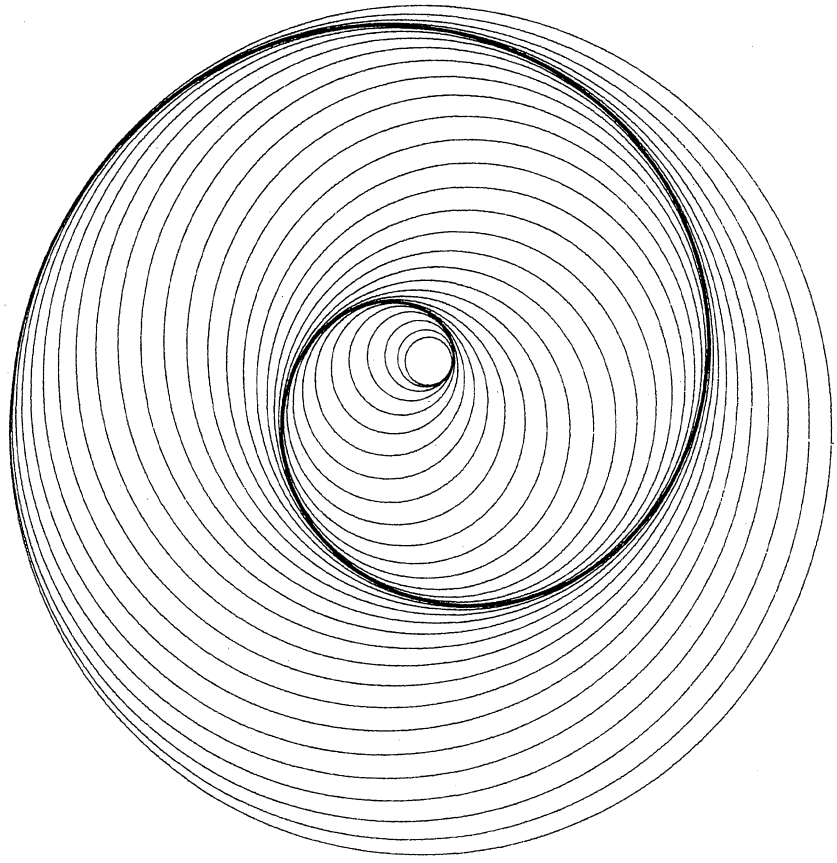
# MATHEMATICS MAGAZINE



Vol. 54, No. 2
March, 1981

FLIPPIN' COINS • INVERSE OF A MATRIX SUM
BILLIARD PATH POLYGONS • OSCULATING CIRCLES

MATHEMATICS MAGAZINE

THE MATHEMATICAL ASSOCIATION OF AMERICA

*Volume 54, Number 2, March 1981*
*(ISSN—0025–570X)*



## ARTICLES



## NOTES



## PROBLEMS



## REVIEWS



## NEWS AND LETTERS

## EDITORIAL POLICY

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

New manuscripts should be sent to: Doris Schattschneider, Editor, Mathematics Magazine, Moravian College, Bethlehem, Pennsylvania 18018. Manuscripts should be prepared in a style consistent with the format of *Mathematics Magazine*. They should be typewritten and double spaced on $8\frac{1}{2}$ by 11 paper. Authors should submit the original and one copy and keep one copy as protection against possible loss. Illustrations should be carefully prepared on separate sheets of paper in black ink, the original without lettering and two copies with lettering added.

Authors planning to submit manuscripts should read the full statement of editorial policy which appears in the News and Letters section of this *Magazine*, Vol. 54, pp. 44–45. Additional copies of the policy are available from the editor.

## AUTHORS

**Bill Newman** ("The Flippin' Coins Problem") was born on 6th May 1936 at Muttaburra, a small town in outback Australia. He received the B.Sc. (honours), B. Ed., and M.Sc. degrees from Queensland University. He received his Ph.D. from James Cook University of North Queensland in 1970 (the first Ph.D. awarded by that institution) for work on the conjugacy problem in one-relator groups. In 1961 he was appointed lecturer in mathematics at Townsville University College. In 1967 he was a visiting professor at Fairleigh Dickenson University, Teaneck, NJ, and during 1968 he was visiting professor at Pahlavi University, Shiraz, Iran. He has been a senior lecturer in mathematics at James Cook University since 1970. His chief interest is in coding and information theory. He was a visiting associate at the California Institute of Technology in 1980, and has worked with Dr. R. Miller at the Jet Propulsion Laboratory on coding problems associated with the Galileo project to send spacecraft to explore Jupiter and its moons.

## ILLUSTRATIONS

The cover design (which could be displayed as op-art) was computer drawn, the program written by **David Gungner**, and was executed on a Cal Comp. 936. With proper coloring, it could be the vortex of a storm; mathematically, it illustrates osculating circles of a spiral (see p. 77).

The illustrations for the poem *Laplace Transforms* are reproductions of drawings by **John Tenniel**, from an early edition of Lewis Carroll's *Through the Looking-Glass*.

Figure 1 in *The Flippin' Coins Problem* was provided by the editor.

All other illustrations were provided by the authors.

**BUSINESS INFORMATION.** Mathematics Magazine is published by the Mathematical Association of America at Washington, D.C., five times a year in January, March, May, September, and November. Ordinary subscriptions are $18 per year. Members of the Mathematical Association of America or of Mu Alpha Theta may subscribe at special reduced rates. Colleges and university mathematics departments may purchase bulk subscriptions (5 or more copies to a single address) for distribution to undergraduate students.

Subscription correspondence and notice of change of address should be sent to A. B. Willcox, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N. W., Washington, D.C. 20036. Back issues may be purchased, when in print, from P. and H. Bliss Co., Middletown, Connecticut 06457.

Advertising correspondence should be addressed to Raoul Hailpern, Mathematical Association of America, SUNY at Buffalo, Buffalo, New York 14214.



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

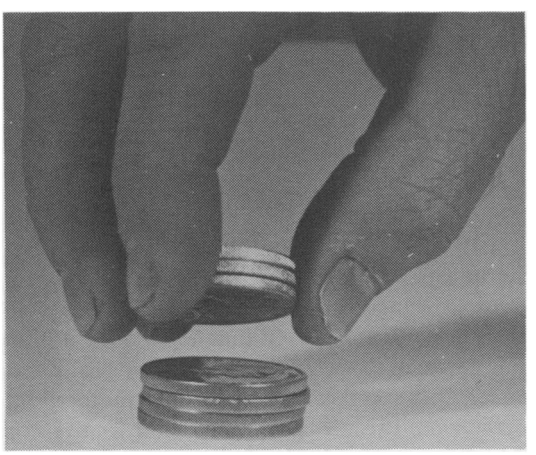
# The Flippin' Coins Problem

*Playing with loose change creates a problem; the solution is found by clever applications of number theory.*

**B. B. NEWMAN**
*James Cook University of North Queensland*
*Queensland, 4811 Australia*

A pile of $M$ coins is arranged so that each coin is heads up. We take the top coin, flip it over and put it back on top of the pile. We then take the top two coins, and holding them together turn them upside down and put them back on top of the pile. We continue, taking in turn the top three, four and so on until the whole pile is turned upside down. We then commence at the top again flipping the top stack of one, two, three coins and continue in this fashion until the pile is again all heads up.

PROBLEM: *How many flips are required*?



For $m = 4$, the pile of coins will undergo these changes (brackets indicate the stack of coins flipped at each stage):

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | H | T | T | T | T | H | H | T | T | H | T | H | | | |
| 2 | H | H | H | T | T | T | T | H | H | H | T | H | | | |
| 3 | H | H | H | H | H | H | H | T | T | T | T | H | | | |
| 4 | H | H | H | H | H | H | H | H | H | H | H | H | | | |

Thus, the number of flips required for a pile of four coins is 11. This problem was conceived by John Gilder and Iain Bride of the University of Manchester Institute of Science and Technology. In private correspondence with the author John Gilder wrote, "The history is rather thin. I conceived the problem while juggling with my small change when waiting for public transport." His original conception was slightly altered and put in its present form by his colleague Iain Bride. Working out particular cases for $M$ is an illuminating exercise in elementary computer programming and was included for that purpose by Graham Birtwistle and others in a book on the Simula language [1]. A table showing the number of flips for different $M$ is shown in TABLE 1. From an examination of this table, Birtwistle made the following conjectures.

CONJECTURE 1. *The number of flips for a pile of M coins is of the form Mk or Mk − 1.*

CONJECTURE 2. *This number is bounded by* $M^2$ $(M > 1)$.

CONJECTURE 3. *For* $2^n \leqslant M \leqslant 2^{n+1} - 1$, *the minimum number of flips occurs when* $M = 2^n$, *and is* $M(n+1) - 1$. *Also if* $M = 2^{n+1} - 1$, *the number is* $M(n+2)$.

The discussion which follows reveals the truth of all three conjectures.

| $M$ (Number in pile) | Number of flips | $M$ (Number in pile) | Number of flips |
|---|---|---|---|
| 1 | 2 | 17 | 204 |
| 2 | 3 | 18 | 323 |
| 3 | 9 | 19 | 228 |
| 4 | 11 | 20 | 199 |
| 5 | 24 | 21 | 146 |
| 6 | 35 | 22 | 264 |
| 7 | 28 | 23 | 529 |
| 8 | 31 | 24 | 504 |
| 9 | 80 | 25 | 200 |
| 10 | 60 | 26 | 675 |
| 11 | 121 | 27 | 540 |
| 12 | 119 | 28 | 251 |
| 13 | 116 | 29 | 840 |
| 14 | 195 | 30 | 899 |
| 15 | 75 | 31 | 186 |
| 16 | 79 | 32 | 191 |

TABLE 1

Associated to each coin in a pile is its *position* in the pile and its *state* (heads up or tails up). Each flip causes a permutation of the positions of the coins in the pile, and changes the state of the coins in the stack turned over. We shall find that knowledge of the permutation of the coins in the pile left after flipping through the *whole* pile is the key tool to confirming the conjectures. Superficially it would appear that a consideration of only the permutations at these particular stages of the flipping process would result in an incomplete analysis, since part of the pile may consist of a stack of heads at some intermediate stage of the flipping process. Although these permutations of the coins provide the key to confirming the conjectures, ultimately it is not the

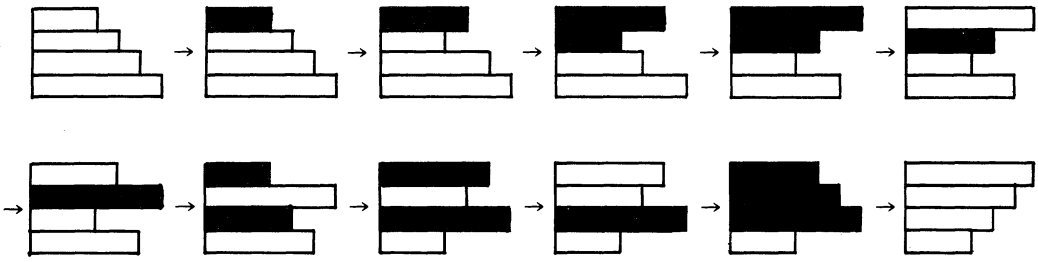


FIGURE 1. **A pile of four coins of different size shows changes in the position of each coin and the state of each coin (white = H, black = T) after each flip. After 11 flips, the pile is all heads, but the positions are in reverse order.**

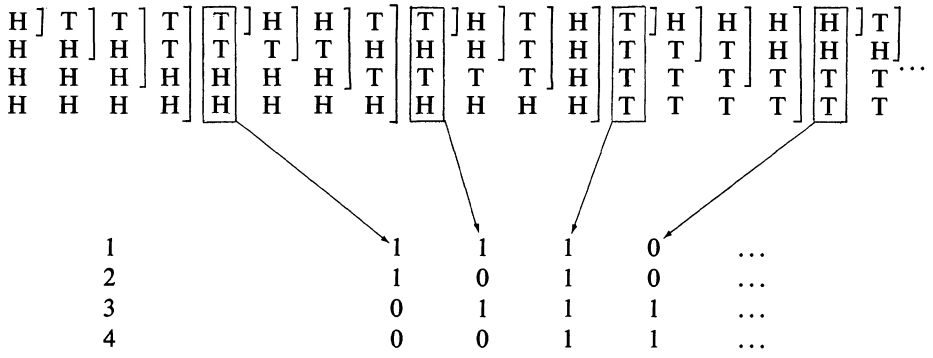
position of a coin in the pile that is of concern, but rather whether it has been flipped heads-up or tails-up (see FIGURE 1).

An interesting and useful characterization of the flip of the coins is the substance of Lemma 1. This Lemma alone may be used to confirm Conjecture 1 and, together with a little elementary group theory, Conjecture 2. We will first confirm Conjectures 1 and 2, using Lemma 1 without proof and then proceed deeper to the analysis of the permutations. We can then confirm Conjecture 3 and prove Lemma 1.

