can be found.

Elliptic curves of high rank might be more likely than curves of lower rank to have three points meeting the conditions of formulation **P3**. (It is a theorem [5] that the group of rational points for an elliptic curve is $T \times Z^r$ where T is the subgroup consisting of all elements of finite order. The *rank* is r.) Wada and Taira [7] compute the ranks of all elliptic curves of the form $y^2 = x^3 - n^2x$ for all but 77 congruent

$n \leq 10{,}000$. The curve has rank three for $n = 1254$, 2605, 2774, 3502, 4199, 4669, 4895, 6286, 6671, 7230, 7766, 8005, 9015, 9430, and 9654. Noda and Wada **[8]** has a table that is an essential part of the results given in **[7]**.

Martin Gardner (**[9, 10]**) also discusses this problem and gives some related results. He offers \$100 to the first person who constructs a three-by-three magic square of distinct squares.

REFERENCES

1. Richard Guy, *Unsolved Problems in Number Theory*, 2nd edition, Springer-Verlag, New York, 1994, Problem D15, pp. 170–171.
2. Martin LaBar, Problem 270, *College Math. J.*, 15 (1984), 69.
3. Martin Gardner, *Riddles of the Sphinx*, Mathematical Association of America, Washington, DC, 1987, pp. 136–137.
4. Leonard Eugene Dickson, *History of the Theory of Numbers*, Volume II, Chelsea, New York, 1952.
5. Neal Koblitz, *Introduction to Elliptic Curves and Modular Forms*, 2nd edition, Springer-Verlag, New York, 1993, pp. 1–50.
6. J. W. S. Cassels, *Lectures on Elliptic Curves*, Cambridge University Press, Cambridge, UK, 1991, pp. 27–31.
7. Hideo Wada and Mayako Taira, Computations of the rank of elliptic curve $y^2 = x^3 - n^2x$, *Proc. Japan Acad.*, 70, Ser. A (1994), 154–157.
8. Kazunari Noda and Hideo Wada, All congruent numbers less than 10000, *Proc. Japan Acad.*, 69, Ser. A (1993), 175–178.
9. Martin Gardner, The magic of 3×3, *Quantum*, 6 (1996), January/February, pp. 24–26.
10. Martin Gardner, Letters, *Quantum*, 6 (1996), March/April, p.60.

General Russian Roulette

GUNNAR BLOM
Department of Mathematical Statistics, University of Lund,
Box 118, S-221 00 Lund, Sweden

JAN-ERIC ENGLUND
Swedish University of Agricultural Sciences,
Box 35, S-230 53 Alnarp, Sweden

DENNIS SANDELL
Biostatistics and Data Processing, Astra Draco AB,
Box 34, S-221 00 Lund, Sweden

1. Russian roulette Russian roulette provides a standard exercise in probability. Let us quote from **[1]**, p. 32:

> Russian roulette is played with a revolver equipped with a rotatable magazine of six shots. The revolver is loaded with one shot. The first duellist, A, rotates the magazine at random, points the revolver at his head and presses the trigger. If, afterwards, he is still alive, he hands the revolver to the other duellist, B, who acts in the same way as A. The players shoot alternately in this manner, until a shot goes off. Determine the probability that A is killed.

The answer is 6/11.

$n \leq 10{,}000$. The curve has rank three for $n = 1254$, 2605, 2774, 3502, 4199, 4669, 4895, 6286, 6671, 7230, 7766, 8005, 9015, 9430, and 9654. Noda and Wada **[8]** has a table that is an essential part of the results given in **[7]**.

Martin Gardner (**[9, 10]**) also discusses this problem and gives some related results. He offers \$100 to the first person who constructs a three-by-three magic square of distinct squares.

REFERENCES

1. Richard Guy, *Unsolved Problems in Number Theory*, 2nd edition, Springer-Verlag, New York, 1994, Problem D15, pp. 170–171.
2. Martin LaBar, Problem 270, *College Math. J.*, 15 (1984), 69.
3. Martin Gardner, *Riddles of the Sphinx*, Mathematical Association of America, Washington, DC, 1987, pp. 136–137.
4. Leonard Eugene Dickson, *History of the Theory of Numbers*, Volume II, Chelsea, New York, 1952.
5. Neal Koblitz, *Introduction to Elliptic Curves and Modular Forms*, 2nd edition, Springer-Verlag, New York, 1993, pp. 1–50.
6. J. W. S. Cassels, *Lectures on Elliptic Curves*, Cambridge University Press, Cambridge, UK, 1991, pp. 27–31.
7. Hideo Wada and Mayako Taira, Computations of the rank of elliptic curve $y^2 = x^3 - n^2x$, *Proc. Japan Acad.*, 70, Ser. A (1994), 154–157.
8. Kazunari Noda and Hideo Wada, All congruent numbers less than 10000, *Proc. Japan Acad.*, 69, Ser. A (1993), 175–178.
9. Martin Gardner, The magic of 3×3, *Quantum*, 6 (1996), January/February, pp. 24–26.
10. Martin Gardner, Letters, *Quantum*, 6 (1996), March/April, p.60.

General Russian Roulette

GUNNAR BLOM
Department of Mathematical Statistics, University of Lund,
Box 118, S-221 00 Lund, Sweden

JAN-ERIC ENGLUND
Swedish University of Agricultural Sciences,
Box 35, S-230 53 Alnarp, Sweden

DENNIS SANDELL
Biostatistics and Data Processing, Astra Draco AB,
Box 34, S-221 00 Lund, Sweden

1. Russian roulette Russian roulette provides a standard exercise in probability. Let us quote from **[1]**, p. 32:

> Russian roulette is played with a revolver equipped with a rotatable magazine of six shots. The revolver is loaded with one shot. The first duellist, A, rotates the magazine at random, points the revolver at his head and presses the trigger. If, afterwards, he is still alive, he hands the revolver to the other duellist, B, who acts in the same way as A. The players shoot alternately in this manner, until a shot goes off. Determine the probability that A is killed.

The answer is 6/11.

2. Generalization In this article we will consider the following generalization. There are n participants $A_1, A_2, \ldots, A_n$, where $n \geq 2$. Each person has one revolver. At each trial the probability is p that a shot goes off, independently of what happens at other trials. The participants shoot in circular order

$$A_1 A_2 \ldots A_n A_1 A_2 \ldots A_n \ldots .$$

First, A_1 uses his revolver, and either dies or survives. Thereafter, A_2 uses his weapon, and either dies or survives, and so on until one person is left; he is the winner. We want to determine the probability P_{in}, $i = 1, 2, \ldots, n$, that A_i is the winner.

In order to avoid unpleasant associations in our subsequent discussions, we will now replace the revolvers with coins, which turn up heads with probability p and tails with probability $q = 1 - p$. When a player tosses his coin and obtains heads, he disappears from the list $A_1 A_2 \ldots A_n A_1 A_2 \ldots$. The last person remaining on the list is the winner.

3. Two players Let two people play. If A_1 obtains heads at the first toss, he disappears and A_2 is the winner. If A_1 obtains tails, the roles of the players become interchanged. These arguments lead to the relation

$$P_{12} = p \cdot 0 + q(1 - P_{12}),$$

and so we find

$$P_{12} = \frac{q}{1+q}; \quad P_{22} = \frac{1}{1+q}.$$

If $p = 1/6$ we obtain classical Russian roulette with the probabilities $5/11$ and $6/11$, respectively.

4. First recursive solution For any number of players, the P_{in}'s can be found recursively, beginning with two players, thereafter continuing with three, and so on.

(a) *Three players.*
The players toss in the order $A_1 A_2 A_3 A_1 A_2 A_3 \ldots$. There are two main cases:

(i) The first toss results in heads. Player A_1 disappears. Players A_2 and A_3 remain, and for the rest of the game they take the places of A_1 and A_2, respectively, in the problem for two players.
(ii) The first toss results in tails. Players A_1, A_2, A_3 remain, and for the rest of the game they take the places of A_3, A_1 and A_2 in the original problem for three players.

Applying these considerations three times, we obtain the system of equations

$$P_{13} = p \cdot 0 + qP_{33}$$

$$P_{23} = pP_{12} + qP_{13}$$

$$P_{33} = pP_{22} + qP_{23}.$$

We already know P_{12} and P_{22}, so solve the system with respect to P_{13}, P_{23} and P_{33}. The solution is

$$P_{13} = \frac{pq}{1+q} + \frac{q^3}{1+q+q^2}$$

$$P_{23} = \frac{q}{1+q+q^2}$$

$$P_{33} = \frac{p}{1+q} + \frac{q^2}{1+q+q^2}.$$

(b) *Four players.*
When four players participate, we first solve the problem for three players and determine the P_{i4}'s from the system of equations

$$P_{14} = p \cdot 0 + qP_{44}$$
$$P_{24} = pP_{13} + qP_{14}$$
$$P_{34} = pP_{23} + qP_{24}$$
$$P_{44} = pP_{33} + qP_{34}.$$

It is now clear how the problem is solved for any given number of players:
We have $P_{1n} = qP_{nn}$ and

$$P_{in} = pP_{i-1,n-1} + qP_{i-1,n},$$

where $i = 2, \ldots, n$.