The permutations which emerge have particularly interesting properties and have arisen in several investigations extending as far back as 1773 [3]. What is new in this article is the relationship between the permutations and the flip of the coins. The techniques developed can be generalized from the two-fold state of the flip of a coin to multiple-fold states to solve such problems as: If each watch in a pile of $M$ watches loses $n$ hours each time it is flipped over, find the number of flips required before the watches show the same time.

## The Proof of Conjectures 1 and 2

If Conjecture 1 is to be correct, then a pile of heads can occur only after a move that flips over the whole pile of $M$ coins, or just prior to such a move. This suggests that we examine the state of the pile after each move that flips over the whole pile. Thus we are interested in the state of the pile just after the $M$th, $2M$th, $3M$th, etc., flip. To examine these states we extract the columns which result from such pile flips and place them together, writing a head as 0 and a tail as 1. For $m = 4$ we obtain




We will interpret the $n$th row which results from this extraction as a binary decimal $p_M(n)$. TABLE 2 shows these rows for $M = 4$.

$p_4(1) = .1110\ldots$
$p_4(2) = .1010\ldots$
$p_4(3) = .0111\ldots$
$p_4(4) = .0011\ldots$

TABLE 2

After plucking these numbers out of the pile flips, we see that the $p_M(n)$ must be recurring decimals, and so represent rational numbers. (After all, for each $M$, the pile is restored to all heads after some finite number of flips of the whole pile. Thus the binary decimals $p_M(n)$ must have repeating blocks of digits.) But it is the unexpected simplicity of the rational representation of these decimals that is the key to confirming the conjectures.

LEMMA 1.

$$p_M(n) = \frac{2M+2-2n}{2M+1}.$$

We prove this lemma in the last section of our article; for now we prefer to assume it is true and show its usefulness. The lemma provides a powerful tool for examining the structure of the pile, and not just for after a move flipping over the whole pile. For $M = 4$, Lemma 1 is illustrated in TABLE 3. (Recall that to obtain the rational representation of a binary decimal, find the sum of the corresponding geometric series. For example, $.\overline{111000} = \sum_{n=0}^{\infty}(2^{-1}+2^{-2}+2^{-3})2^{-6n}$.)

| $n$ | $p_M(n)$ | $\frac{2M+2-2n}{2M+1}$ |
|---|---|---|
| 1 | $.\overline{111\ 000}$ | 8/9 |
| 2 | $.\overline{101\ 010}$ | 6/9 |
| 3 | $.\overline{011\ 100}$ | 4/9 |
| 4 | $.\overline{001\ 110}$ | 2/9 |

TABLE 3

These rational numbers provide us with an easy representation of the digits occurring in each row of TABLE 2, but more important is a knowledge of the digits occurring in the columns. This is easily obtained, for if $p_M(n) \geq \frac{1}{2}$ then '1' will occur in the first position after the decimal point, and if $p_M(n) < \frac{1}{2}$ then '0' will occur in the first position. Thus in TABLE 3 corresponding to $8/9, 6/9, 4/9, 2/9$ there is the column of first binary digits $1, 1, 0, 0$. To examine the $k$th column of the decimals of TABLE 2 we may use a similar argument. We first shift the decimal point up to this position by multiplying the binary decimal by $2^{k-1}$, and then examine the decimal part. Again a decimal part greater than $\frac{1}{2}$ will indicate a 1, and less than $\frac{1}{2}$ will indicate a 0. But using the rational representation we see that shifting the decimal point and taking the fractional part is equivalent to multiplying the number $(2M+2-2n)/(2M+1)$ by $2^{k-1}$ and then reducing the numerator mod $2M+1$.

For $M = 4$ the fractions which will identify the $k$th digit of $p_M(n)$ will be those shown in TABLE 4.

| $n$ | $k=1$ | $k=2$ | $k=3$ | $k=4$ | $k=5$ | $k=6$ |
|---|---|---|---|---|---|---|
| 1 | 8/9 | 7/9 | 5/9 | 1/9 | 2/9 | 4/9 |
| 2 | 6/9 | 3/9 | 6/9 | 3/9 | 6/9 | 3/9 |
| 3 | 4/9 | 8/9 | 7/9 | 5/9 | 1/9 | 2/9 |
| 4 | 2/9 | 4/9 | 8/9 | 7/9 | 5/9 | 1/9 |

TABLE 4

The numerators of the fractions in the $k$th column form an arithmetic sequence mod $2M+1$ from top to bottom

$$M\times 2^k,(M-1)\times 2^k,\ldots,3\times 2^k,2\times 2^k,1\times 2^k.$$

Whenever the numerators mod $2M+1$ are less than $M+1$ there will be a 0 in the corresponding position for $p_M(n)$. In order for the $k$th column to be a column of zeros, we must have the sequence of numerators mod $2M+1$ all less than $M+1$. The only such sequence is

$$M,M-1,\ldots,3,2,1.$$

This implies $2^k\equiv 1 \bmod 2M+1$. In order to have the $k$th column all ones, we must have the sequence of numerators

$$M+1,M+2,\ldots,2M-2,2M-1,2M.$$

This implies $2^k\equiv 2M\equiv -1 \bmod 2M+1$.

If we interpret this discussion of binary digits in terms of the original pile of coins, we see that a pile of heads occurs immediately after flipping over the whole pile of $M$ coins $k$ times, when $2^k\equiv 1 \bmod 2M+1$, or immediately before flipping over the whole pile $k$ times, when $2^k\equiv -1 \bmod 2M+1$.

This discussion does not quite confirm Conjecture 1, for there still remains the possibility of a column of heads occurring more than one move before the next pile flip. This could only occur if flipping over the whole pile resulted in an arrangement consisting of at least two heads at the bottom, with the remaining coins alternating heads and tails. As a hypothetical case, suppose that for $M=8$, a pile flip resulted in the first column shown in TABLE 5. Then after five succeeding flips, the pile would be all heads, as shown. We will show that the state illustrated in column 1 of TABLE 5 cannot occur (for any $M$) after a pile flip.

| $M=8$ | | | | | | |
|---|---|---|---|---|---|---|
| T | H | T | H | T | H |
| H | H | T | H | T | H |
| T | T | T | H | T | H |
| H | H | H | H | T | H |
| T | T | T | T | T | H |
| H | H | H | H | H | H |
| H | H | H | H | H | H |
| H | H | H | H | H | H |

TABLE 5

Notice that if one tail only occurred after a pile flip, and it was the top coin, this would imply that the pile was in a state of all heads two moves earlier. A column arrangement after a pile flip such as shown in column 1 of TABLE 5, corresponds to an arithmetic sequence:

$$Ma,(M-1)a,\ldots,3a,2a,a \quad \bmod 2M+1.$$

Suppose heads occur in the bottom $r-1$ positions and a tail in the $r$th position from the bottom. Then mod $2M+1$

$$a,2a,\ldots,(r-1)a<M.$$

This however implies that

$$a,2a,\ldots,(r-1)a<M$$

without having to perform reduction mod $2M+1$. Since a tail occurs somewhere other than at the top, $r<M$. We will show that the terms $(r+1)a$ as well as $ra$ are mod $2M+1$ greater than $M$, and so prove a pattern of heads and tails alternating cannot occur. Since

$$ra=(r-1)a+a\leqslant M+M$$

then no reduction mod $2M+1$ is required. Similarly $(r+1)a=(r-1)a+2a\leqslant M+M$ and $(r+1)a$ requires no reduction mod $2M+1$. Clearly $(r+1)a>ra$ and so $(r+1)a>M$. This

shows that $\bmod 2M+1$, $(r+1)a$ is also greater than $M$. Thus we have shown that *a whole pile of heads will occur if and only if Mk flips are performed where* $2^k \equiv 1 \bmod 2M+1$, *or* $Mk-1$ *flips when* $2^k \equiv -1 \bmod 2M+1$, and so Conjecture 1 is proved.

It is interesting to look at the possible values of $M$ such that there exists an integer $k$ with $2^k \equiv -1 \bmod 2M+1$, for in this case the number of flips is not a multiple of $M$. If $k$ is even, then $-1$ is a quadratic residue of $2M+1$ and so is a quadratic residue of every prime dividing $2M+1$. If $k$ is odd, then $-2$ is a quadratic residue of $2M+1$ and so is a quadratic residue of every prime dividing $2M+1$. We now use the following two results from number theory. (See [2], p. 135, 139.)

LEMMA 2. *The number* $-1$ *is a quadratic residue of all primes of the form* $8n-1$ *or* $8n+5$ *and a quadratic nonresidue of all primes of the form* $8n+3$ *or* $8n+7$.

LEMMA 3. *The number* $-2$ *is a quadratic residue of all primes of the form* $8n+1$ *or* $8n+3$, *and a quadratic nonresidue of all primes of the form* $8n+5$ *or* $8n+7$.

These results imply the following: if $p$ is a prime dividing $2M+1$ of the form

(i) $8n+7$, then $2^k \equiv -1 \bmod 2M+1$ has no solutions,

(ii) $8n+5$, then $2^k \equiv -1 \bmod 2M+1$ has solutions only for $k$ even,

(iii) $8n+3$, then $2^k \equiv -1 \bmod 2M+1$ has solutions only for $k$ odd.

Thus if $2M+1$ is divisible by any prime of the form $8n+7$ or by the product of two primes of the form $8n+3$ and $8n+5$, then no $k$ exists with $2^k \equiv -1 \bmod 2M+1$. The product of two primes of the form $8n+3$ and $8n+5$ is a number of the form $8n+7$. Also $8n+7$ must have prime factors of the form (a) $8n+7$ or (b) $8n+3$ and $8n+5$. Hence we can combine these two cases: *If* $2M+1$ *is divisible by any number of the form* $8n+7$ *then* $2^k \not\equiv -1 \bmod 2M+1$.

For example, if $M$ is of one of the forms $4K-1, 7K-4, 12K-2$, then $2M+1$ is of the form $8K-1, 14K-7, 24K-3$, respectively, all of which are clearly divisible by a number of the form $8n+7$. For such values of $M$, the number of flips will be a multiple of $M$. These few forms account for all the cases in TABLE 1 where $M$ divides the number of flips except the case $M=25$. We can now prove Conjecture 2. To show that the number of flips for a pile of $M$ coins is at most $M^2$, it will suffice, in the light of our results above, to show that there exists an integer $k \leqslant M$ such that $2^k \equiv \pm 1 \bmod 2M+1$. To this end we introduce some elementary group theory.

Consider the positive integers less than $2M+1$ and relatively prime to $2M+1$. Taking as the group operation multiplication modulo $2M+1$, we obtain a group $G$. The order of $G$ must be an even integer $2n$, because if $x$ is relatively prime to $2M+1$, then so is the different integer $2M+1-x$, and the elements of the group can be paired off. Clearly $2n \leqslant 2M$. We are interested in the subgroup of $G$ generated by 2. There are two situations that can arise, illustrated by $M=4$, $M=7$.

$$M=4: \quad G=\{1,2,4,5,7,8\}, \quad \text{and 2 generates } G.$$

$$M=7: \quad G=\{1,2,4,7,8,11,13,14\}, \quad \text{and 2 generates a proper subgroup } \{2, 2^2=4, 2^3=8, 2^4=1\}.$$

If the powers of 2 generate the whole group $G$ as they do for the case $M=4$, then $2^{2n}-1 \equiv 0 \bmod 2M+1$ or $(2^n+1)(2^n-1) \equiv 0 \bmod 2M+1$. Since $2n$ is the order of $G$, $2^n \not\equiv 1$, hence $2^n \equiv -1 \bmod 2M+1$. Thus there exists an integer $k(=n) \leqslant M$ with the required properties.

If the powers of 2 generate a proper subgroup of $G$, then by Lagrange's theorem the order $k$ of this subgroup is a divisor of the order of $G$. Since $2^k \equiv 1 \bmod 2M+1$, and $k$ must satisfy $k \leqslant \frac{1}{2} \cdot 2n \leqslant M$, we have shown there exists an integer $k$ with the required properties.

## The Proofs of Conjecture 3 and Lemma 1

Until now we have been concerned only with coins being heads up or tails up. To confirm the

third conjecture, and to prove Lemma 1, we will examine the rearrangement of the coins which results as we proceed from one pile flip to another. Because the same process is repeated after each move which flips over the whole pile, the resultant rearrangements after succeeding pile flips will be obtained by taking powers of the permutation of the positions of the coins after the first pile flip. We can keep track of the movement of the coins as shown in TABLE 6.

| | | | | |
|---|---|---|---|---|
| 1 | 1 | 2 | 3 | 4 |
| 2 | 2 | 1 | 1 | 2 |
| 3 | 3 | 3 | 2 | 1 |
| 4 | 4 | 4 | 4 | 3 |

TABLE 6

After the first pile flip (the last column in TABLE 6), coin 1 is in position 3, coin 2 is in position 2, and so on where positions are counted down from the top of the pile. The permutation which represents the change of positions of the coins in the pile after the first pile flip may be written

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 4 & 1 \end{pmatrix}.$$

We will find it more convenient to use the inverse of such permutations. Thus

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 4 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 3 & 2 & 4 & 1 \\ 1 & 2 & 3 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 2 & 1 & 3 \end{pmatrix}.$$

Compare the second row of this last term with the last column of TABLE 6. We define $\phi_M$ to be the inverse of the permutation of the positions of the coins $1, 2, \ldots, M$ performed from one pile flip to the next. Thus the coin moving into the $n$th position from the top is that which was in position $n\phi_M$ at the outset. If one carries out the flips, one quickly sees the pattern in $\phi_M$. For example

$$\phi_8 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 8 & 6 & 4 & 2 & 1 & 3 & 5 & 7 \end{pmatrix}$$

$$\phi_9 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 9 & 7 & 5 & 3 & 1 & 2 & 4 & 6 & 8 \end{pmatrix}.$$

The pattern observed for these $\phi_M$ suggests the following lemma.

LEMMA 4.

$$n\phi_M = \begin{cases} M + 2 - 2n, & \text{if } n < \dfrac{M}{2} + 1, \\ 2n - M - 1, & \text{if } n \geqslant \dfrac{M}{2} + 1 \end{cases}.$$

We leave the proof of this lemma to the reader; it is proved by induction on $M$. We illustrate by finding $\phi_9$ given $\phi_8$. In evaluating $\phi_9$ the flips performed are exactly the same as those for $M = 8$, until the move involving the 9th coin. Thus we can represent the flips for $M = 9$ as shown below:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1 | | | | | | 8 | | 9 |
| 2 | | | | | | 6 | | 7 |
| 3 | · | | | | | 4 | | 5 |
| 4 | | · | | | | 2 | | 3 |
| 5 | | | · | | | 1 | | 1 |
| 6 | | | | | | 3 | | 2 |
| 7 | | | | | | 5 | | 4 |
| 8 | | | | | | 7 | | 6 |
| 9 | | | | | | 9 | | 8 |

The second last column is known from $\phi_8$, and the final column gives $\phi_9$.

| $M$ | Order of cycles of $\phi_M$ | $k$ (order of $\phi_M$) | $M_k$ | Number of flips for $M$ coins |
|---|---|---|---|---|
| 8 | 4,4 | 4 | 32 | 31 |
| 9 | 9 | 9 | 81 | 80 |
| 10 | 6,3,1 | 6 | 60 | 60 |
| 11 | 11 | 11 | 121 | 121 |
| 12 | 10,2 | 10 | 120 | 119 |
| 13 | 9,3,1 | 9 | 117 | 116 |
| 14 | 14 | 14 | 196 | 195 |
| 15 | 5,5,5 | 5 | 75 | 75 |

TABLE 7

The same argument may be used to find which coins are flipped over as a result of the moves from one pile flip to the next. This is the content of the next lemma.

LEMMA 5. *The coin moving into the nth position is flipped over if and only if* $n < M/2 + 1$.

If a pack of $2n$ cards is shuffled, as is not unusual, by placing the second card on the first, the third below these, the fourth above them, and so on, the permutation associated with shuffling the pack is precisely $\phi_{2n}$. The theory of this system of shuffling appeared in 1773 [3]. Rouse Ball [4] mentions eight papers dealing with this card shuffling problem.

Every permutation can be written as a product of disjoint cycles. For example

$$\phi_8 = (1\ 8\ 7\ 5)\ (2\ 6\ 3\ 4)$$

$$\phi_9 = (1\ 9\ 8\ 6\ 2\ 7\ 4\ 3\ 5).$$

The order of any permutation is the lcm of the orders of the disjoint cycles. The relevance of the order of $\phi_M$ to the problem in hand is shown in TABLE 7.

Notice that the order of $\phi_M$ appears to be the order of one of the cycles of $\phi_M$. This is proved in Lemma 6.

LEMMA 6. *The order of the permutation* $\phi_M$ *is the order of the cycle containing* $M$.

*Proof.* From Lemma 4, the permutation $\phi_M$ may be written $n\phi_M = (-1)^\epsilon(2n - M - 1) + \epsilon$, $\epsilon = 0$ or 1, which implies $2(n\phi_M) = (-1)^\epsilon 2(2n-1) - 2(-1)^\epsilon M + 2\epsilon$. From this we obtain $2(n\phi_M) - 1 = (-1)^\epsilon 2(2n-1) - (-1)^\epsilon(2M+1)$, which implies

$$2(n\phi_M) - 1 \equiv (-1)^\epsilon 2(2n-1) \bmod 2M + 1. \tag{1}$$

Define

$$\begin{aligned} x_0(n) &= 2n - 1 \\ x_1(n) &= 2(n\phi_M) - 1 \\ x_2(n) &= 2(n\phi_M{}^2) - 1 \\ &\vdots \\ x_k(n) &= 2(n\phi_M{}^k) - 1. \end{aligned}$$

Then for $\epsilon_i = 0$ or 1, the congrugence in (1) implies

$$\begin{aligned} x_1(n) &\equiv (-1)^{\epsilon_1} 2x_0(n) \mod 2M + 1 \\ x_2(n) &\equiv (-1)^{\epsilon_2} 2x_1(n) \mod 2M + 1 \\ &\vdots \\ x_k(n) &\equiv (-1)^{\epsilon_k} 2x_{k-1}(n) \mod 2M + 1. \end{aligned}$$

Multiplying together and cancelling we obtain

$$x_k(n) \equiv (-1)^{\Sigma \epsilon_i} 2^k x_0(n) \bmod 2M+1.$$

Thus if $2^k \equiv \pm 1$, then $x_k(n) \equiv \pm x_0(n) \bmod 2M+1$. The minus sign cannot occur, for if it did, then $x_k(n) = 2(n\phi_M{}^k) - 1 \equiv -(2n-1) \bmod 2M+1$, and so $n\phi_M{}^k \equiv 1-n$ since $2M+1$ is odd. But the left hand side is $\leqslant M$ and the right hand side is $> M$. Thus if $2^k \equiv \pm 1$, then $n\phi_M{}^k = n$. Also $1\phi_M{}^k = 1$ if and only if $2^k \equiv \pm 1$. This proves that if $k$ is the order of the cycle containing 1, then it is the order of the permutation $\phi_M$. Since the cycle containing 1 must also contain $M$, this proves the lemma.

It is interesting that the order of $\phi_M$ is the same $k$ obtained in the discussion of Conjecture 1. What this means is that a column of heads or a column of tails can occur only when all the coins have been returned to their original order, or that order reversed.

By considering the structure of the cycle containing $M$ we are able to obtain further information about the relationship between $M$ and the order of $\phi_M$. From Lemma 4, the cycle containing $M$ may be written ending with 1 and beginning

$$M, M-1, M-1-2, M-1-2-2^2, \ldots$$

and continuing in this fashion until an integer $< M/2+1$ is reached. Let the $l$th term of the cycle be the first integer $< M/2+1$. Then the $(l-1)$th term is $M-(1+2+\cdots+2^{l-3}) = M-2^{l-2}+1 \geqslant M/2+1$, and the $l$th term is $M-(1+2+\cdots+2^{l-2}) = M-2^{l-1}+1 < M/2+1$.

These two inequalities imply $2^{l-1} \leqslant M \leqslant 2^l - 1$. If $M = 2^{l-1}$, then the $l$th term is 1, so the order of $\phi_M$ is $l$, which is the minimum possible order for $M$ in the stated binary range, and so Conjecture 3 is confirmed. Note that other parts of Conjecture 3 have already been confirmed by the discussion of Conjecture 1.

Finally, let us prove Lemma 1. Let $n_k$ denote the $k$th digit in the binary decimal $p_M(n)$. The proof that for all $k$, $n_k$ equals the $k$th digit in the binary expansion of $X = (2M+2-2n)/(2M+1)$ will be by induction on $k$.

For $k=1$, Lemma 5 implies $n_k = 0$ if and only if $n \geqslant M/2+1$, or equivalently, $2M+2-2n \leqslant M$. This is the same condition for the first digit in the binary expansion of $X$ to be zero. Assume the result for $k$. Since the coin that moves to the $n$th position is $n\phi_M$ and is flipped over if and only if $n < M/2+1$, we have

$$n_{k+1} = \begin{cases} (n\phi_M)_k & \text{if } n \geqslant M/2+1, \\ 1-(n\phi_M)_k & \text{if } n < M/2+1. \end{cases}$$

If $n \geqslant M/2+1$, $(n\phi_M)_k = 0$ if and only if the following inequality is true, $\bmod 2M+1$: $2^{k-1}(2M+2-2(2n-M-1)) \leqslant M$, that is, $\bmod 2M+1$, $2^k(2M+2-2n) \leqslant M$, which is the necessary and sufficient condition for the $(k+1)$th digit in the binary expansion of $X$ to be zero.

If $n < M/2+1$, $1-(n\phi_M)_k = 1$ if and only if the following inequality is true $\bmod 2M+1$: $2^{k-1}(2M+2-2(M+2-2n)) \leqslant M$, that is, $\bmod 2M+1$, $2^k(2M+2-2n) > M$, which is the necessary and sufficient condition for the $(k+1)$th digit in the binary expansion of $X$ to be 1.

This proves that $n_{k+1}$ is the $(k+1)$th digit of the binary expansion of $X$, and the lemma follows by induction.

**References**


[1] Graham Birtwistle, Ole-Johan Dahl, Bjørn Myhrhaug, Kristen Nygaard, Simula Begin, Auerback, Philadelphia, PA, 1973.

[2] Trygve Nagell, Introduction to Number Theory, Chelsea, New York, 1964.

[3] Mémoires de l'Académie des Sciences, Paris, 1773.

[4] W. W. Rouse Ball, Mathematical Recreations and Essays, Macmillan, London, 1959, p. 310.

# NOTES

## Modified Farey Sequences and Continued Fractions

MAURICE SHRADER-FRECHETTE
*Defense Mapping Agency*
*600 Federal Place*
*Louisville, KY 40202*

The purpose of this note is to investigate the structure of certain lists of fractions which are similar to Farey sequences. We shall find that these modified Farey sequences are closely related to the notion of the "mass" of a continued fraction. Furthermore we shall see that the new entries of each modified sequence can be generated recursively from the new entries of the previous sequence.

The **Farey sequence of order** $n$ is the list of rationals in the interval [0, 1], arranged in ascending order, whose denominators are less than or equal to $n$. For example, the following are the sequences of orders 3 and 4.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| order 3: | 0/1 | | 1/3 | 1/2 | 2/3 | | 1/1 |
| order 4: | 0/1 | 1/4 | 1/3 | 1/2 | 2/3 | 3/4 | 1/1 |

An interesting and fairly well-known property of Farey sequences is that there is a recursive process for generating the fractions of order $n+1$ from those of order $n$. The first Farey sequence consists of $0/1$ and $1/1$; and the sequence of order $n+1$ can be obtained from the sequence of order $n$ by inserting between each successive pair $p/q$ and $p'/q'$ the fraction $(p+p')/(q+q')$, if and only if its denominator $q+q'$ equals $n+1$ [5]. Thus in the example above, the sequence of order 4 is obtained from the sequence of order 3 by inserting $1/4$ and $3/4$, but not $2/5$ and $3/5$, since 5 is too large a denominator.

A fraction such as $(p+p')/(q+q')$ which is formed by adding the numerators and denominators of two fractions is called their **mediant**. The recursive procedure for generating Farey fractions requires that some mediants be held back as each new list (after the third one) is generated. Thus $2/5$ and $3/5$ first appear in the fifth sequence instead of in the fourth.

In this note we shall investigate the effect of using all of the mediants in generating new lists. We shall call the sequences obtained in this manner **modified Farey sequences**: the sequence of order 1 is $0/1$, $1/1$; the sequence of order $n+1$ is generated by inserting the mediant between each consecutive pair in the sequence of order $n$, regardless of the size of the denominator. For example, the modified Farey sequence of order 4 includes the rationals $2/5$ and $3/5$, which are omitted from the Farey sequence of order 4. Because the mediant of two fractions lies between them, it follows that the fractions in each modified Farey sequence are all in the interval [0, 1] and are in ascending order.

Since the patterns which emerge from the modified Farey sequences are best expressed in continued fraction notation, we review this topic briefly. A finite, regular, simple continued fraction is an expression of the form

$$a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots \cfrac{1}{a_m}}},$$

where each $a_i$ $(i \geqslant 2)$ is a positive integer and $a_1$ is an integer; such an expression will be denoted by $(a_1, \ldots, a_m)$. The integers $a_1, \ldots, a_m$ are called its **terms**, and the successive rational numbers $(a_1)$, $(a_1, a_2)$, $(a_1, a_2, a_3) \ldots$ the **convergents**. We will denote the $n$th convergent $(a_1, \ldots, a_n)$ when reduced to lowest terms by $p_n/q_n$.