5. Second recursive solution We begin the second recursive solution by constructing a recursion for P_{1n}.

If at the first toss A_1 obtains heads, he does not win the game; on the other hand, if he obtains tails, he will appear at the beginning of the second round. Suppose that there are $k+1$ people on the list after the first round. This happens if k of the players $A_2, \ldots, A_n$ obtain tails during the first round; according to the binomial distribution this happens with probability

$$\binom{n-1}{k} q^k p^{n-1-k}.$$

On the other hand, when there are $k+1$ people on the list, the probability that, counted from the second round onwards, A_1 wins the games is $P_{1,k+1}$. Summing over the binomial probabilities we obtain the recursion

$$P_{1n} = q \sum_{k=0}^{n-1} \binom{n-1}{k} q^k p^{n-1-k} P_{1,k+1},$$

starting with $P_{11} = 1$.

We are now able to construct a recursion for P_{in}, $i \geq 2$. Suppose that, in the first round, k of the players $A_1, \ldots, A_{i-1}$ obtain heads. This happens with probability

$$\binom{i-1}{k} p^k q^{i-1-k}.$$

When A_i tosses his coin in the first round, he is first in a game with $n-k$ people, so he wins with probability $P_{1,n-k}$. Summing over the binomial probabilities, we obtain the recursion

$$P_{in} = \sum_{k=0}^{i-1} \binom{i-1}{k} p^k q^{i-1-k} P_{1,n-k}.$$

By first computing a suitable number of P_{1j}'s, we are able to find P_{in} for any $i \geq 2$ and n.

This recursive method requires a smaller number of operations than the method described in the previous section.

6. Explicit solution We will now derive an explicit expression for the probability P_{in} that A_i wins. Let us then suppose that the game is prolonged until the winner, though being alone, goes on tossing until he obtains heads. In the main part of the solution we will assume that $1 < i < n$.

Let B_j be the event that A_i obtains heads for the first time in the $(j+1)$st round, where $j = 0, 1, \ldots$. (Remember the prolongation of the game.) The events B_j are, of course, disjoint, and we have $P(B_j) = q^j p$. If B_0 occurs, A_i can never win, so we exclude this case. Given that B_j, $j > 0$, occurs, A_i wins if the following events C_j and D_j occur:

C_j: Players $A_1, A_2, \ldots, A_{i-1}$ obtain heads before A_i, that is, in the $(j+1)$st round or earlier.

D_j: Players $A_{i+1}, A_{i+2}, \ldots, A_n$ obtain heads before A_i, that is, in the jth round or earlier.

The probability that A_1, say, obtains heads at the $(j+1)$st round or earlier is $1 - q^{j+1}$. Hence we have

$$P(C_j) = (1 - q^{j+1})^{i-1}.$$

Similarly we find

$$P(D_j) = (1 - q^j)^{n-i}.$$

The three events B_j, C_j and D_j are independent. Summing over j we obtain

$$P_{in} = \sum_{j=1}^{\infty} P(B_j C_j D_j) = \sum_{j=1}^{\infty} P(B_j) P(C_j) P(D_j),$$

and so we arrive at the final expression

$$P_{in} = p \sum_{j=1}^{\infty} (1 - q^{j+1})^{i-1} (1 - q^j)^{n-i} q^j. \tag{1}$$

We leave it as an exercise to the reader to verify that this expression holds also for $i = 1$. For $i = n$ the summation runs from 0 to ∞.

It follows from (1) that if $0 < p < 1$ then $P_{1n} < P_{2n} < \cdots < P_{nn}$. This is no surprise: Remember that A_1 begins and hence has the smallest chance to win. We also note

that $P_{1n} = qP_{nn}$; this also follows directly from the recursive relations at the end of Section 4. As a consequence, when p is small and q is therefore near 1, the P_{in}'s are almost equal.

7. Asymptotics Russian roulette with very many people involved seems unlikely. Nevertheless, friends of asymptotic solutions may like to study the behavior of (1) when n is large.

For example, when $i = 1$ it is found that

$$P_{1n} = p \sum_{j=1}^{\infty} (1 - q^j)^{n-1} q^j.$$

Replacing the sum with an integral and performing the integration we obtain

$$P_{1n} \approx -\frac{p}{n \ln q}.$$

More generally, we have

$$P_{in} \approx -\frac{p}{\ln q} \cdot \frac{1}{n - (i-1)p}.$$

The approximations become better when n grows and/or p decreases; see Table 1 for some very good values for $n = 5$.

TABLE 1. Exact and approximate winning probabilities for the two cases $n = 5$, $p = 1/6$ and $n = 5$, $p = 1/2$.

P_{i5}	$p = 1/6$		$p = 1/2$	
i	*Exact*	*Approx.*	*Exact*	*Approx.*
1	0.1828	0.1828	0.1447	0.1443
2	0.1904	0.1891	0.1628	0.1603
3	0.1989	0.1959	0.1862	0.1803
4	0.2084	0.2031	0.2169	0.2061
5	0.2194	0.2110	0.2894	0.2404

REFERENCE

1. G. Blom, *Probability and Statistics: Theory and Applications*, Springer-Verlag, New York, NY, 1989.

Parallels on the Sphere

J. SCHAER
University of Calgary
Calgary, Alberta, Canada T2N 1N4

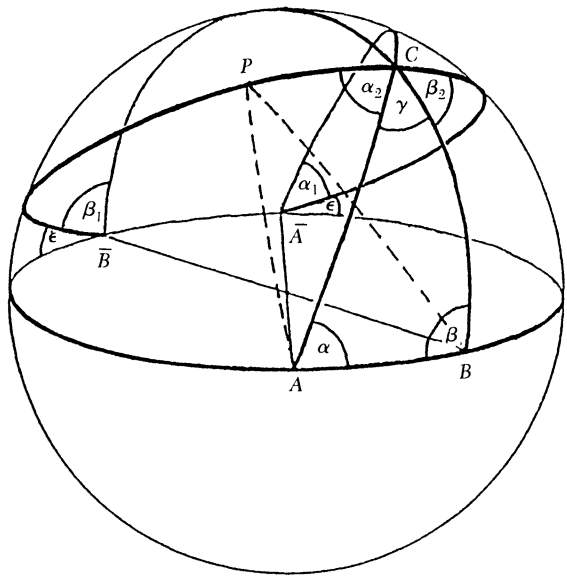
In the plane, parallels are usually defined as lines that do not meet. On the sphere, what corresponds to a line in the plane is a great circle, a straightest possible curve. And since any two distinct great circles intersect in two antipodal points, there are no parallels on the sphere. (In projective geometry antipodal points are identified, so any two distinct "lines" always intersect in exactly one "point.")

One can define a parallel to line L in another way, as the locus of all points on one side of L that have a constant distance from L. This definition would make the famous parallel axiom a theorem if we knew that such a parallel was a line. On the sphere that locus is not a great circle but rather a parallel circle.

A third way to define a parallel uses area: The locus of all points P on one side of the line AB, that form a triangle ABP with given fixed base and constant area. Since the base is fixed, the height of the triangle must be constant, and we fall back to the second definition. On the sphere, the area of a triangle is not simply half of base $\times$ height and so this "parallel" is a different curve.

THEOREM. *Given a spherical triangle ABC (with all sides* $< \pi$*). If P is any point on the circular arc* $\overline{A}C\overline{B}$*, where* $\overline{A}, \overline{B}$ *are the antipodal points of A, B, then the triangles ABP and ABC have equal area.*

Proof. (See FIGURE.) $\alpha = \alpha_1 + \epsilon$, $\beta = \beta_1 + \epsilon$; but $\alpha_2 = \alpha_1$, $\beta_2 = \beta_1$, and $\alpha_2 + \beta_2 + \gamma = \pi$. So $\alpha + \beta + \gamma = \alpha_2 + \epsilon + \beta_2 + \epsilon + \gamma = \pi + 2\epsilon$, hence the area of the triangle ABC is 2ϵ. This area depends only on the locations of $\overline{A}$ and $\overline{B}$, and on ϵ; hence if P is on the circular arc $\overline{A}C\overline{B}$ then the triangle $\overline{A}P\overline{B}$ has also area 2ϵ. So the circular arc $\overline{A}C\overline{B}$ is the "parallel" (in the third sense) through C of the "line" AB.



FIGURE

On the Convergence of Hillam's Iteration Scheme

BERND-JÜRGEN FALKOWSKI
FAST
Arabellastr. 17
D-81925 München, Germany

In their excellent paper [1] on digital halftoning, Geist, et al., develop an interesting approach to neural network simulation. In this context they reaffirm a conjecture on the convergence of a certain numerical iteration scheme originally due to Hillam, cf. [2], since they found substantial numerical evidence supporting it. In this note we prove the conjecture under a more restrictive condition than the one given in [1]. Moreover, we present a numerical example providing evidence that the original conjecture does not hold.