The following two properties (see [5]) will be useful for our purposes. If $p_m/q_m = (a_1, \ldots, a_m)$ and $a_m > 1$, then $p_m/q_m$ is also equal to $(a_1, \ldots, a_{m-1}, a_m - 1, 1)$ so every rational can be expressed as a continued fraction in exactly two ways. (To get one representation note that $p/q = 1/(a + b/p)$ where $ap + b = q$; a direct argument shows that if $(a_1, \ldots, a_n) = (b_1, \ldots, b_m)$ that either $a_i = b_i$ for all $i$ or $m = n + 1$ and $b_m = 1$, $b_{m-1} = a_m - 1$.) For example

$$\frac{30}{7} = 4 + \frac{1}{3 + \frac{1}{2}} = 4 + \frac{1}{3 + \frac{1}{1 + \frac{1}{1}}}.$$

Also the $p_k$ and $q_k$ of successive convergents may be expressed recursively: begin with $p_0 = q_{-1} = 1$ and $q_0 = p_{-1} = 0$, and then

$$p_k = a_k p_{k-1} + p_{k-2}$$
$$q_k = a_k q_{k-1} + q_{k-2}.$$

If $p_m/q_m = (a_1, \ldots, a_m)$, then the **mass** $m(p_m/q_m)$ of $p_m/q_m$ is defined to be $a_1 + \cdots + a_m$. This concept is well defined, even though every rational number has two continued fraction representations $(a_1, \ldots, a_{m-1}, 1)$ and $(a_1, \ldots, a_{m-1} + 1)$, for the sum of the terms is the same in each case.

In the interval $(0, 1)$ it turns out that the mass of any rational $x/y$ is the same as that of its reciprocal and also of its complement $1 - x/y$. To see this, note that the first term of any continued fraction in $(0, 1)$ must be 0. Since $(0, a_2, \ldots, a_m) = 0 + 1/(a_2, \ldots, a_m)$, the reciprocal of $(0, a_2, \ldots, a_m)$ is $(a_2, \ldots, a_m)$. Because the sum of terms is the same in each of these, $m(x/y) = m(y/x)$. For the mass of $1 - x/y$ first suppose that $a_2 > 1$; then $y/x = (a_2, \ldots, a_m) = 1 + (a_2 - 1, a_3, \ldots, a_m)$, and it follows that $(y - x)/x = (a_2 - 1, a_3, \ldots, a_m)$. Taking the reciprocal and adding 1 to both sides, we have $y/(y - x) = (1, a_2 - 1, a_3, \ldots, a_m)$. Taking the reciprocal again, we obtain the formula $1 - x/y = (0, 1, a_2 - 1, a_3, \ldots, a_m)$ hence $m(x/y) = m(1 - x/y)$. Finally if $a_2 = 1$, i.e., $x/y = (0, 1, a_3, \ldots, a_m)$, then $x/y$ is the complement of $(0, a_3 + 1, a_4, \ldots, a_m)$. So, again, $m(x/y) = m(1 - x/y)$.

If we make ordered lists of the rationals in the interval $[0, 1]$ having mass less than or equal to 4 (TABLE 1) we see that these lists are exactly the same as the first four modified Farey sequences! This is our central result: *The ordered list of rationals, lying in the interval* $[0, 1]$ *and having mass* $\leqslant n$, *is the same as the modified Farey sequence of order* $n$. We prove this Theorem by a series of lemmas.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| mass $\leqslant 1$: | $(0)$ | | | | | | | | $(0, 1)$ |
| | $= 0/1$ | | | | | | | | $= 1/1$ |
| mass $< 2$: | $(0)$ | | | | $(0, 2)$ | | | | $(0, 1)$ |
| | $= 0/1$ | | | | $= 1/2$ | | | | $= 1/1$ |
| mass $\leqslant 3$: | $(0)$ | | $(0, 3)$ | | $(0, 2)$ | | $(0, 1, 2)$ | | $(0, 1)$ |
| | $= 0/1$ | | $= 1/3$ | | $= 1/2$ | | $= 2/3$ | | $= 1/1$ |
| mass $\leqslant 4$: | $(0)$ | $(0, 4)$ | $(0, 3)$ | $(0, 2, 2)$ | $(0, 2)$ | $(0, 1, 1, 2)$ | $(0, 1, 2)$ | $(0, 1, 3)$ | $(0, 1)$ |
| | $= 0/1$ | $= 1/4$ | $= 1/3$ | $= 2/5$ | $= 1/2$ | $= 3/5$ | $= 2/3$ | $= 3/4$ | $= 1/1$ |

TABLE 1

LEMMA 1. *The mediant of* $p_{m-1}/q_{m-1} = (a_1, \ldots, a_{m-1})$ *and* $p_m/q_m = (a_1, \ldots, a_m)$ *is* $p/q = (a_1, \ldots, a_m, 1)$.

*Proof.* By the recursion relations for the convergents of a continued fraction, $p = 1 \cdot p_m + p_{m-1}$ and $q = 1 \cdot q_m + q_{m-1}$.

LEMMA 2. *In any modified Farey sequence, successive pairs of rationals have continued fraction representations either of the form* $(a_1, \ldots, a_{m-1})$, $(a_1, \ldots, a_{m-1}, a_m)$ *or the reverse* $(a_1, \ldots, a_{m-1}, a_m)$, $(a_1, \ldots, a_{m-1})$.

*Proof.* The entire sequence of order 1 is (0), (0, 1). Clearly this pair satisfies the lemma. Suppose that the property holds for the sequence of order $k$ and let $p/q$ and $p'/q'$ be any consecutive pair of rationals in the sequence of order $k+1$. Further let $p/q$ be the mediant of $p'/q'$ and $p''/q''$. By our inductive assumption, $p'/q'$ can be denoted $(a_1, \ldots, a_{m-1})$ and $p''/q''$ can be denoted $(a_1, \ldots, a_{m-1}, a_m)$ or vice versa. In either case the mediant $p/q$ has the continued fraction representation $(a_1, \ldots, a_{m-1}, a_m, 1)$, according to Lemma 1. If $p'/q' = (a_1, \ldots, a_{m-1})$, then we denote $p/q$ as $(a_1, \ldots, a_{m-1}, a_m + 1)$ and have our result. But if $p'/q' = (a_1, \ldots, a_{m-1}, a_m)$, then we denote $p/q$ as $(a_1, \ldots, a_{m-1}, a_m, 1)$ and also have the result.

LEMMA 3. *If $p/q$ is a new entry in the modified Farey sequence of order $n \geqslant 2$ (i.e., if $p/q$ is not found in any of the sequences of order $< n$), then $m(p/q) = n$.*

*Proof.* This is evidently true in the first few sequences, which we have written above. Suppose that it is true in the sequence of order $k$. Let $p/q$ be any new entry in the sequence of order $k+1$; let it be the mediant of $p'/q' = (a_1, \ldots, a_{m-1})$ and $p''/q'' = (a_1, \ldots, a_{m-1}, a_m)$. Since only one of the pairs $p'/q'$ and $p''/q''$ is new in the sequence of order $k$, then by the inductive assumption it has mass $k$ and the other has a smaller mass. Thus $m(p''/q'') = k$, since $p''/q''$ has one more term than $p'/q'$. But $p/q$ has all the terms of $p''/q''$ and an additional 1, so that $m(p/q) = m(p''/q'') + 1 = k + 1$.

We still cannot conclude that this sequence is the same as the ordered list of all rationals in the interval [0, 1] with mass $\leqslant n$, because we do not yet know whether any rationals with mass $n$ have been omitted. However, there are $2^{n-2}$ new entries in the sequence of order $n$, and the following lemma shows that this is also the number of rationals in the interval [0, 1] which have mass $n$.

LEMMA 4. *For each $n \geqslant 2$ there are $2^{n-2}$ rationals in the interval* [0, 1] *which have mass $n$.*

*Proof.* Continued fractions for the rationals with mass $n$ can be constructed as follows. Write down a sequence of $n$ 1's separated by $n-1$ commas. (For example, if $n$ is 6, write 1, 1, 1, 1, 1, 1.) For each of the first $n-2$ commas, either substitute a "+" or leave the comma alone. (In the example, write 1 + 1, 1, 1 + 1, 1.) Do not change the last comma. Write a "0" and a comma to the left of the resulting arrangement and enclose everything in parentheses. (Our example then becomes (0, 1 + 1, 1, 1 + 1, 1).) In this way, one can construct continued fractions (always ending in 1) for all the rationals having mass $n$ in the interval [0, 1]. Since there are $2^{n-2}$ binary decisions to be made, there are $2^{n-2}$ such rationals.

Our main result follows immediately from the previous lemmas, and we state it formally:

THEOREM 1. *The modified Farey sequence of order n is the same as the ascending list of rationals in the interval* [0, 1] *having mass $\leqslant n$.*

We next show that the modified sequences have an advantage over Farey fractions in generating rationals in that the *new* entries in the sequence of order $n$ can be generated recursively from the new entries of order $n-1$. Nothing of the sort is possible with Farey sequences. The first few lists of new entries (see TABLE 1) in the modified Farey sequences exhibit a pattern, which we formulate as Theorem 2.

THEOREM 2. *For $n \geqslant 0$, the ordered list of rationals with mass $n+1$ in the interval* [0, 1] *can be written from the ordered list of those with mass n by the following algorithm:*

1. *Add* 1 *to the second (first nonzero) entry in the continued fraction expression for each rational*

*of mass n. (For a rational x/y in reduced form, this yields x/(x+y).)*

2. *If n = 0 or 1, there is nothing more to do. Otherwise use the relation* $1-(0,a_2,a_3,\ldots,a_m)=(0,1,a_2-1,a_3,\ldots,a_m)$ *to extend the list by writing in reverse order the complements of the numbers produced in step* 1. *(For the rational x/(x+y), this yields y/(x+y).)*

*Proof of the theorem.* It is obvious that the steps are sufficient to generate the rows of mass 1 and 2. Suppose that the procedure is also sufficient to list the row of mass $k$. Since the application of step 1 to the row with mass $k$ requires the addition of a 1, the masses of the resulting rationals will all be $k+1$, and since the second step consists of finding complements, the resulting masses will also be $k+1$. By Lemma 4, there are $2^{k-2}$ rationals with mass $k$ in the interval (0, 1); hence the procedure produces twice as many, or $2^{k-1}$ continued fractions with mass $k+1$. This is exactly the number of rationals in the interval (0, 1) with mass $k+1$. However, we must make sure that we have not listed two different continued fractions for the same number. Once we have shown that the algorithm yields numbers arranged in strictly ascending order, it will follow that they are all different.

It is clear from TABLE 1 that the results for masses 0 through 4 are in ascending order. Suppose the two-step procedure lists the rationals of mass $k$ in strictly ascending order; let $(0,a_2,\ldots)$ and $(0,b_2,\ldots)$ be two rationals in this list. Some arithmetic manipulation shows that $(0,a_2,\ldots)<(0,b_2,\ldots)$ implies $(0,a_2+1,a_3,\ldots)<(0,b_2+1,b_3,\ldots)$. Thus the numbers in the first half of row $k+1$ are also in strictly ascending order and are all different. Because the numbers in the second half of this row are the complements of those in the first half *in reverse order*, they also increase in value from left to right and are all different. Finally the numbers in the second half of row $k+1$ are all larger than those in the first half because each second entry is 1, whereas in the first half it is $\geqslant 2$. Therefore this two-step procedure produces the entire ordered list of rationals in [0, 1] with mass $k+1$ from the ordered list of those with mass $k$.

Let us look at some examples of this procedure (see TABLE 1). According to the first step, in row 1, (0, 1) comes from (0); in row 2, (0, 2) comes from (0, 1); in row 4, (0, 4) comes from (0, 3) and (0, 2, 2) comes from (0, 1, 2). By step 2, since the new entries of the first half of row 4 are (0, 4) and (0, 2, 2), the formula is first applied to (0, 2, 2) to yield (0, 1, 1, 2); next it is applied to (0, 4) to yield (0, 1, 3). Further when these steps are used to generate the elements of mass 5 from those of mass 4, this is the result:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| mass 4: | 1/4 | 2/5 | 3/5 | 3/4 | | | | |
| mass 5: | 1/5 | 2/7 | 3/8 | 3/7 | 4/7 | 5/8 | 5/7 | 4/5. |

Note that because of part 2 of Theorem 2, in the ascending arrangement of fractions with mass $n$, the list of numerators in the second half of any row is symmetric.

We close with some open questions about the concepts of mass, modified Farey sequences and our recursive processes. For example, can the concept of mass be extended meaningfully to continued fractions which are not regular (i.e., those whose terms may be negative or zero)? Does the "alternating mass" $(a_1-a_2+a_3-\cdots)$ have any interesting properties? And can the recursive generation of fractions with mass $n$ be used in inductive proofs of significant theorems about the rational numbers?

**References**


[1] G. Chrystal, Algebra, Part II, 2nd ed., A&C Black, London, 1939, pp. 423–444.

[2] H. Davenport, The Higher Arithmetic, 2nd ed., Harper Torchbooks, New York, 1960, pp. 79–114.

[3] G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, 4th ed., Clarendon Press, Oxford, England, 1960.

[4] Charles G. Moore, An Introduction to Continued Fractions, National Council of Teachers of Mathematics, Washington, 1964, pp. 1–18.

[5] Ivan Niven and Herbert S. Zuckerman, An Introduction to the Theory of Numbers, 3rd ed., Wiley, New York, 1972, pp. 134–177.

[6] Oskar Perron, Die Lehre von den Kettenbrüchen, Chelsea, New York, 1950, pp. 1–37.

[7] Harry N. Wright, First Course in Theory of Numbers, Dover, New York, 1971, pp. 15–42.

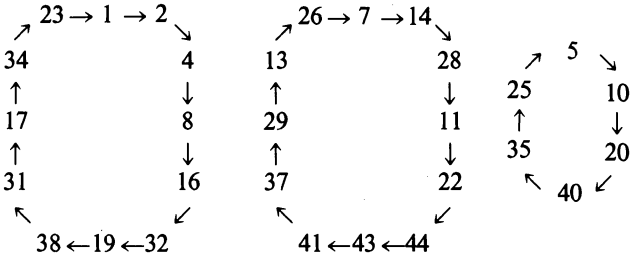
# Card Shuffling

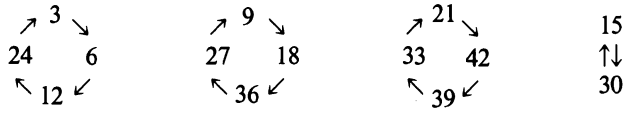
JOHN W. ROSENTHAL
*Ithaca College*
*Ithaca, NY 14850*

Our purpose here is to show that portions of a theory of card shuffling offer a nice application of number theory. It is especially suitable for an introduction to number theory in a general or liberal arts mathematics course, both since it helps motivate and illustrate various basic results in number theory and since students can easily be led to make many interesting conjectures. Indeed, most students are able to follow many proofs, and some are even able to provide proofs themselves. The latter part of this note is more advanced, and is more appropriate to an undergraduate number theory course.

Our topic is the theory of perfect shuffles. In a **perfect shuffle** of a deck with an even number $2n$ of cards, the top $n$ cards are perfectly interleaved with the bottom $n$. The result may be pictured schematically as in FIGURE 1. Define $s = 2n - 1$ to be the **size** of the deck, and for convenient reference, let's number the positions in the deck from 0 to $2n - 1$. Then the basic connection between perfect shuffles and number theory is that *a perfect shuffle sends the card in position k of a deck of size s to position* $2k \bmod s$ *for* $0 < k < s$. (Positions 0 and $s$ are uninteresting because cards in these positions are not moved by a perfect shuffle.)

Using this basic principle one can readily compute what repeated perfect shuffles do to various size decks. Eventually each card returns to its original position. We call the sequence of positions it goes through its **orbit**. For example, if $s = 45$ the orbits are




Examination of this and other cases leads to the conjecture that the orbit of 1 is always the longest orbit. We denote its length by $L(s)$. Using elementary algebra of arithmetic modulo $s$ one may show that if each position in the orbit of 1 is multiplied modulo $s$ by $k$, then one obtains the orbit of $k$, perhaps run through repeatedly. From this we obtain the **first orbit rule:** *the length of each orbit of a deck of size s is a factor of* $L(s)$. Hence, a deck of size $s$ is restored to its original order for the first time after $L(s)$ perfect shuffles.

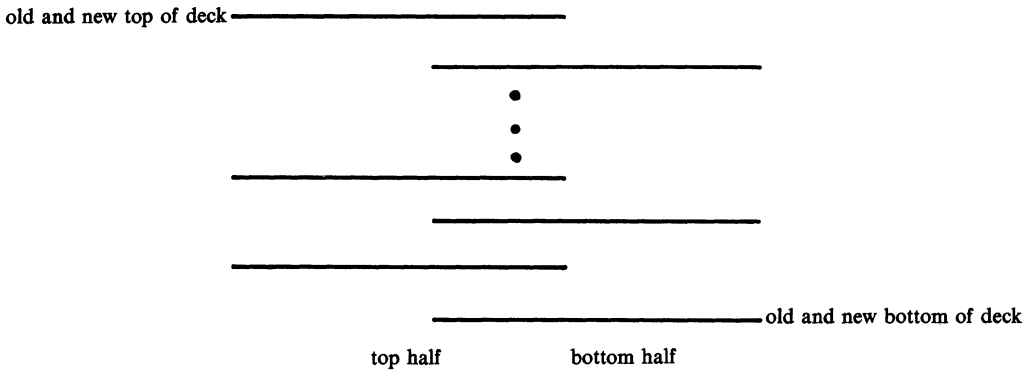
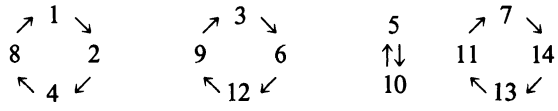



FIGURE 1

We will actually obtain a far more efficient means to compute $L(s)$ in many cases. To do this we first study the relation between the orbits of various size decks. Comparing the orbits of a deck of size $t$ with the orbits of decks of sizes which are factors of $t$ suggests the following conjecture: If $s$ is both a factor of $t$ and also a factor of $b$, a position in a deck of size $t$, then the orbit of $b$ can be obtained by multiplying by $s$ each position in the orbit of $b/s$ in the deck of size $t/s$. In particular we have the **induced orbit rule:** *the length of the orbit of b in the deck of size t is the same as the length of the orbit of b/s in the deck of size t/s.*

An excellent example is given by looking at the orbits of a deck of size 15:




If we multiply each entry by 3 then we obtain the orbits of those positions in a deck of size 45 which have 3 as a factor. (We leave to the reader actual proofs of these rules.)

As a consequence of the first orbit and induced orbit rules we see that $L(s)$ is a factor of $L(t)$ for each proper factor $s$ of $t$, and hence that $\operatorname{lcm}\{L(s) | s \text{ is a proper factor of } t\}$ *is a factor of* $L(t)$. In many cases $\operatorname{lcm}\{L(s) | s \text{ is a proper factor of } t\} = L(t)$. It is natural to ask when equality holds. Examination of various cases (say $t = 3$ to 99) leads to the observation that we have equality except when $t$ is a prime or a power of a prime. (A computer is very helpful in examining cases efficiently.) More careful examination of the cases of a power of a prime suggests that $L(p^2) = pL(p)$, and, more generally, that $L(p^{n+1}) = p^n L(p)$. The next simplest case is $L(p_1 p_2) = \operatorname{lcm}(L(p_1), L(p_2))$ where $p_1$ and $p_2$ are distinct primes. This may be proven by showing that the orbit of 1 in a deck of size $s = p_1 p_2$ may be obtained by adding together modulo $s$ the entries in repeated copies of the orbits of $p_1 x$ mod $s$ and $p_2 y$ mod $s$ where $x$ and $y$ are integers such that $1 = p_1 x + p_2 y$. The same proof shows if $s_1$ and $s_2$ are relatively prime, then $L(s_1 s_2) = \operatorname{lcm}(L(s_1), L(s_2))$. An inductive consequence of this is $L(s) = \operatorname{lcm}(L(p_1^{n_1}), \ldots, L(p_k^{n_k}))$ where $p_1^{n_1} \cdots p_k^{n_k}$ is the prime factorization of $s$.

Using $L(s) = \operatorname{lcm}(L(p_1^{n_1}), \ldots, L(p_k^{n_k}))$ one can readily show, for example, that since $999{,}999 = 3^3 \cdot 7 \cdot 11 \cdot 13 \cdot 37$, $L(999{,}999) = 180$. In other words, 180 perfect shuffles restore a million card deck to its original order. The entire computation should take less than 5 minutes! On the other hand, the conjecture that $L(p^2) = pL(p)$ is an excellent example of a statement which, although true for small primes, is not true in general. Use of a computer reveals that $L(p^2) = pL(p)$ for all primes $p < 1093$, but that $L(1093^2) = L(1093) = 364$.

So far we have discussed applications of number theory to the study of perfect shuffles. One can also apply the theory of perfect shuffles to prove results in number theory. For example, one

can use perfect shuffles to prove the following case of Fermat's theorem: $2^{\phi(s)} \equiv 1 \bmod s$ *for* $s$ *odd* (where $\phi$ is the classical Euler $\phi$ function). The proof comes from a detailed examination of the orbits of the deck of size $s$, especially of the orbits of positions which are relatively prime to $s$. Examination of various cases of the orbits of positions which are relatively prime to $s$ suggests that these orbits all have the same length. Suppose $k$ is relatively prime to $s$. Then there is an $x$ such that $kx \equiv 1 \bmod s$. One can obtain the orbit of 1 by multiplying by $x$ modulo $s$ each entry in the orbit of $k$. As a result, $L(s)$ is a factor of $\phi(s)$. Since $2^{L(s)} \equiv 1 \bmod s$, $2^{\phi(s)} \equiv 1 \bmod s$.

In fact, one can modify the above proof to derive the general version of Fermat's theorem: *If* $m$ *and* $s$ *are relatively prime, then* $m^{\phi(s)} \equiv 1 \bmod s$. To derive this, we define for relatively prime $m$ and $s$ an **m-handed perfect shuffle** of a deck of size $s$ to be the rearrangement of the deck in which the card in position $n$ is sent to position $mn \bmod s$. (A moment's thought will reveal the aptness of this name.) One now just needs to observe that all the arguments suggested to prove various general results about 2-handed perfect shuffles, in fact, can be routinely modified to prove analogous results about $m$-handed perfect shuffles.

Another application to number theory concerns primitive roots of unity. This arises from the further study of $L(p^n)$ for $p$ prime. In a naive effort to prove $L(p^2) = pL(p)$ we may observe that by definition $2^{L(p)} \equiv 1 \bmod p$ and hence $2^{L(p)} \equiv (ap + 1) \bmod p^2$ for some $a$ where $0 \leqslant a < p$. If $a \neq 0$, then $2^{kL(p)} \equiv (ap + 1)^k \bmod p^2 \equiv (kap + 1) \bmod p^2$ as all other factors in the binomial expansion of $(ap + 1)^k$ have a factor of $p^2$. Thus, $2^{kL(p)} \equiv 1 \bmod p^2$ if and only if $ka \equiv 0 \bmod p$, or equivalently if and only if $k \equiv 0 \bmod p$. Thus we may conclude that $L(p^2) = pL(p)$. On the other hand if $a = 0$, $L(p^2) = L(p)$. Thus, $L(p^2) = L(p)$ or $pL(p)$. More generally one may prove similarly that $L(p^{n+1}) = L(p^n)$ or $pL(p^n)$.

A similar but more refined argument yields: if $L(p^{n+k}) = p^n L(p)$, then $L(p^k) = L(p)$. For $k = 2$ the argument goes as follows. Say $2^{L(p)} = a + cp^2$. We must show $a \equiv 1 \bmod p^2$. We first observe that $a^{p^n} \equiv 1 \bmod p^{n+2}$ since by hypothesis $2^{p^n L(p)} \equiv 1 \bmod p^{n+2}$ and $2^{p^n L(p)} \equiv (a + cp^2)^{p^n} \equiv a^{p^n} \bmod p^{n+2}$ as all other terms in the binomial expansion of $(a + cp^2)^{p^n}$ have a factor of $p^{n+2}$. This implies that $a^{p^n} \equiv 1 \bmod p$, and hence, by Fermat's theorem, $a \equiv 1 \bmod p$. So let $a = 1 + bp$. It remains to show $b \equiv 0 \bmod p$. We first observe that $a^{p^n} \equiv (1 + bp)^{p^n} \equiv (1 + bp^{n+1}) \bmod p^{n+2}$, as careful examination reveals that all other terms of the binomial expansion of $(1 + bp)^{p^n}$ have factors of $p^{n+2}$. (This careful examination consists of looking at how many factors of $p$ occur in $\binom{p^n}{k}$, especially for $k$ a multiple of $p$. For more details, see, e.g., volume 1, chapter 9 of [1].) Since $a^{p^n} \equiv 1 \bmod p^{n+2}$, $bp^{n+1} \equiv 0 \bmod p^{n+2}$, and hence $b \equiv 0 \bmod p$.

In the case where $k > 2$ one proves that $a \equiv 1 \bmod p^i$ for $1 \leqslant i \leqslant k$ by an inductive argument in which the inductive step is similar to the conclusion of the above proof.