1. Hillam's Theorem In [2], Hillam established the following, at first sight remarkable, result for functions on the real line:

1.1. THEOREM. *If* $f\colon [a,b] \to [a,b]$ *satisfies a Lipschitz condition with constant* K, *i.e., if*

$$|f(x) - f(y)| \le K|x - y|$$

holds for all x, y *in* $[a,b]$, *then the iteration scheme*

$$x_{n+1} := (1-\lambda)x_n + \lambda f(x_n)$$

where $\lambda = 1/(K+1)$, *converges to a fixed point of* f.[1]

On the conjecture that this result might extend to higher dimensions, Hillam noted that a completely new approach would be needed, since his proof relied heavily on the total ordering of the real line. In [1] Geist, et al., restate this conjecture and offer numerical evidence for its support. As generalization of the Lipschitz condition they use

$$|f(x) - f(y)|_{\max} \le K|x - y|_{\max} \tag{1}$$

where $|\cdot|_{\max}$ denotes the maximum norm on $\mathbb{R}^n$. In order to deal with Hillam's remark, first we define a new function $F\colon [a,b] \to [a,b]$ by

$$F(x) := (1-\lambda)x + \lambda f(x).$$

The iteration scheme may then be rewritten as

$$x_{n+1} := F(x_n).$$

With this definition of F we obtain

1.2. LEMMA. *F is monotonically increasing.*

[1]Note: We are only concerned with the case $K > 1$ since otherwise there are iteration schemes known that converge to a fixed point; cf. [3].

Proof Suppose $x \geq y$. Then we have

$$f(y) - f(x) \leq |f(y) - f(x)| \leq K|y - x| = K(x - y). \tag{2}$$

Setting $\lambda := 1/(K+1)$ as before, we get $K = (1-\lambda)/\lambda$ and thus inequality (2) reduces to

$$(1-\lambda)x + \lambda f(x) \geq (1-\lambda)y + \lambda f(y). \tag{3}$$

Inequality (3) says that $F(x) \geq F(y)$.

We observe as a somewhat surprising effect that the Lipschitz condition on f implies the monotonicity of F. Now, of course, Hillam's result seems rather less unusual in view of the well-known Tarski fixed-point theorem for lattices, cf. [4]. As far as the convergence of our iteration scheme is concerned, we obtain from (1.2) the following.

1.3. LEMMA. *The sequence* (x_n) *defined inductively by*

$$x_0 := a$$

$$x_{n+1} := F(x_n)$$

converges to a fixed-point x *of* F, *which is also a fixed-point for* f.

Proof. $F(a) \geq a$ by the definition of F. Hence $x_1 \geq x_0$ and thus $F(x_1) = x_2 \geq F(x_0) = x_1$ by monotonicity of F. By induction, $x_{n+1} \geq x_n$ for all n. Since a bounded monotonic sequence converges, $\lim_{n\to\infty} x_n = x$ exists. From the continuity of F, which is a consequence of the continuity of f, which in turn follows from the Lipschitz condition, we clearly have $F(x) = x$. Finally we compute $f(x) = x$ from the definition of F.

Note at this stage that the result of (1.3) is slightly weaker than Hillam's result since we have to start with $x_0 = a$. On the other hand, the proof imitates the proof of Tarski's fixed-point theorem, and admits an easy generalization to higher dimensions since it does not use the fact that the real line is totally ordered.

2. A generalization of Hillam's result Let us first fix some notation: For $\underline{x} := (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ we set

$$|\underline{x}|_{vn} := (|x_1|, |x_2|, \ldots, |x_n|) \in \mathbb{R}^n.$$

Moreover for $\underline{a} := (a_1, a_2, \ldots, a_n)$ and $\underline{b} := (b_1, b_2, \ldots, b_n) \in \mathbb{R}^n$ we define $[\underline{a}, \underline{b}] := \{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \mid a_i \leq x_i \leq b_i \text{ for } 1 \leq i \leq n\}$. Further let the partial order relation on $\mathbb{R}^n$ be defined as usual, i.e.,

$$(x_1, x_2, \ldots, x_n) \leq (y_1, y_2, \ldots, y_n)$$

if, and only if,

$$x_i \leq y_i \text{ for } 1 \leq i \leq n.$$

Then we say that a function $g\colon \mathbb{R}^n \to \mathbb{R}^n$ satisfies a *modified Lipschitz condition* with constant L if

$$|g(\underline{x}) - g(\underline{y})|_{vn} \leq L|\underline{x} - \underline{y}|_{vn}. \tag{4}$$

With these definitions we can easily prove the following generalization of Hillam's result to n dimensions.

2.1. THEOREM. *Suppose that* $g\colon [\underline{a}, \underline{b}] \to [\underline{a}, \underline{b}]$ *satisfies the modified Lipschitz condition* (4). *Then the iteration scheme defined by*

$$\underline{x}_0 := \underline{a}$$
$$\underline{x}_{n+1} := (1-\mu)\underline{x}_n + \mu g(\underline{x}_n)$$

where $\mu := 1/(L+1)$, *converges to a fixed-point* $\underline{x}$ *of* g.

Proof. Repeating the calculations of 1.2, we see that the function G defined by $G(\underline{x}) := (1-\mu)\underline{x} + \mu g(\underline{x})$ is monotonic with respect to the partial order defined on $\mathbb{R}^n$. This in turn immediately implies that the sequence

$$\underline{x}_0 := \underline{a}$$
$$\underline{x}_{n+1} := G(\underline{x}_n)$$

is monotonically increasing (as in 1.3). Hence all coordinate sequences are monotonically increasing and thus convergent and so $\lim_{n\to\infty} \underline{x}_n = \underline{x}$ exists. By continuity again $\underline{x}$ must be a fixed-point for G and thus for g.

Remark 1. Our modified Lipschitz condition is one possible "natural" extension of the 1-dimensional Lipschitz condition. Unfortunately, it is obviously stronger than condition (1) that was suggested in [**1**]. Nevertheless, we feel that it makes sense to use it, since it guarantees the monotonicity of the function G as in the 1-dimensional case, which is crucial for convergence. Indeed, one might well ask, whether monotonicity would not be a more natural condition to use in the first place!

Remark 2. Functions satisfying the modified Lipschitz condition (4) may easily be constructed as follows. Let $g_i\colon [a_i, b_i] \to [a_i, b_i]$ be functions satisfying the Lipschitz conditions $|g_i(x) - g_i(y)| \le L_i|x-y|$. Let $\pi_i\colon \mathbb{R}^n \to \mathbb{R}$ be defined by $\pi_i(x_1, x_2, \ldots, x_n) := x_i$, and let L be given by $L := \max_i L_i$. Then

$$g(\underline{x}) := \big(g_1(\pi_1(\underline{x})), g_2(\pi_2(\underline{x})), \ldots, g_n(\pi_n(\underline{x}))\big)$$

obviously obeys condition (4).

3. Numerical evidence against the Hillam conjecture Since our modified Lipschitz condition (4) seems rather restrictive, we shall provide numerical evidence against the original conjecture using a function $g\colon [\underline{a}, \underline{b}] \to [\underline{a}, \underline{b}]$ in $\mathbb{R}^2$, where $\underline{a} := (0,0)$ and $\underline{b} := (1,1)$. We define g as follows. Let

$$f_1(x) := \begin{cases} 1-3x & \text{for } 0 \le x \le \frac{1}{3} \\ 0 & \text{for } \frac{1}{3} \le x \le 1 \end{cases}$$
$$f_2(x) := \sin(\pi x) \quad \text{for } 0 \le x \le 1$$

and set

$$g(x_1, x_2) := (f_1(x_2), f_2(x_1)).$$

Then g satisfies the Lipschitz condition

$$|g(\underline{x}) - g(\underline{y})|_{\max} \le \pi |\underline{x} - \underline{y}|_{\max}.$$

Using the value 3.141592654 for π we find that the function

$$G(\underline{x}) := (1-\mu)\underline{x} + \mu g(\underline{x}),$$

where

$$\mu := \frac{1}{(\pi+1)},$$

gives rise to a sequence defined inductively by

$$\underline{x_0} := (0,0)$$
$$\underline{x_{n+1}} := G(\underline{x_n}).$$

The sequence begins to cycle for $n \approx 250$ through the nine different values given by

$$(0.0922915121, 0.247146055),$$
$$(0.132437981, 0.25650258),$$
$$(0.156113482, 0.29215620),$$
$$(0.148246434, 0.335343035),$$
$$(0.112451887, 0.362803981),$$
$$(0.0853000404, 0.358740632),$$
$$(0.0647040891, 0.336054069),$$
$$(0.0490810923, 0.303656586),$$
$$(0.0587269346, 0.267420753).$$

This somewhat surprising result provides strong numerical evidence that Hillam's conjecture does not hold if the Lipschitz condition is defined using the maximum norm.