As a consequence of our observations about $L(p^{n+1})$ and $L(p^{n+k})$, if $L(p^2) = pL(p)$, then $L(p^{n+1}) = p^n L(p)$ for all $n$. This is true not only for 2-handed perfect shuffles but also for $m$-handed perfect shuffles, provided $p \neq 2$. (For $p = 2$, the binomial coefficients in some binomial expansions don't provide enough factors of $p$.)

The relation of the above to primitive roots of unity is that if $L(p^n) = \phi(p^n)$ for an $m$-handed perfect shuffle, then $m$ is a primitive root of unity modulo $p^n$. Hence for $p \neq 2$, if $m$ is a primitive root of unity modulo $p^2$, then it is a primitive root of unity modulo $p^n$ for every $n$. This is a classical result of Arndt to which Dickson [1] gives special attention.

One further reason to use perfect shuffles in an introduction to number theory is that they give rise to famous unsolved problems in mathematics. First, finding $L(p^n)$ for various primes $p$ is closely related to finding the prime factorization of $2^m - 1$ for various integers $m$, and hence to Mersenne primes and perfect numbers. Further, it is striking that $L(p) = \phi(p) = p - 1$ (i.e., 2 is a primitive root of unity modulo $p$) for many small primes. It is natural to ask whether this happens for infinitely many primes, for 2-handed and more generally for $m$-handed perfect shuffles (for $m$ not a square). Artin conjectured that it does. Hooley [2] has shown that Artin's conjecture follows from generalizations of the Riemann Hypothesis.

**References**

[1] Leonard E. Dickson, History of the Theory of Numbers, vol. 1, Chelsea, New York, 1952.
[2] Christopher Hooley, On Artin's conjecture, J. Reine Angew. Math., 225 (1967) 209–220.

# On the Inverse of the Sum of Matrices

KENNETH S. MILLER
*Riverside Research Institute*
*80 West End Avenue*
*New York, NY 10023*

If $\mathbf{G}$ and $\mathbf{H}$ are arbitrary nonsingular square matrices of the same dimension, then the inverse of their product $\mathbf{GH}$ is well known to be $\mathbf{H}^{-1}\mathbf{G}^{-1}$. If $\mathbf{G}$ and $\mathbf{G}+\mathbf{H}$ are nonsingular, then the problem of finding a simple expression for the inverse of $\mathbf{G}+\mathbf{H}$, say in terms of $\mathbf{G}^{-1}$, is more difficult. This is the problem we wish to address. In the theorem below we shall establish a recursive form for $(\mathbf{G}+\mathbf{H})^{-1}$.

The key result in proving this theorem is a fundamental Lemma which states that if $\mathbf{H}$ has rank one, then

$$(\mathbf{G}+\mathbf{H})^{-1}=\mathbf{G}^{-1}-\frac{1}{1+g}\mathbf{G}^{-1}\mathbf{H}\mathbf{G}^{-1} \tag{1}$$

where $g=\operatorname{tr}\mathbf{H}\mathbf{G}^{-1}$. For example, if

$$\mathbf{A}=\begin{bmatrix} a & 1 & 1 \\ 1 & a & 1 \\ 1 & 1 & a \end{bmatrix}, \qquad a\neq 1,-2$$

is a $3\times 3$ nonsingular matrix, then we may write $\mathbf{A}=\mathbf{G}+\mathbf{H}$ where $\mathbf{G}$ is $(a-1)$ times the identity matrix $\mathbf{I}$ and $\mathbf{H}$ is a $3\times 3$ matrix with ones everywhere (and hence of rank one). Since $\operatorname{tr}\mathbf{H}\mathbf{G}^{-1}=3/(a-1)$, equation (1) yields

$$\begin{aligned}\mathbf{A}^{-1}&=\frac{1}{a-1}\left[\mathbf{I}-\frac{1}{a+2}\mathbf{H}\right]\\ &=\frac{1}{(a-1)(a+2)}\begin{bmatrix} a+1 & -1 & -1 \\ -1 & a+1 & -1 \\ -1 & -1 & a+1 \end{bmatrix}.\end{aligned}$$

That the above formula indeed represents the inverse of $\mathbf{A}$ may be verified by a direct calculation.

Our program will be as follows. First we shall prove (1). Then we shall show that any matrix may be decomposed into the sum of matrices of rank one. This result will enable us to iteratively apply the Lemma to obtain our main theorem. We also shall consider some ancillary results, including some simple generalizations to the inverse of sums of Kronecker products of matrices.

It is appropriate to begin our analysis by considering matrices of rank one. Suppose, then, that $\mathbf{E}$ is a square matrix of rank one. Then all except possibly one eigenvalue of $\mathbf{E}$ is zero. Since the sum of the eigenvalues of $\mathbf{E}$ is the trace of $\mathbf{E}$, we see that the remaining eigenvalue is $\operatorname{tr}\mathbf{E}$. Matrices of rank one may be constructed by beginning with two nonzero column vectors $\mathbf{u}$ and $\mathbf{v}$ of the same dimension. Since $\mathbf{u}$ and $\mathbf{v}$ have rank one, the matrix $\mathbf{E}=\mathbf{u}\mathbf{v}'$ has rank one. (We shall use primes to indicate transposes of vectors and matrices. In particular, $\mathbf{v}'$ is a row vector.) Conversely, if $\mathbf{E}$ has rank one, then there exist nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ such that $\mathbf{E}=\mathbf{u}\mathbf{v}'$ [5, page 92]. We give some examples to indicate the usefulness of this representation.

EXAMPLE 1. (Norm of a matrix) If $\mathbf{A}$ is any square matrix, then the **norm** $|\mathbf{A}|$ of $\mathbf{A}$ is the square root of the largest eigenvalue of $\mathbf{A}'\mathbf{A}$. Now suppose $\mathbf{E}$ is a square matrix of rank one, and we write $\mathbf{E} = \mathbf{u}\mathbf{v}'$ where $\mathbf{u}$ and $\mathbf{v}$ are nonzero vectors. Then $\mathbf{E}'\mathbf{E} = \mathbf{v}\mathbf{u}'\mathbf{u}\mathbf{v}' = |\mathbf{u}|^2\mathbf{v}\mathbf{v}'$ and the only nonzero eigenvalue of $\mathbf{E}'\mathbf{E}$ is $|\mathbf{u}|^2 \operatorname{tr} \mathbf{v}\mathbf{v}' = |\mathbf{u}|^2|\mathbf{v}|^2$. Hence

$$|\mathbf{E}| = |\mathbf{u}||\mathbf{v}|. \tag{2}$$

As a corollary we deduce from the Schwarz inequality $|\mathbf{u}'\mathbf{v}| \leqslant |\mathbf{u}||\mathbf{v}|$ that

$$|\operatorname{tr} \mathbf{E}| \leqslant |\mathbf{E}|. \tag{3}$$

This result will be used later.

EXAMPLE 2. (Reproducing property) Let $\mathbf{A}$ be a square matrix and $\mathbf{E}$ a square matrix of rank one. Then if we write $\mathbf{E} = \mathbf{u}\mathbf{v}'$ where $\mathbf{u}$ and $\mathbf{v}$ are nonzero vectors, we have the interesting and useful reproducing property

$$\mathbf{E}\mathbf{A}\mathbf{E} = \mathbf{u}\mathbf{v}'\mathbf{A}\mathbf{u}\mathbf{v}' = \beta\mathbf{u}\mathbf{v}' = \beta\mathbf{E} \tag{4}$$

where $\beta$ is the bilinear form $\mathbf{v}'\mathbf{A}\mathbf{u}$. We also may write

$$\beta = \operatorname{tr} \mathbf{v}'\mathbf{A}\mathbf{u} = \operatorname{tr} \mathbf{u}\mathbf{v}'\mathbf{A} = \operatorname{tr} \mathbf{E}\mathbf{A}.$$

Thus (4) may be written as

$$\mathbf{E}\mathbf{A}\mathbf{E} = (\operatorname{tr} \mathbf{E}\mathbf{A})\mathbf{E} \tag{5}$$

and no explicit mention of the vectors $\mathbf{u}$ and $\mathbf{v}$ appears. As a corollary we see that

$$(\mathbf{E}\mathbf{A})^n = \beta^{n-1}\mathbf{E}\mathbf{A} \quad \text{and} \quad (\mathbf{A}\mathbf{E})^n = \beta^{n-1}\mathbf{A}\mathbf{E} \tag{6}$$

for all positive integers $n$.

Equation (4) is the key formula in proving the fundamental Lemma mentioned at the beginning of this paper. Suppose, then, that $\mathbf{G}$ and $\mathbf{G} + \mathbf{E}$ are nonsingular matrices where $\mathbf{E}$ has rank one. We may write $\mathbf{G} + \mathbf{E} = (\mathbf{I} + \mathbf{E}\mathbf{G}^{-1})\mathbf{G}$, and since $\mathbf{G}$ is nonsingular, the matrix $\mathbf{E}\mathbf{G}^{-1}$ has rank one. Hence $1 + \operatorname{tr} \mathbf{E}\mathbf{G}^{-1}$ is an eigenvalue of $\mathbf{I} + \mathbf{E}\mathbf{G}^{-1}$, the remaining eigenvalues all being one. Thus we see that $\mathbf{G} + \mathbf{E}$ is nonsingular if and only if $\operatorname{tr} \mathbf{E}\mathbf{G}^{-1} \neq -1$.

We search for an inverse of $\mathbf{G} + \mathbf{E}$ in the form $\mathbf{G}^{-1} - \nu\mathbf{G}^{-1}\mathbf{E}\mathbf{G}^{-1}$ where $\nu$ is a scalar. (A possible motivation to justify this choice stems from the following argument. Let $f(\zeta)$ be continuous on $[0, x]$ and differentiable on $(0, x)$. Let $x < a$. Then by the Law of the Mean, $f(x) = f(0) + f'(\lambda a)x$ where $0 < \lambda < 1$. Now let $f(\zeta) = (\zeta + a)^{-1}$. Then we may write

$$(x + a)^{-1} = a^{-1} - \nu a^{-1} x a^{-1}$$

where $\nu = (1 + \lambda)^{-2}$. Identify $x$ with $\mathbf{E}$ and $a$ with $\mathbf{G}$.)

If $\mathbf{G}^{-1} - \nu\mathbf{G}^{-1}\mathbf{E}\mathbf{G}^{-1}$ is to be the inverse of $\mathbf{G} + \mathbf{E}$, their product

$$(\mathbf{G} + \mathbf{E})(\mathbf{G}^{-1} - \nu\mathbf{G}^{-1}\mathbf{E}\mathbf{G}^{-1}) = \mathbf{I} - \nu\mathbf{E}\mathbf{G}^{-1} + \mathbf{E}\mathbf{G}^{-1} - \nu\mathbf{E}\mathbf{G}^{-1}\mathbf{E}\mathbf{G}^{-1}$$

must be the identity matrix. This in turn implies that

$$\nu\mathbf{E}\mathbf{G}^{-1} - \mathbf{E}\mathbf{G}^{-1} + \nu(\mathbf{E}\mathbf{G}^{-1})^2 = \mathbf{0}. \tag{7}$$

By (6), $(\mathbf{E}\mathbf{G}^{-1})^2 = g\mathbf{E}\mathbf{G}^{-1}$ where $g = \operatorname{tr} \mathbf{E}\mathbf{G}^{-1}$ and (7) may be written in the form

$$(\nu - 1 + \nu g)\mathbf{E}\mathbf{G}^{-1} = \mathbf{0}. \tag{8}$$

A sufficient condition for (8) to be valid is to have $\nu - 1 + \nu g = 0$ or

$$\nu = \frac{1}{1 + g}.$$

(Since $\mathbf{G} + \mathbf{E}$ is nonsingular, we have seen that $1 + g \neq 0$.) Thus we have proved:

LEMMA. *Let* $\mathbf{G}$ *and* $\mathbf{G} + \mathbf{E}$ *be nonsingular matrices where* $\mathbf{E}$ *is a matrix of rank one. Let* $g = \operatorname{tr} \mathbf{E}\mathbf{G}^{-1}$. *Then* $g \neq -1$ *and*

$$\mathbf{G}^{-1} - \frac{1}{1 + g}\mathbf{G}^{-1}\mathbf{E}\mathbf{G}^{-1}$$

*is the inverse of* $\mathbf{G}+\mathbf{E}$.

The above equation is essentially the Sherman-Morrison formula (see [1, page 161]).

Before continuing to the general case of finding the inverse of $\mathbf{G}+\mathbf{H}$ where $\mathbf{H}$ is not necessarily of rank one, let us show the relation of this Lemma to the Neumann series expansion of a matrix.