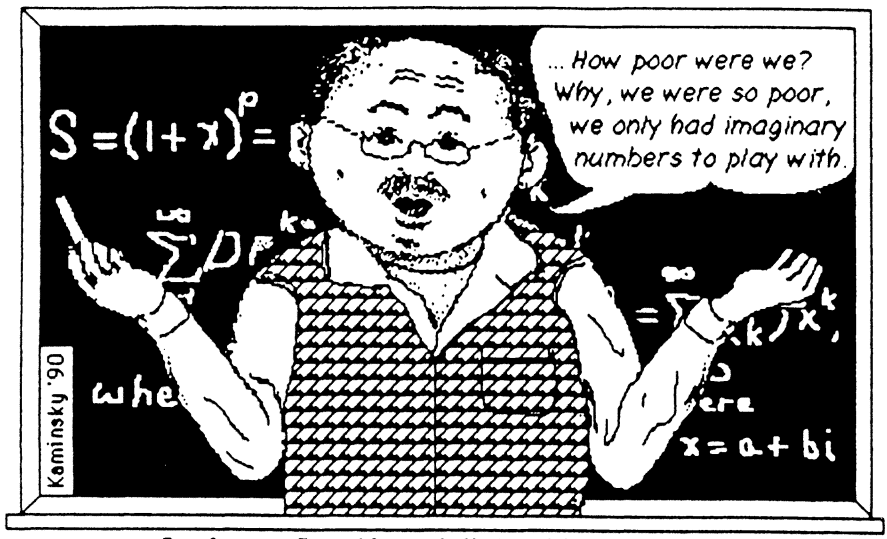
Concluding Remark. It would be most interesting to find other conditions that guarantee convergence of the iteration scheme since 2.1 doesn't seem to cover the special case considered in [**1**].

REFERENCES

1. Geist, R., Reynolds, R., and Suggs, D., A Markovian framework for digital halftoning, *ACM Transactions on Graphics*, 12 (1993), 136–159.
2. Hillam, B. P., A generalization of Krasnoselski's Theorem on the real line, this MAGAZINE 48 (1975), 167–168.
3. Krasnoselski, M. A., Two remarks on the method of successive approximations, *Uspehi Math. Nauk* (N.S.), 10 (1955), 123–127.
4. Tarski, A., A lattice theoretical fix-point theorem and its applications, *Pacific Journal of Mathematics*, 5 (1955), 285–309.

Professor Fogelfroe

Professor F. Fogelfroe is Professor of Mathematics at ValuPak™ University, in Margo's Forehead, Minnesota.



...Professor Fogelfroe delights his students in the style of Johnny Carson, too...

—Kenneth Kaminsky
Augsburg College
Minneapolis, MN 55454

PROBLEMS

GEORGE T. GILBERT, *Editor*
Texas Christian University

ZE-LI DOU, KEN RICHARDSON, and SUSAN G. STAPLES, *Assistant Editors*
Texas Christian University

Proposals

To be considered for publication, solutions should be received by March 1, 1997.

1504. *Proposed by Erwin Just, emeritus, Bronx Community College, Bronx, New York.*

For which positive integers n does there exist a set of n distinct positive integers such that
(i) each member of the set divides the sum of all members of the set, and
(ii) none of its proper subsets with two or more elements satisfies (i)?

1505. *Proposed by Murray S. Klamkin, University of Alberta, Edmonton, Canada, and Cecil C. Rousseau, The University of Memphis, Memphis, Tennessee.*

Let a and b be positive numbers satisfying $a + b \geq (a - b)^2$. Prove that

$$x^a(1-x)^b + x^b(1-x)^a \leq \frac{1}{2^{a+b-1}}$$

for $0 \leq x \leq 1$, with equality if and only if $x = 1/2$.

1506. *Proposed by Wu Wei Chao, He Nan Normal University, Xin Xiang City, He Nan Province, China.*

Let I and O denote the incenter and circumcenter, respectively, of ΔABC. Assume ΔABC is not equilateral. Prove that $\angle AIO \leq 90°$ if and only if $2BC \leq AB + CA$, with equality holding only simultaneously.

1507. *Proposed by Howard Morris, Ridgeland, Mississippi.*

For what real values of a and b_0 does the sequence $(b_n)_{n \geq 0}$ defined by $b_{n+1} = e^{ab_n}$ converge?

We invite readers to submit problems believed to be new and appealing to students and teachers of advanced undergraduate mathematics. Proposals must, in general, be accompanied by solutions and by any bibliographical information that will assist the editors and referees. A problem submitted as a Quickie should have an unexpected, succinct solution.

Solutions should be written in a style appropriate for this MAGAZINE. *Each solution should begin on a separate sheet containing the solver's name and full address.*

Solutions and new proposals should be mailed to George T. Gilbert, Problems Editor, Department of Mathematics, Box 298900, Texas Christian University, Fort Worth, TX 76129, or mailed electronically (ideally as a LaTeX file) to g.gilbert@tcu.edu. Readers who use e-mail should also provide an e-mail address.

1508. *Proposed by Saul Stahl, University of Kansas, Lawrence, Kansas.*

Let $\det_n$ denote the determinant of the $n \times n$ matrix whose entries are independent random variables each of which has value 1 with probability p and value 0 with probability $1 - p$. Compute the mean and variance of $\det_n$ for each positive integer n.

Quickies

Answers to the Quickies are on page 000.

Q853. *Proposed by S. B. Karmakar, Piscataway, New Jersey, and Murray S. Klamkin, University of Alberta, Edmonton, Canada.*

Are there any positive integral solutions to the Fermat-type equation

$$x^{m/3} + y^{m/3} = z^{m/3},$$

where $m > 3$ is a given positive integer relatively prime to 3?

Q854. *Proposed by Eugene Sard, Huntington, New York.*

In acute triangle ABC with sides $AB < AC < BC$, which of the three inscribed squares has largest area?

Q855. *Proposed by Jens Peter Reus Christensen and Mogens Esrom Larsen, Københavns Universitet, København, Denmark.*

For positive integers m and n, prove that

$$\sum_{k=0}^{4n-3} e^{2\pi i k^m/(4n-2)} = 0.$$

Solutions

A Recursive Optimization — October 1995

1479. *Proposed by Donald E. Knuth, Stanford University, Stanford, California.*

Let m_n be the maximum value of the quantity

$$\frac{x_1}{(1+x_1+x_2+\cdots+x_n)^2} + \frac{x_2}{(1+x_2+\cdots+x_n)^2} + \cdots + \frac{x_n}{(1+x_n)^2}$$

over all nonnegative real numbers $(x_1, \ldots, x_n)$. At what point(s) does the maximum occur? Express m_n in terms of m_{n-1}, and find $\lim_{n\to\infty} m_n$.

Solution by David Zhu, Jet Propulsion Laboratory, Pasadena, California.

Consider

$$g(x) = \frac{a}{x+b} + \frac{x}{(x+b)^2},$$

where $a \geq 0$ and $b \geq 1$. Then $g(x)$ attains its (absolute) maximum $(1+a)^2/(4b)$ at $x = b(1-a)/(1+a)$.

Let

$$f_n(x_1, x_2, \ldots, x_n) = \frac{x_1}{(1+x_1+x_2+\cdots+x_n)^2} + \frac{x_2}{(1+x_2+\cdots+x_n)^2} + \cdots + \frac{x_n}{(1+x_n)^2}.$$

Fix $x_2, x_3, \ldots, x_n$, and view f_n as a function of x_1. Its maximum value is

$$\frac{(1+a_1)^2}{4} \cdot \frac{1}{1+x_2+\cdots+x_n} + \frac{x_2}{(1+x_2+\cdots+x_n)^2} + \cdots + \frac{x_n}{(1+x_n)^2},$$

for $x_1 = (1+x_2+\cdots+x_n)(1-a_1)/(1+a_1)$, where $a_1 = 0$.

As a function of x_2, the above expression attains its maximum value,

$$\frac{(1+a_2)^2}{4} \cdot \frac{1}{1+x_3+\cdots+x_n} + \frac{x_3}{(1+x_3+\cdots+x_n)^2} + \cdots + \frac{x_n}{(1+x_n)^2},$$

at $x_2 = (1+x_3+\cdots+x_n)(1-a_2)/(1+a_2)$, where $a_2 = (1+a_1)^2/4$.

Repeating this process leads to

$$\frac{(1+a_{n-1})^2}{4} \cdot \frac{1}{1+x_n} + \frac{x_n}{(1+x_n)^2},$$

which attains its maximum value, $(1+a_n)^2/4$, at $x_n = (1-a_n)/(1+a_n)$, where $a_n = (1+a_{n-1})^2/4$.

Thus, a_{n+1} is the maximum value of f_n, where a_n is defined by

$$a_1 = 0, \quad \text{and} \quad a_{n+1} = \frac{(1+a_n)^2}{4}, \text{ for } n \geq 1.$$

The maximum of f_n occurs at the point $(x_1, x_2, \ldots, x_n)$ which satisfies

$$x_n = \frac{1-a_n}{1+a_n},$$

$$x_{n-1} = (1+x_n)\frac{1-a_{n-1}}{1+a_{n-1}},$$

$$\cdots$$

$$x_1 = (1+x_2+\cdots+x_n)\frac{1-a_1}{1+a_1}.$$

It is easily verified that $a_n \geq a_{n-1}$ and $0 \leq a_n \leq 1$ if $0 \leq a_{n-1} \leq 1$. Hence, $x_1, x_2, \ldots, x_n$ are nonnegative. Since (a_n) is a bounded and monotonically increasing sequence, it converges. Let $\lim_{n\to\infty} a_n = a$. Thus a satisfies $a = (1+a)^2/4$, which implies $a = 1$.

Also solved by Anchorage Math Solutions Group, Rich Bauer, Robin Chapman (U.K.), Con Amore Problem Group (Denmark), Steve Deckelman, Robert L. Doucette, L. R. King, M. S. Klamkin (Canada), Bogdan Kotkowski, O. P. Lossers (The Netherlands), Heinz-Jürgen Seiffert (Germany), WMC Problems Group, and the proposer. There were two incomplete solutions and one incorrect solution.

A Partition Identity — October 1995

1480. *Proposed by Ron Rietz and John Holte, Gustavus Adolphus College, St. Peter, Minnesota.*

Prove that

$$\sum_{i_1=0}^{n}\sum_{i_2=0}^{i_1}\cdots\sum_{i_k=0}^{i_{k-1}} r^{i_1+i_2+\cdots+i_k} = \prod_{j=1}^{k}\frac{1-r^{n+j}}{1-r^j}$$

for $r \neq \pm 1$, $k = 1, 2, 3, \ldots$, and $n = 0, 1, 2, \ldots$.

Solution by F. C. Rembis, Clifton, New Jersey.

For notational simplicity let

$$\sigma(k,n) = \sum_{i_1=0}^{n}\sum_{i_2=0}^{i_1}\cdots\sum_{i_{k=0}}^{i_{k-1}} r^{i_1+i_2+\cdots+i_k} \text{ and } \pi(k,n) = \prod_{j=1}^{k}\frac{1-r^{n+j}}{1-r^j}.$$

We will show $\sigma(k,n) = \pi(k,n)$ by induction on $k+n$. The cases when $k=1$ or $n=0$ are easily checked, which include the case $k+n=1$. Suppose $\sigma(l,m) = \pi(l,m)$ for $l+m < k+n$. Since

$$\begin{aligned}\sigma(k,n) &= \sigma(k,n-1) + r^n\sigma(k-1,n)\\ &= \pi(k,n-1) + r^n\pi(k-1,n),\end{aligned}$$

for $k > 1$ and $n > 0$, we need to show $\pi(k,n) - \pi(k,n-1) = r^n\pi(k-1,n)$. Now

$$\pi(k,n) = \frac{1-r^{n+k}}{1-r^k}\pi(k-1,n) \text{ and } \pi(k,n-1) = \frac{1-r^n}{1-r^k}\pi(k-1,n),$$

so

$$\begin{aligned}\pi(k,n) - \pi(k,n-1) &= \left(\frac{1-r^{n+k}}{1-r^k} - \frac{1-r^n}{1-r^k}\right)\pi(k-1,n)\\ &= r^n\pi(k-1,n),\end{aligned}$$

and the proposition holds.

Comment. Harald Fripertinger points out that the coefficient of r^j in the expansion of the left-hand side is the number of partitions of j into at most k parts such that the largest part is at most n, and that the right-hand side is the product expansion of the q-binomial coefficient, citing Proposition 1.3.19 of R. P. Stanley's *Enumerative Combinatorics* for a proof of the equality.

Also solved by Anchorage Math Solutions Group, Michael H. Andreoli, Nirdosh Bhatnagar, David Callan, Robin Chapman (U.K.), Con Amore Problem Group (Denmark), Qais Haider Darwish (Oman), Jesse I. Deutsch, Robert L. Doucette, Harald Fripertinger (Austria), Brad Gubser, Bogdan Kotkowski, Kee-Wai Lau (Hong Kong), O. P. Lossers (The Netherlands), Can A. Minh (student), Jean-Claude Ndogmo (Cameroun) and Pavel Winternitz (Canada), Michael Vowe (Switzerland), and the proposer.

A Characterization of Constant Acceleration **October 1995**

1481. *Proposed by Murray S. Klamkin, University of Alberta, Edmonton, Alberta, Canada.*

It is known that if a point moves on a straight line with constant acceleration and s_1, s_2, s_3 are its positions at times t_1, t_2, t_3, respectively, then the constant acceleration is given by

$$2\left(\frac{(s_2-s_3)t_1+(s_3-s_1)t_2+(s_1-s_2)t_3}{(t_1-t_2)(t_2-t_3)(t_3-t_1)}\right).$$

Show that this property characterizes uniformly accelerated motion; that is, if a particle moves on a straight line and s_1, s_2, s_3 are its positions at any times t_1, t_2, t_3, respectively, then if

$$\frac{(s_2-s_3)t_1+(s_3-s_1)t_2+(s_1-s_2)t_3}{(t_1-t_2)(t_2-t_3)(t_3-t_1)} = \text{constant},$$

the motion is one of constant acceleration.

Solution by Victor Kutsenok, St. Francis College, Fort Wayne, Indiana.

Fix $t_2 \neq t_3$. Then

$$\frac{(s_2-s_3)t+(s_3-s)t_2+(s-s_2)t_3}{(t-t_2)(t_2-t_3)(t_3-t)} = \frac{a}{2}$$

for some real number a and $t \neq t_2, t_3$, where s is the position corresponding to time t. Then, $(s_2-s_3)t+(s_3-s)t_2+(s-s_2)t_3 = (a/2)(t-t_2)(t_2-t_3)(t_3-t)$ for all t. Solving for s yields a quadratic in t, so the given motion is one of constant acceleration with $s'' = a$.

Also solved by Anchorage Math Solutions Group, Stanley J. Becker, Joseph E. Chance, Robin Chapman (U.K.), John Christopher, Con Amore Problem Group (Denmark), Robert L. Doucette, Mordechai Falkowitz (Canada), Bogdan Kotkowski, Nick Lord (England), Jean-Claude Ndogmo (Cameroun) and Pavel Winternitz (Canada), F. C. Rembis, Xavier Retnam, Nora S. Thornber, Michael Vowe (Switzerland), Robert J. Wagner, WMC Problems Group, David Zhu, and the proposer.

Perfect Numbers in Terms of Triangular Numbers **October 1995**

1482. *Proposed by C. F. Eaton, Pepperell, Massachusetts.*

Show that all even perfect numbers, $P > 6$, are of the form $P = 1 + 9T_n$, where T_n is a triangular number of the form $T_n = n(n+1)/2$, $n = 8j + 2$.

Solution by Bogdan Kotkowski, Kent State University, Tuscarawas Campus, New Philadelphia, Ohio.

Let P be an even perfect number greater than 6. Then there exists a prime number $p \geq 3$ such that $P = 2^{p-1}(2^p - 1)$. Because $p - 3$ is even, $2^{p-3} - 1$ is divisible by 3. A simple calculation shows that

$$P = 1 + 9 \cdot \frac{1}{2} \cdot \left(8 \cdot \frac{2^{p-3}-1}{3} + 2\right)\left(8 \cdot \frac{2^{p-3}-1}{3} + 3\right).$$

Also solved by William B. Adams, Anchorage Math Solutions Group, Rich Bauer, Ryan Buschert (student), David Callan, Robin Chapman (U.K.), John Christopher, Con Amore Problem Group (Denmark), Charles R. Diminnie, Robert L. Doucette, Hugh Edgar, Roger B. Eggleton, L. L. Foster, Joe

Howard, D. E. Iannucci (Virgin Islands), Hans Kappus (Switzerland), Sidney Kravitz, Vernon J. Kunz, Kee-Wai Lau (Hong Kong), S. C. Locke, Nick Lord (England), O. P. Lossers (The Netherlands), David E. Manes, Don Redmond, F. C. Rembis, R. P. Sealy, Jamie Simpson (Australia), Lawrence Somer, Selvaratnam Sridharma, David R. Stone, Michael Vowe (Switzerland), Monte J. Zerger, David Zhu, and the proposer.

A Trigonometric Relation in Triangles — October 1995

1483. *Proposed by Alexandru Teodorescu-Frumosu, student, Boston University, Boston, Massachusetts.*

Let ABC be an arbitrary triangle, and let a, b, c, be the lengths of the sides BC, AC, AB, respectively. Let M be the midpoint of the segment BC, let $\alpha = \angle BAM$, $\beta = \angle CAM$ and $x = \angle AMB$. Show that

$$\frac{b}{\sin \alpha} = \frac{a \cos x}{\sin(\alpha - \beta)}.$$

Solution by Catherine Taylor, student, San Francisco University High School, San Francisco, California.