EXAMPLE 3. (Neumann series) If $\mathbf{P}$ is a square matrix and $|\mathbf{P}|<1$, then $(\mathbf{I}-\mathbf{P})^{-1}$ has the Neumann series expansion

$$(\mathbf{I}-\mathbf{P})^{-1}=\mathbf{I}+\mathbf{P}+\mathbf{P}^2+\cdots+\mathbf{P}^n+\cdots \tag{9}$$

(see [5, page 186]). Now let us suppose that $\mathbf{P}$ also has rank one. Then by (6), $\mathbf{P}^n=\alpha^{n-1}\mathbf{P}$, $n=1,2,\ldots$ where $\alpha=\operatorname{tr}\mathbf{P}$. Thus we may write (9) as

$$(\mathbf{I}-\mathbf{P})^{-1}=\mathbf{I}+(1+\alpha+\cdots+\alpha^{n-1}+\cdots)\mathbf{P}. \tag{10}$$

The series on the right-hand side of (10) is a geometric series with ratio $\alpha$, and by (3) $|\alpha|=|\operatorname{tr}\mathbf{P}|\leqslant|\mathbf{P}|<1$. Hence we may sum the series to obtain

$$(\mathbf{I}-\mathbf{P})^{-1}=\mathbf{I}+\frac{1}{1-\alpha}\mathbf{P}.$$

This formula is just a special case of the Lemma. One may consider the above identity for $(\mathbf{I}-\mathbf{P})^{-1}$ to be an "analytic continuation" of (9).

Now let us return to our main problem. We wish to find the inverse of $\mathbf{G}+\mathbf{H}$ where $\mathbf{G}$ and $\mathbf{G}+\mathbf{H}$ are nonsingular. Our basic argument in the Theorem below is to decompose $\mathbf{H}$ into the sum of matrices of rank one and iteratively apply the Lemma. It is known that if $\mathbf{H}$ has positive rank $r$, then we may write $\mathbf{H}$ in the form

$$\mathbf{H}=\mathbf{E}_1+\mathbf{E}_2+\cdots+\mathbf{E}_r \tag{11}$$

where each $\mathbf{E}_k$, $1\leqslant k\leqslant r$, has rank one [5, page 93]. (This decomposition is not unique.) Thus we may write

$$\mathbf{G}+\mathbf{H}=\mathbf{G}+\mathbf{E}_1+\cdots+\mathbf{E}_r.$$

If we are to recursively apply the Lemma, we need $\mathbf{G}+\mathbf{E}_1+\cdots+\mathbf{E}_k$ to be nonsingular for all $k$. This will not necessarily be true for an arbitrary decomposition (11) of $\mathbf{H}$. However, we can show that there *does* exist a decomposition (11) such that each of the "partial sums" $\mathbf{C}_{k+1}=\mathbf{G}+\mathbf{E}_1+\cdots+\mathbf{E}_k$ is nonsingular for $k=1,\ldots,r$.

To prove this contention we first note that since $\mathbf{G}+\mathbf{H}=(\mathbf{I}+\mathbf{H}\mathbf{G}^{-1})\mathbf{G}$ is nonsingular, no eigenvalue of $\mathbf{H}\mathbf{G}^{-1}$ can be $-1$ (as we observed in the proof of the Lemma.) Now let $\mathbf{Q}$ be a nonsingular matrix such that $\mathbf{J}=\mathbf{Q}(\mathbf{H}\mathbf{G}^{-1})\mathbf{Q}^{-1}$ is the Jordan normal form of $\mathbf{H}\mathbf{G}^{-1}$, and let

$$\mathbf{J}=\mathbf{F}_1+\mathbf{F}_2+\cdots+\mathbf{F}_r$$

where the $k$th row of $\mathbf{F}_k$ is the $k$th row of $\mathbf{J}$, and the remaining rows of $\mathbf{F}_k$ are all zero. Then every $\mathbf{F}_j$ has rank one and $\mathbf{I}+\mathbf{F}_1+\cdots+\mathbf{F}_k$ is nonsingular for $k=1,\ldots,r$. Thus

$$\left[\mathbf{Q}^{-1}(\mathbf{I}+\mathbf{F}_1+\cdots+\mathbf{F}_k)\mathbf{Q}\right]\mathbf{G}=\mathbf{G}+\mathbf{Q}^{-1}\mathbf{F}_1\mathbf{Q}\mathbf{G}+\cdots+\mathbf{Q}^{-1}\mathbf{F}_k\mathbf{Q}\mathbf{G}$$

is also nonsingular for $k=1,\ldots,r$. Since each $\mathbf{Q}^{-1}\mathbf{F}_j\mathbf{Q}\mathbf{G}$ has rank one, let $\mathbf{E}_j=\mathbf{Q}^{-1}\mathbf{F}_j\mathbf{Q}\mathbf{G}$. Then $\mathbf{C}_{k+1}=\mathbf{G}+\mathbf{E}_1+\cdots+\mathbf{E}_k$ is nonsingular for $k=1,\ldots,r$ and $\mathbf{C}_{r+1}=\mathbf{G}+\mathbf{H}$ where $\mathbf{H}=\mathbf{E}_1+\mathbf{E}_2+\cdots+\mathbf{E}_r$.

We are now in a position to give a precise statement of our main result.

THEOREM. *Let* $\mathbf{G}$ *and* $\mathbf{G}+\mathbf{H}$ *be nonsingular matrices and let* $\mathbf{H}$ *have positive rank* $r$. *Let* $\mathbf{H}=\mathbf{E}_1+\mathbf{E}_2+\cdots+\mathbf{E}_r$ *where each* $\mathbf{E}_k$ *has rank one and* $\mathbf{C}_{k+1}=\mathbf{G}+\mathbf{E}_1+\cdots+\mathbf{E}_k$ *is nonsingular for* $k=1,\ldots,r$. *Then if* $\mathbf{C}_1=\mathbf{G}$,

$$\mathbf{C}_{k+1}^{-1}=\mathbf{C}_k^{-1}-\nu_k\mathbf{C}_k^{-1}\mathbf{E}_k\mathbf{C}_k^{-1},\quad k=1,\ldots,r$$

*where*

$$\nu_k = \frac{1}{1 + \operatorname{tr} \mathbf{C}_k^{-1}\mathbf{E}_k}.$$

*In particular*

$$(\mathbf{G} + \mathbf{H})^{-1} = \mathbf{C}_r^{-1} - \nu_r \mathbf{C}_r^{-1}\mathbf{E}_r\mathbf{C}_r^{-1}.$$

To prove this result we first write $\mathbf{C}_2 = \mathbf{C}_1 + \mathbf{E}_1 = \mathbf{G} + \mathbf{E}_1$ and recall that $\mathbf{G}$ and $\mathbf{C}_2$ are nonsingular. Then by the Lemma,

$$\mathbf{C}_2^{-1} = \mathbf{G}^{-1} - \nu_1 \mathbf{G}^{-1}\mathbf{E}_1\mathbf{G}^{-1} \tag{12}$$

and we have calculated $\mathbf{C}_2^{-1}$ in terms of $\mathbf{G}^{-1}$. Now $\mathbf{C}_3 = \mathbf{G} + \mathbf{E}_1 + \mathbf{E}_2 = \mathbf{C}_2 + \mathbf{E}_2$. Hence since $\mathbf{C}_2$ and $\mathbf{C}_3$ are nonsingular, we again may invoke the Lemma to write $\mathbf{C}_3^{-1}$ in terms of $\mathbf{C}_2^{-1}$, viz.:

$$\mathbf{C}_3^{-1} = \mathbf{C}_2^{-1} - \nu_2 \mathbf{C}_2^{-1}\mathbf{E}_2\mathbf{C}_2^{-1}.$$

But $\mathbf{C}_2^{-1}$ is known from (12). If we continue this process $r$ times (where $r$ is the rank of $\mathbf{H}$) we obtain

$$\mathbf{C}_{r+1}^{-1} = \mathbf{C}_r^{-1} - \nu_r \mathbf{C}_r^{-1}\mathbf{E}_r\mathbf{C}_r^{-1}.$$

But $\mathbf{C}_{r+1} = \mathbf{G} + \mathbf{H}$, and thus our Theorem is proved.

As an illustration let us consider the problem of finding the inverse of $\mathbf{I} + \mathbf{H}$ where $\mathbf{I}$ is the identity matrix and $\mathbf{H}$ has rank two. Applying the above theorem with two (the rank of $\mathbf{H}$) iterations, some algebra will show that

$$(\mathbf{I} + \mathbf{H})^{-1} = \mathbf{I} - \frac{1}{a+b}(a\mathbf{H} - \mathbf{H}^2) \tag{13}$$

where $a = 1 + \operatorname{tr}\mathbf{H}$ and $2b = (\operatorname{tr}\mathbf{H})^2 - \operatorname{tr}\mathbf{H}^2$. Equation (13) also is valid if $\mathbf{H}$ has rank one. For in this case $\mathbf{H}^2 = (\operatorname{tr}\mathbf{H})\mathbf{H}$ by (5) and hence $\operatorname{tr}\mathbf{H}^2 = (\operatorname{tr}\mathbf{H})^2$. Thus $b = 0$ and (13) becomes

$$(\mathbf{I} + \mathbf{H})^{-1} = \mathbf{I} - \frac{1}{a}[a\mathbf{H} - (\operatorname{tr}\mathbf{H})\mathbf{H}] = \mathbf{I} - \frac{1}{a}\mathbf{H}$$

since $\operatorname{tr}\mathbf{H} = a - 1$. But this is just the Lemma with $\mathbf{G} = \mathbf{I}$ and $\mathbf{E} = \mathbf{H}$.

We also may exploit the above theorem to find the determinant of $\mathbf{G} + \mathbf{H}$ in terms of the $\nu_k$.

EXAMPLE 4. (Determinant of $\mathbf{G} + \mathbf{H}$) From the Theorem we see that $\mathbf{C}_{k+1} = \mathbf{C}_k + \mathbf{E}_k = \mathbf{C}_k(\mathbf{I} + \mathbf{C}_k^{-1}\mathbf{E}_k)$ and hence $\det \mathbf{C}_{k+1} = (\det \mathbf{C}_k)\det(\mathbf{I} + \mathbf{C}_k^{-1}\mathbf{E}_k)$. But $\mathbf{C}_k^{-1}\mathbf{E}_k$ has rank one. Thus $\det(\mathbf{I} + \mathbf{C}_k^{-1}\mathbf{E}_k) = 1 + \operatorname{tr}\mathbf{C}_k^{-1}\mathbf{E}_k = \nu_k^{-1}$. Hence $\det \mathbf{C}_{k+1} = \nu_k^{-1}(\det \mathbf{C}_k)$, $k = 1, \ldots, r$ and inductively

$$\det \mathbf{C}_{r+1} = \frac{1}{\nu_1\nu_2 \cdots \nu_r} \det \mathbf{C}_1.$$

But $\mathbf{C}_{r+1} = \mathbf{G} + \mathbf{H}$ and $\mathbf{C}_1 = \mathbf{G}$. Hence $(\nu_1\nu_2 \cdots \nu_r)\det(\mathbf{G} + \mathbf{H}) = \det \mathbf{G}$.

In many practical problems, for example, in the theory of Kalman filtering (see [4], [6], [8]), one wishes to find the inverse of $\mathbf{G} + \mathbf{H}$ for various matrices $\mathbf{H}$ when $\mathbf{G}^{-1}$ is known. The usefulness of the above formulas in such cases is readily apparent. Frequently in such problems $\mathbf{G}$ is positive definite and $\mathbf{H}$ is diagonal and nonnegative definite. In such cases $\mathbf{G} + \mathbf{H}$ is always nonsingular and the decomposition of $\mathbf{H}$ into matrices of rank one is trivial. It is also to be noted that the recursive form is especially convenient for computer utilization. The techniques also may be applied to the problem of finding the inverse of $\mathbf{G}$ itself. For example, for any square matrix $\mathbf{G}$ we may choose a matrix $\mathbf{G}_1$ and write $\mathbf{G} = \mathbf{G}_1 + (\mathbf{G} - \mathbf{G}_1)$. If $\mathbf{G}_1$ is chosen to be nonsingular, then we may determine $\mathbf{G}^{-1}$ in terms of $\mathbf{G}_1^{-1}$. In particular $\mathbf{G}_1$ may be chosen to be the identity matrix.