Applying the law of sines to ΔACM, we find

$$\frac{2 \sin \beta}{a} = \frac{\sin(\pi - x)}{b} = \frac{\sin x}{b}.$$

Applying the law of sines to ΔABC, we find

$$\frac{\sin(\alpha + \beta)}{a} = \frac{\sin(\pi - (\alpha + x))}{b} = \frac{\sin(\alpha + x)}{b},$$

or

$$\frac{\sin \alpha \cos \beta + \sin \beta \cos \alpha}{a} = \frac{\sin \alpha \cos x + \sin x \cos \alpha}{b}.$$

Subtracting $2 \sin \beta \cos \alpha / a = \sin x \cos \alpha / b$ from both sides, we get

$$\frac{\sin \alpha \cos \beta - \sin \beta \cos \alpha}{a} = \frac{\sin \alpha \cos x}{b},$$

or

$$\frac{\sin(\alpha - \beta)}{a} = \frac{\sin \alpha \cos x}{b}.$$

Dividing both sides by $\sin \alpha \sin(\alpha - \beta)/(ab)$ when $\alpha \neq \beta$, we get

$$\frac{b}{\sin \alpha} = \frac{a \cos x}{\sin(\alpha - \beta)}.$$

Also solved by Reza Akhlaghi, Anchorage Math Solutions Group, John Andraos (Australia), Francisco Bellot Rosado (Spain), Kenneth Bernstein, Nirdosh Bhatnagar, J. C. Binz (Switzerland), Robin Chapman (U.K.), John Christopher, Con Amore Problem Group (Denmark), Charles K. Cook, Robert L. Doucette, Milton P. Eisner, Mordechai Falkowitz (Canada), R. Govindaraj (India), Joe Howard, Paul Irwin, Hans Kappus (Switzerland), Jahangeer Kholdi, Bogdan Kotkowski, Victor Kutsenok, Kee-Wai Lau (Hong Kong), Nick Lord (England), O. P. Lossers (The Netherlands), V. S. "Mano" Manoranjan, Robert C. Maxell, Shoeleh Mutameni, S. A. Obaid, F. C. Rembis, Noah Rosenberg (student), Alan Shettel, Selvaratnam Sridharma, Maggie Tran (student), Ian VanderBurgh (Canada), Michael Vowe (Switzerland), Harry Weingarten, Andrew Ho Kuen Wu, Robert L. Young, Monte J. Zerger, Ted Zerger, David Zhu, and the proposer.

Answers

Solutions to the Quickies on page

A853. There are no positive integral solutions. If there were a solution, there would be one with $(x, y, z) = 1$. On cubing the equation, we get

$$x^m + y^m + 3(xyz)^{m/3} = z^m,$$

so that xyz must be a perfect cube. Either each of x, y, z must be a perfect cube or else two of them have a common prime factor p. In the former case we get Fermat's equation which is now known to have no solutions. In the latter case $(x, y, z) \neq 1$. Consequently, there are no solutions.

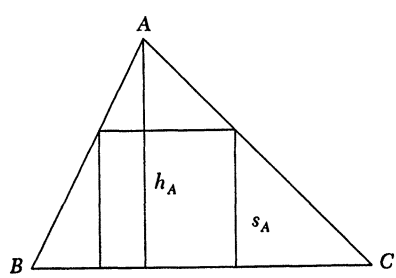
A854. Let h_A denote the length of the altitude from A to BC, and let s_A be the length of the side of the inscribed square with two vertices on BC, and so forth. From similar triangles, we see that

$$\frac{h_A - s_A}{h_A} = \frac{s_A}{BC} \quad \text{or} \quad s_A = \frac{BC \cdot h_A}{BC + h_A}.$$

Since $BC \cdot h_A$ is twice the area of ΔABC, the largest square corresponds to the smallest of $BC + h_A$, $AC + h_B$, and $AB + h_C$. Now,

$$\begin{aligned}(BC + h_A) - (AC + h_B) &= (BC + AC \sin\angle C) - (AC + BC \sin\angle C) \\ &= (BC - AC)(1 - \sin\angle C) > 0.\end{aligned}$$

Hence $s_A < s_B$. Similar reasoning implies $s_B < s_C$, so that the square inscribed on AB has largest area.



A855. We have

$$\sum_{k=0}^{4n-3} e^{2\pi i k^m/(4n-2)} = \sum_{k=0}^{2n-2} \left(e^{2\pi i k^m/(4n-2)} + e^{2\pi i (k+2n-1)^m/(4n-2)}\right).$$

Because $(k + 2n - 1) - k = 2n - 1$ is a factor of $(k + 2n - 1)^m - k^m$, it follows that $(k + 2n - 1)^m - k^m \equiv 2n - 1 \pmod{2n-1}$. It is clear that $(k + 2n - 1)^m - k^m \equiv 1 \equiv 2n - 1 \pmod 2$ as well. Thus $e^{2\pi i(k+2n-1)^m/(4n-2)} = -e^{2\pi i k^m/(4n-2)}$, hence

$$\sum_{k=0}^{4n-3} e^{2\pi i k^m/(4n-2)} = 0.$$

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 mathematics literature. Readers are invited to suggest items for review to the editors.

Conway, John H., and Allyn Jackson, Budding mathematician wins Westinghouse Competition, *Notices of the American Mathematical Society* 43 (7) (July 1996) 776–779.

Jacob Lurie of Bethesda, Maryland, won first prize in the 1996 Westinghouse Science Talent Search, a competition for high-school students. (Unfortunately, the first prize of $40,000 does not cover even two years of tuition at a leading university.) His paper treats computable sets in the surreal numbers, which are the "most natural collection of numbers that includes both the usual real numbers and the infinite ordinal numbers." They were discovered by author Conway in 1969, who applied them to analyzing combinatorial games. Surreal numbers were also featured in Donald Knuth's *Surreal Numbers: How Two Ex-Students Turned on to Pure Mathematics and Found Total Happiness* (1974). Mention of the book in historical vein in this column (December 1995: "Older readers (and younger ones who have explored the library) will remember ... ") prompted Knuth to advise that not only is the book still in print but it also has its own home page, at

`http://www-cs-faculty.stanford.edu/~knuth/sn.html`

which offers errata, translations, and hints for some of the exercises.

Thwaites, Bryan, Two conjectures or how to win £1100, *Mathematical Gazette* 80 (March 1996) 35–36.

Thwaites reminds readers of his origination in 1952 of what has become known variously as the $3n+1$ conjecture, Collatz conjecture, and many other names besides the "Thwaites conjecture." It says that starting from any positive integer n, iteration of the map

$$n \longrightarrow \begin{cases} 3n+1, & \text{if } n \text{ is odd;} \\ k, & \text{if } n \text{ is even and } n = k2^m \text{ with } k \text{ odd.} \end{cases}$$

always leads eventually to 1. Thwaites offers £1,000 reward for resolving the conjecture, and £100 for resolving another conjecture: Given any finite sequence of rational numbers, take the positive differences of successive members (including differencing the last member with the first); iteration of this operation eventually produces a set of zeros iff the size of the set is a power of 2.

Centenary Issue, *Mathematical Gazette* 80 (March 1996).

This issue celebrates 100 years of the *Gazette* (may all our readers still be enjoying a happy retirement when THIS MAGAZINE celebrates its centennial!). Commemorative articles discuss production of the *Gazette* and its history, plus reminiscences and recollections. Notable articles survey mathematics teaching and twentieth-century mathematics: Jean Dieudonné on "Mathematics of our day," Michael Atiyah on "Geometry and physics," David Lindley on statistics in the last 100 years, Peter Neumann on "A hundred years of finite group theory," and others.

Sterrett, Andrew (ed.), *101 Careers in Mathematics*, MAA, 1996; x + 260 pp, $20 (less to MAA members). ISBN 0–88385–704–9.

"What can I do with a major in mathematics?" The answer, of course, is "anything"; only 9% of male mathematics majors employed in the U.S. work in the mathematical sciences (but their average salary is second only to engineers). This is a book that the admissions staff of your institution need to pass around and have on their bookshelf (do yourself a favor and buy them a copy or two), and that your department needs to have in its common room. The book contains two-page first-person vignettes with photos of 101 people in "a wide variety of careers for which a background in the mathematical sciences is useful," plus articles on career and job-seeking advice reprinted from *Math Horizons*. Very few of the people featured appear to be over 50 (does this suggest that math majors die young?), and with one exception all have degrees from U.S. institutions and work in the U.S. William Perry (U.S. Secretary of State; Ph.D. in operations research); Alberto Fujimori (President of Peru; M.S., University of Wisconsin–Milwaukee), Alexander Solzhenitsyn (Nobel Prize for Literature; B.S., University of Rostov), and notables in general are not included, which is not to say that a background in the mathematical sciences has not been useful to them (e.g., it saved Solzhenitsyn from some forms of prison labor).

Courant, Richard, and Herbert Robbins, *What Is Mathematics? An Elementary Approach to Ideas and Methods*, 2nd ed., revised by Ian Stewart, Oxford University Press, 1996; xix + 566 pp, $18.95 (P). ISBN 0–19–510519–2.

Ian Stewart has brought this wonderful classic (first published in 1942) up to date by adding a new chapter of several pages each on a dozen topics: polynomials that produce all primes and the Jones knot polynomial; progress on the Goldbach conjecture and on soap films; proofs of Fermat's Last Theorem, the Four Color Theorem, and the Steiner ratio conjecture; the independence of the Continuum Hypothesis, fractals, and the rehabilitation of infinitesimals via nonstandard analysis.

Stewart, Ian, *From Here to Infinity: A Guide to Today's Mathematics*, Oxford Univ. Pr., 1996; x + 310 pp, $12.50 (P). ISBN 0–19–283202–6.

This is a revised and retitled edition of the magnificent *The Problems of Mathematics* (1987; 2nd ed., 1992) and complements the Courant and Robbins book above. "The new title is supposed to indicate that mathematics combines relevance to everyday life ('here') with sweeping intellectual invention ('infinity')." Mathematics majors will find it affordable easy reading about exciting contemporary mathematics; ask your bookstore to stock it.

Cipra, Barry, Lattices may put security codes on a firmer footing, *Science* 273 (23 August 1996) 1047–1048.

Given a lattice in Euclidean n-space, find a set of spanning vectors that have the shortest lengths. Milos Ajtai (IBM Almaden Research Center) has shown that there is no efficient algorithm for any positive fraction of such problems unless there is an efficient algorithm for all of them (and none is known). Hence randomly generated lattices could form the basis for digital signatures and authentication (append to your document a lattice problem whose solution you know) or even new codes.

Yam, Philip, Profile: Martin Gardner, The mathematical gamester, *Scientific American* 273 (6) (December 1995), 38–41; Puzzling with Martin Gardner, 26.

Biographical and personal profiles of the amateur magician, self-taught mathematician, Lewis Carroll expert, and long-time author of *Scientific American*'s Mathematical Games column. Three of his favorite puzzles are included.

Cole, K.C., Fairness by the numbers, *Los Angeles Times* (Washington Edition) (26 April 1996) A1, A8. Peterson, Ivars, Formulas for fairness: Applying the math of cake cutting to conflict resolution, *Science News* 149 (4 May 1996) 284–285. Brams, Steven J., and Alan D. Taylor, *Fair Division: From Cake-Cutting to Conflict Resolution*, Cambridge Univ. Pr., 1996; xiv + 272 pp, $18.95 (P). ISBN 0–521–55644–9. Taylor, Alan D., Fair division, Chapter 13 in *For All Practical Purposes*, 4th ed., edited by Solomon Garfunkel, W.H. Freeman, 1996. ISBN 0–387–94612–8.

Agreements on division of marital property or of an estate can falter on the different values that parties impute to indivisible goods. Countries in conflict over borders face a similar problem. Authors Brams and Taylor offer algorithms for such disputes that result in "envy-free" allocations, i.e., allocations in which everyone is satisfied that he or she has received more than anybody else. For two or three parties, the procedures are fairly simple; to go to four or more involves a great leap in complexity. The key question, however, is: Will a population of divorcing couples and their lawyers who are ignorant of—and deeply dislike—mathematics trust their welfare to mathematical procedures that they do not understand? As Robert E.D. "Gene" Woolsey (Colorado School of Mines) has often noted, "A manager would rather live with a problem he can't solve than with a solution he doesn't understand."

Cipra, Barry, A proof to please Pythagoras, *Science* 271 (22 March 1996) 1669.

Can the positive integer N be the area of a right triangle with rational sides? The integers 5 and 6 are, but 1, 2, 3, and 4 are not. The key to this problem lies not in Euclidean geometry or elementary number theory but in elliptic curves: Each such right triangle corresponds to a rational point on the elliptic curve $y^2 = x^3 - N^2x$. Such a curve has either infinitely many rational points with $y \neq 0$ or none. A particular criterion function is zero in the first case. Hence, if the criterion function is nonzero, N is not the area of a right triangle with rational sides. The new techniques about elliptic curves that were used in Wiles's proof of Fermat's Last Theorem may lead to a proof of the converse, that when the criterion function is zero, there is such a right triangle.

Stewart, Ian, Tales of a neglected number, *Scientific American* 274 (6) (June 1996) 102–103.

The number of the title is the *plastic number*, so-named because of a genesis similar to the golden ratio. The plastic number is the limiting ratio of successive terms of the *Padovan sequence* described by the recursion $P(n+1) = P(n-1) + P(n-2)$ with initial conditions $P(0) = P(1) = P(2) = 1$. Architect Richard Padovan used the plastic number in design; but unlike the golden ratio, the plastic ratio does not seem to have any manifestations in nature, and the sequence itself seems have no connections with other mathematics. However, the sequence with the same recursion but initial conditions $P(0) = 3$, $P(1) = 0$, and $P(2) = 2$, called the *Perrin sequence*, has an interesting property noticed by Édouard Lucas in 1876: If n is a prime, n divides $P(n)$. This result provides a speedy test (in $\log n$ steps) for nonprimality, but it is still unknown if there can be a composite n that divides $P(n)$ (called a *Perrin pseudoprime*).

Kulig, Christopher J., Winning at Quarto!, *Mathematics Teacher* 89 (5) (May 1996) 374–375.

Quarto! is a relatively new board game played on a 4×4 grid. Each of the 16 playing pieces displays a different combination of four binary properties: short/tall, light/dark, round/square, and solid/hollow. The players take turns placing a piece (chosen by the other player!), trying to be the first to create a row, column, or main diagonal of four pieces with the same property. Author Kulig shows how to create positions in which neither player wins.

Carl B. Allendoerfer Awards – 1996

The Carl B. Allendoerfer Awards, established in 1976, are made to authors of expository articles published in *Mathematics Magazine*. The Awards are named for Carl B. Allendoerfer, a distinguished mathematician at the University of Washington, and President of the Mathematical Association of America, 1959–60.

This year's awards were presented at the August 1996 Prizes and Awards Banquet, held in Seattle as part of the Joint Summer Meetings.

Judith Grabiner
"Descartes and Problem-Solving"
Mathematics Magazine **68 (1995) pp. 83–97**

This article deals with a big subject in a fascinating way, and the writing is superb. The subject is the "method" of Descartes. We learn, for example, that one of the fundamental and pervasive aspects of the method is working backward from an assumed solution—the original meaning of the word "analysis." Equally fundamental and pervasive is the idea that mathematics consists of solving problems, not deriving logical systems from first principles. We see what kind of geometric problems Descartes addressed, how he used his method to analyze them, and how his methods now pervade the practice of mathematics—very much as he intended they should. The article is an excellent example of the insight we can gain from an historical view of mathematics.

Biographical Note. Judith V. Grabiner is currently the Flora Sanborn Pitzer Professor of Mathematics and Professor of Science, Technology & Society at Pitzer College, Claremont, California. Educated at the University of Chicago (B.S. (Honors) in 1960) and Radcliffe College and Harvard University (M.A., 1962; Ph.D., 1966), Professor Grabiner is a leading historian of mathematics, having written two well-known books, *The Origins of Cauchy's Rigorous Calculus*, and *The Calculus of Algebra: J.-L. Lagrange*, as well as numerous articles. The MAA has previously honored her with a Lester R. Ford Award for an article on Cauchy in the *American Mathematical Monthly* in 1984. And this is her third Carl B. Allendoerfer Award, the earlier two for an article on the derivative from Fermat to Weierstrass, in 1984, and an article on the centrality of mathematics in the history of Western thought, in 1989. Her current work is on the mathematics of Maclaurin.

Response from Professor Grabiner. I owe this award first to Professor Tatiana Deretsky, who suggested the topic and who invited me to speak about it at a conference on the 350th anniversary of Descartes's Geometry at San Jose State University in 1987. I would also like to thank Paul Halmos and Jerry Alexanderson for their en-

couraging words about the talk, and Alfred Bloom (now president at Swarthmore) for a valuable debate about Descartes's role in European thought, which sharpened some of the ideas. I again thank the *Mathematics Magazine*'s referees, who made helpful suggestions for improvement; my husband Sandy for reading several drafts and saying "think about the audience," and also my precalculus, calculus, and history of mathematics students for listening to my discussions of problem-solving in the Cartesian manner. Finally, I thank the Allendoerfer Award Committee and the MAA.

Daniel J. Velleman and Gregory S. Call
"Permutations and Combination Locks"
Mathematics Magazine **68 (1995) pp. 243–253**

This article immediately draws in the reader with a nice conjunction of combinatorial and analytical reasoning. The authors take a simple, practical problem and develop the mathematics clearly and thoroughly. They bring forward techniques from diverse fields as they need them. The resolution includes a number of interesting combinatorial concepts, including asymptotic estimates. In particular, calculus and discrete mathematics are integrated in ways that an undergraduate might find surprising and intriguing. The writing is lucid and brisk; the reader is swept along but is never disoriented. Both the solution to the problem and the exposition are models for how these things should be done.

Biographical Notes. Dan Velleman received his bachelor's degree from Dartmouth College in 1976 and his doctorate from the University of Wisconsin in 1980. He taught at the University of Texas and the University of Toronto before joining the faculty of Amherst College in 1983. Dan is interested in logic, philosophy of mathematics, and the foundations of quantum mechanics. He is the author of the book *How to Prove It*, and a coauthor, with Joe Konhauser and Stan Wagon, of the forthcoming problem collection *Which Way Did the Bicycle Go?*. In 1994 he received a Lester R. Ford Award for the paper "Versatile Coins."

A native of Hanover, New Hampshire, Greg Call completed his A.B. degree at Dartmouth College in 1980. He did his graduate work at Harvard under John Tate, receiving his A.M. in 1981 and his Ph.D. in 1986. He taught for two years at Tufts University, before meeting and then joining Dan Velleman at Amherst College, where Greg is now an Associate Professor of Mathematics. With a dozen student members, he founded Amherst's Student Chapter of the MAA in the spring of 1990 and, except for sabbatical years at Brown in 1991–92 and Harvard in 1995–96, has served as the Chapter's Faculty Advisor ever since. While his primary research interests are in Diophantine geometry and algebraic number theory, Greg is always ready to collaborate with his good friend Dan on an interesting problem-of-the-week.

Response from Daniel J. Velleman. I am very pleased and honored to have been chosen for this award. I would like to thank my coauthor, Greg Call, for making this project what a mathematical collaboration should be: an enjoyable and productive exchange of ideas leading to a paper that is better than either of us would have written alone.

Response from Gregory S. Call. I would like to extend my sincere thanks to the MAA and to the Committee on

Allendoerfer Awards, in particular, for their generous recognition.

As mathematicians one of our great pleasures is working in collaboration to solve a challenging problem. Sharing our results and, whenever possible, explaining how they were discovered is an equal joy. Writing "Permutations and Combination Locks" gave me the opportunity to enjoy each of these pleasures. In addition, Dan Velleman and I tried to provide our students with an accessible model of mathematical research which we hope will encourage them to undertake their own investigations. I look forward to those investigations and the opportunity to share them with my colleagues in the MAA.

Letters to the Editor

Dear Editor:

I thank Josh Nichols-Barrer (Letters to the Editor, June 1996, p. 238) for bringing to light an error in my article *Continued powers and a sufficient condition for their convergence* (this MAGAZINE, December 1995, pp. 387–392). He points out that since it does not in fact violate my convergence condition for continued squares, my Example III doesn't show that the condition for general powers $p > 1$ is not necessary.

As my penance for publicly transgressing first-year calculus, I offer the following replacement for the lightly-conceived and ill-fated Example III. Consider the continued square

$$S = b+^2(0+^2(b+^2(0+^2(b+^2(0+^2(\ldots))))))$$

with $b = 3/(4^{4/3})$. This fails the convergence test for a continued square. With $p = 2$, we have $R = (p-1)/p^{p/(p-1)} = 1/4$, $x_n = b$ for n even and 0 for n odd, and

$$\left(\frac{x_n}{R}\right)^{p^n} = \begin{cases} [3/(4^{1/3})]^{2^n} & n \text{ even;} \\ 0 & n \text{ odd.} \end{cases}$$

The dominant subsequence of even terms results in an unbounded expression, and the test fails.

However, S is equivalent to the continued fourth power

$$b+^4(b+^4(b+^4(\ldots))),$$

which converges by the boundedness test: with $p = 4$, one has $R = 3/(4^{4/3})$ and $(x_n/R)^{p^n} = 1^{4^n} = 1$. The continued square S therefore converges, but since it fails the continued squares convergence test, the test remains sufficient but not necessary.

Dixon J. Jones
5112 Fairchild Avenue
Fairbanks, Alaska 99709–4523

Dear Editor:

Lawrence Zalcman has called my attention to references [2] and [3] below, related to the question at the bottom of page 92 in my article *Inverse problems for central forces*, this MAGAZINE 69 (April 1996), pp. 83–93. The answer is yes if the surface is smooth ([3]), but no in general ([2]).

He also pointed out the following corrections. Zagier obtained his proof in 1982 or 1983, not 1987, and the earlier proof required no smoothness assumption. On page 92, below display (12), "integrand" should be replaced by "integral." Finally, reference [11] in the paper should be [1], below.

REFERENCES

1. A. V. Kandraskov, On the uniqueness of the reconstruction of certain regions from their exterior gravitational potential, *Ill-posed Mathematical Problems and Problems of Geophysics*, Novosibirsk (1976), pp. 122–129 (Russian)

2. John L. Lewis and Andrew Vogel, On pseudospheres, *Revista Matematica Iberoamericana* 7 (1991), pp. 25–54

3. Henrik Shahgholian, A characterization of the sphere in terms of single-layer potentials, *PAMS* 115 (1992), pp. 1167–1168

S. K. Stein
University of California - Davis
Davis, California 95616–8633

Dear Editor:

Two recent MAGAZINE articles ([1] and [2]) show that for integers $a_1 < a_2 < \ldots < a_n$, the product of differences

$$\prod_{1 \le i < j \le n} (a_j - a_i)$$

is evenly divisible by

$$\prod_{1 \le i < j \le n} (j - i).$$

This problem has appeared in the USSR Mathematical Olympiad ([3], Problem 62). Readers may also be interested to know that when the a_i are all positive, the quotient counts a rather concrete class of combinatorial objects: the number of collections of n pairwise disjoint lattice paths (with unit steps *east* or *south*) joining $(0, a_i)$ to $(i-1, i-1)$ for $1 \le i \le n$ (see [4]).

REFERENCES

1. B. Sury, An integral polynomial, this MAGAZINE 68 (1995), pp. 134–135

2. Robin Chapman, A polynomial taking integer values, this MAGAZINE 69 (1996), p. 121

3. Shklarsky, Chentzov, and Yaglom, *USSR Olympiad Problem Book*, Dover, New York, NY, 1993

4. Ira Gessel and Gérard Vennot, Binomial determinants, paths, and hook length formulae, *Advances in Mathematics* 58 (1985), pp. 300–321

David Callan
Department of Statistics
University of Wisconsin - Madison
Madison, Wisconsin 53706

MicroCalc, ver. 7.0

Interactive Calculus Software

MicroCalc covers almost all topics needed for teaching and learning calculus: single variable, several variable, differential equations. (Programmed by Harley Flanders)

MicroCalc is menu-driven; there is no language to learn; its input is the way you write mathematics. *MicroCalc* is dedicated to calculus, and has desired calculus topics ready to run. Examples: generation of sine and tangent, graphs of Riemann sums, solids of revolution by slabs and by shells, implicit curves and surfaces, Newton steps for solving equations, graphical Lagrange multipliers. About 70 other topics, symbolic, numerical, graphical, are on *MicroCalc's* menus, plus several utilities.

MS-DOS platform only. 3.5" diskette with *install* program.

Site license fees:
First 50 workstations: $850; up to 50 additional: $10 each; still more workstations: $5 each. Upgrade from previous version: 40% off. (First class mail free. Express mail at cost.)

Sample disk:
$15 paid in advance; $25 by purchase order.

Purchase orders:
(Please include the full name of the licensing school/department.)

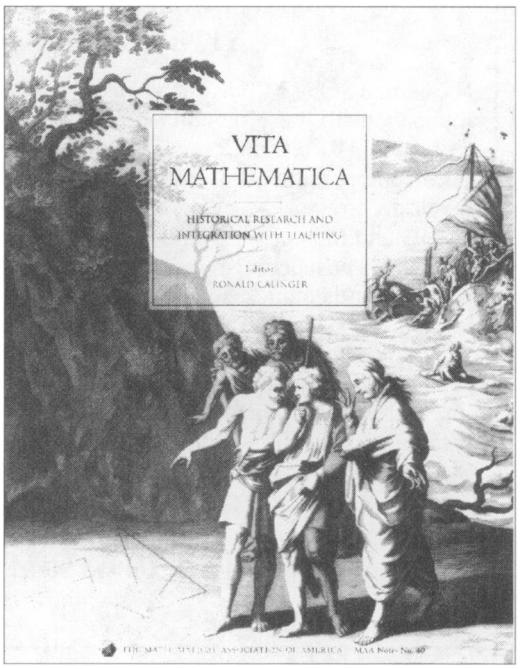
MathCalcEduc
1449 Covington Drive
Ann Arbor, MI 48103-5630

Federal EIN: 38-2740157
Telephone: 313 761 4666

Vita Mathematica

Historical Research and Integration with Teaching

Ronald Calinger, Editor



The use of the history of mathematics in the teaching of mathematics at all levels is an idea whose time has come. To use history in the teaching of undergraduate mathematics, the instructor must be familiar with the history as well as the mathematics. *Vita Mathematica* will enable college teachers to learn the relevant history of various topics in the undergraduate curriculum and help them incorporate this history in their teaching.

For example, should calculus be approached from a geometric or an algebraic point of view? The book shows us how two important eighteenth century mathematicians, Colin Maclaurin and Joseph-Louis Lagrange, understood the calculus from these different standpoints and how their legacy is still important in teaching calculus today. We also learn why Lagrange's algebraic approach dominated teaching in Germany in the nineteenth century. Some of the reasons for this are related to the appropriate foundations of the calculus, and so the book traces the ancient history of one of the possible foundations, the concept of indivisibles. Even though we generally do not use this concept formally today, many ideas for a heuristic approach to the calculus can be developed out of his study.

Vita Mathematica contains numerous other articles dealing with calculus, with algebra, combinatorics, graph theory, and geometry, as well as more general articles on teaching courses for prospective teachers.

This volume, then, demonstrates that the history of mathematics is no longer tangential to the mathematics curriculum, but in fact deserves a central role.

Catalog Code: NTE40
350 pp., Paperbound, 1996, ISBN 0-88385-097-4
List: $34.95 MAA Member: $29.00

ORDER FROM:
THE MATHEMATICAL ASSOCIATION OF AMERICA
P.O. Box 91112, Washington, DC 20090-1112
1-800-331-1622 (301) 617-7800 FAX (301) 206-9789

Membership Code:

____ ____ ____ ____ ____ ____

Name _____

Address _____

City _____

State _____ Zip _____

Qty.	Catalog Code	Price	Amount
_____	NTE40	_____	_____
_____	_____	_____	_____
		TOTAL	_____

Payment ☐ Check ☐ VISA ☐ MasterCard

Credit Card No. _____ Expires __/__

Signature _____

CONTENTS

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, NW
Washington, D.C. 20036