Before continuing we consider a specific application of the above Theorem. Let

$$\Lambda = \begin{bmatrix} 1 & \lambda & \lambda^2 \\ \lambda & 1 & \lambda \\ \lambda^2 & \lambda & 1 \end{bmatrix}, \quad 0 < \lambda < 1$$

be a $3 \times 3$ Markoffian matrix. We shall find $\Lambda^{-1}$. We first write $\Lambda = \mathbf{I} + \mathbf{E}_1 + \mathbf{E}_2 + \mathbf{E}_3$ where $\mathbf{I}$ is the identity matrix and

$$\mathbf{E}_1 = \begin{bmatrix} 0 & 0 & 0 \\ \lambda & 0 & 0 \\ \lambda^2 & 0 & 0 \end{bmatrix}, \quad \mathbf{E}_2 = \begin{bmatrix} 0 & \lambda & 0 \\ 0 & 0 & 0 \\ 0 & \lambda & 0 \end{bmatrix}, \quad \mathbf{E}_3 = \begin{bmatrix} 0 & 0 & \lambda^2 \\ 0 & 0 & \lambda \\ 0 & 0 & 0 \end{bmatrix}.$$

Then $\mathbf{C}_1 = \mathbf{I}$ and the formulas for $\nu_k$ and $\mathbf{C}_{k+1}^{-1}$ yield $\nu_1 = 1$ and $\mathbf{C}_2^{-1} = \mathbf{I} - \mathbf{E}_1$. Continuing the computation,

$$\nu_2 = \frac{1}{1 + \operatorname{tr} \mathbf{C}_2^{-1}\mathbf{E}_2} = \frac{1}{1-\lambda^2}$$

and

$$\begin{aligned} \mathbf{C}_3^{-1} &= \mathbf{C}_2^{-1} - \nu_2 \mathbf{C}_2^{-1}\mathbf{E}_2\mathbf{C}_2^{-1} \\ &= (\mathbf{I} - \mathbf{E}_1) - \frac{1}{1-\lambda^2}(\mathbf{I} - \mathbf{E}_1)\mathbf{E}_2(\mathbf{I} - \mathbf{E}_1) \\ &= \begin{bmatrix} \dfrac{1}{1-\lambda^2} & -\dfrac{\lambda}{1-\lambda^2} & 0 \\ -\dfrac{\lambda}{1-\lambda^2} & \dfrac{1}{1-\lambda^2} & 0 \\ 0 & -\lambda & 1 \end{bmatrix}. \end{aligned}$$

Finally,

$$\nu_3 = \frac{1}{1 + \operatorname{tr} \mathbf{C}_3^{-1}\mathbf{E}_3} = \frac{1}{1-\lambda^2}$$

and

$$\begin{aligned} \Lambda^{-1} &= \mathbf{C}_3^{-1} - \nu_3 \mathbf{C}_3^{-1}\mathbf{E}_3\mathbf{C}_3^{-1} \\ &= \mathbf{C}_3^{-1} - \frac{1}{1-\lambda^2}\begin{bmatrix} 0 & 0 & 0 \\ 0 & -\lambda^2 & \lambda \\ 0 & \lambda^3 & -\lambda^2 \end{bmatrix} \\ &= \frac{1}{1-\lambda^2}\begin{bmatrix} 1 & -\lambda & 0 \\ -\lambda & 1+\lambda^2 & -\lambda \\ 0 & -\lambda & 1 \end{bmatrix}. \end{aligned}$$

Let us now consider some applications involving the Kronecker product of matrices. Suppose $\mathbf{A}$ is an $M \times M$ matrix and $\mathbf{G} = \|g_{mn}\|$ an $N \times N$ matrix. Then the $MN \times MN$ dimensional partitioned matrix

$$\begin{bmatrix} g_{11}\mathbf{A} & g_{12}\mathbf{A} & \cdots & g_{1N}\mathbf{A} \\ g_{21}\mathbf{A} & g_{22}\mathbf{A} & \cdots & g_{2N}\mathbf{A} \\ \cdots & \cdots & \cdots & \cdots \\ g_{N1}\mathbf{A} & g_{N2}\mathbf{A} & \cdots & g_{NN}\mathbf{A} \end{bmatrix} \tag{14}$$

is called the **Kronecker product** or **direct product** of $\mathbf{A}$ and $\mathbf{G}$ and is frequently written as $\mathbf{A} \otimes \mathbf{G}$ [7, page 81]. The algebra of direct products is interesting. For example, if $\mathbf{A}$ and $\mathbf{G}$ are nonsingular, then

$$(\mathbf{A} \otimes \mathbf{G})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{G}^{-1}. \tag{15}$$

(Note that the order of $\mathbf{A}$ and $\mathbf{G}$ is not reversed.) However, the only nontrivial property we really need in our analysis is the formula

$$(\mathbf{A} \otimes \mathbf{G})(\mathbf{B} \otimes \mathbf{H}) = \mathbf{AB} \otimes \mathbf{GH} \tag{16}$$

where $\mathbf{A}$ and $\mathbf{B}$ are $M \times M$ matrices and $\mathbf{G}$ and $\mathbf{H}$ are $N \times N$ matrices.

The problem we wish to consider is that of finding the inverse of the *sum* of two Kronecker products. We begin by considering the matrix

$$\mathbf{W} = \mathbf{A} \otimes \mathbf{G} + \mathbf{B} \otimes \mathbf{E} \tag{17}$$

where $\mathbf{E}$ is an $N \times N$ matrix of rank one, and $\mathbf{A}$, $\mathbf{G}$ and $\mathbf{W}$ are nonsingular. Our previous analyses suggest that we search for an inverse in the form

$$\mathbf{W}^{-1} = \mathbf{A}^{-1} \otimes \mathbf{G}^{-1} - \mathbf{T} \otimes \mathbf{G}^{-1}\mathbf{E}\mathbf{G}^{-1} \tag{18}$$

where $\mathbf{T}$ is a matrix to be determined. If we multiply (17) and (18) together and use (15) and (16) we obtain

$$\mathbf{W}\mathbf{W}^{-1} = \mathbf{I} \otimes \mathbf{I} - \mathbf{A}\mathbf{T} \otimes \mathbf{E}\mathbf{G}^{-1} + \mathbf{B}\mathbf{A}^{-1} \otimes \mathbf{E}\mathbf{G}^{-1} - \mathbf{B}\mathbf{T} \otimes \mathbf{E}\mathbf{G}^{-1}\mathbf{E}\mathbf{G}^{-1}.$$

But by our reproducing property (6), $\mathbf{E}\mathbf{G}^{-1}\mathbf{E}\mathbf{G}^{-1} = g\mathbf{E}\mathbf{G}^{-1}$ where $g = \operatorname{tr}\mathbf{E}\mathbf{G}^{-1}$. Hence if $\mathbf{W}\mathbf{W}^{-1}$ is to be the identity matrix, then $(-\mathbf{A}\mathbf{T} + \mathbf{B}\mathbf{A}^{-1} - g\mathbf{B}\mathbf{T}) \otimes \mathbf{E}\mathbf{G}^{-1}$ must be the zero matrix. Thus we choose $\mathbf{T}$ to be $\mathbf{T} = (\mathbf{A} + g\mathbf{B})^{-1}\mathbf{B}\mathbf{A}^{-1}$.

Certain special cases are of interest.

EXAMPLE 5. ($\mathbf{G}$ the identity matrix) In this case (17) may be written

$$\mathbf{W} = \mathbf{A} \otimes \mathbf{I} + \mathbf{B} \otimes \mathbf{E} \tag{19}$$

and $\mathbf{T} = [\mathbf{A} + (\operatorname{tr}\mathbf{E})\mathbf{B}]^{-1}\mathbf{B}\mathbf{A}^{-1}$. Thus

$$\mathbf{W}^{-1} = \mathbf{A}^{-1} \otimes \mathbf{I} - [\mathbf{A} + (\operatorname{tr}\mathbf{E})\mathbf{B}]^{-1}\mathbf{B}\mathbf{A}^{-1} \otimes \mathbf{E}. \tag{20}$$

We also observe that $\det\mathbf{W} = (\det\mathbf{A})^{N-1}\det[\mathbf{A} + (\operatorname{tr}\mathbf{E})\mathbf{B}]$; and if $\lambda_1, \ldots, \lambda_M$ are the eigenvalues of $\mathbf{A}$, and $\mu_1, \ldots, \mu_M$ the eigenvalues of $\mathbf{A} + (\operatorname{tr}\mathbf{E})\mathbf{B}$, then each $\lambda_m$, $1 \leqslant m \leqslant M$, is an eigenvalue of multiplicity $N - 1$ of $\mathbf{W}$ and the $\mu_m$, $1 \leqslant m \leqslant M$, are simple eigenvalues of $\mathbf{W}$.

EXAMPLE 6. ($N = 1$) If $N = 1$ in (19), then $\mathbf{W} = \mathbf{A} + e\mathbf{B}$ where $e = \operatorname{tr}\mathbf{E} \neq 0$. Equation (20) then yields the well-known identity $(\mathbf{A} + e\mathbf{B})^{-1} = \mathbf{A}^{-1} - e(\mathbf{A} + e\mathbf{B})^{-1}\mathbf{B}\mathbf{A}^{-1}$. (Compare this with the result given by the fundamental Lemma.)

Now, returning to (17), suppose $\mathbf{B} = \mathbf{F}$ is also a matrix of rank one. Then $\mathbf{W} = \mathbf{A} \otimes \mathbf{G} + \mathbf{F} \otimes \mathbf{E}$ and by (18) its inverse is

$$\mathbf{W}^{-1} = \mathbf{A}^{-1} \otimes \mathbf{G}^{-1} - (\mathbf{A} + g\mathbf{F})^{-1}\mathbf{F}\mathbf{A}^{-1} \otimes \mathbf{G}^{-1}\mathbf{E}\mathbf{G}^{-1}. \tag{21}$$

We notice now that since $\mathbf{F}$ also is of rank one, the fundamental Lemma may be applied to $\mathbf{A} + g\mathbf{F}$, viz.:

$$(\mathbf{A} + g\mathbf{F})^{-1} = \mathbf{A}^{-1} - \frac{g}{1 + ga}\mathbf{A}^{-1}\mathbf{F}\mathbf{A}^{-1}$$

where $a = \operatorname{tr}\mathbf{F}\mathbf{A}^{-1}$. Substituting this result in (21) and simplifying yields the elegant formula:

$$(\mathbf{A} \otimes \mathbf{G} + \mathbf{F} \otimes \mathbf{E})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{G}^{-1} - \frac{1}{1 + ga}\mathbf{A}^{-1}\mathbf{F}\mathbf{A}^{-1} \otimes \mathbf{G}^{-1}\mathbf{E}\mathbf{G}^{-1}.$$

**References**


[1] G. Dahlquist and Å. Björck, Numerical Methods, Prentice-Hall, Englewood Cliffs, NJ, 1974.

[2] P. S. Dwyer and F. V. Waugh, On errors in matrix inversion, J. Amer. Statist. Assoc., 48 (1953) 289–319.

[3] D. K. Faddeev and V. N. Faddeeva, Computational Methods of Linear Algebra, Freeman, San Francisco, 1963.

[4] A. Gelb, Applied Optimal Estimation, M.I.T. Press, Cambridge, 1974.

[5] P. R. Halmos, Finite-Dimensional Vector Spaces, Van Nostrand, Princeton, 1958.

[6] D. M. Leskiw and K. S. Miller, A comparison of some Kalman estimators, IEEE Trans., IT-25 (1979) 491–495.

[7] C. C. Mac Duffee, The Theory of Matrices, Chelsea, New York, 1946.

[8] K. S. Miller, Vector Stochastic Processes, Krieger, Huntington, NY, 1980.

# A Billiard Path Characterization of Regular Polygons

DUANE W. DETEMPLE
JACK M. ROBERTSON
*Washington State University*
*Pullman, WA 99164*

It is not unexpected that geometric figures with special characteristics will also have special billiard path properties and, in fact, billiard paths may be used to characterize classes of figures. Robert Sine and Vladislav Kreĭnovič have given a billiard characterization of curves of constant diameter [2].

The intuitive meaning of **billiard path** is made precise as follows. Let $K$ be a convex body (a closed bounded convex set with nonempty interior) in $\mathbb{R}^n$, with a piecewise smooth boundary $\partial K$. The billiard ball is a point in $K$ which moves with constant velocity in $K$ until striking a point $p \in \partial K$. If $p$ is a regular point, the billiard ball bounces in the direction determined by reflection off the unique supporting hyperplane at $p$. For our purposes no rule of reflection need be formulated if $p$ is a singular point of $\partial K$. The orbit generated by a billiard ball is a billiard path. A billiard path is **periodic** if the orbit is a closed polygon with finitely many vertices at regular points of $\partial K$.

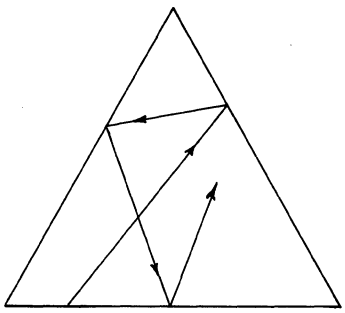


FIGURE 1

FIGURE 1 illustrates a billiard path in an equilateral triangle. We observe that if the billiard ball starts at a midpoint of one side and is directed toward the midpoint of an adjacent side, then the billiard path retraces itself after three bounces and is an equilateral triangle. Indeed this property is true for *any* regular polygon of $n$ sides: if the path begins at the midpoint of any side and is directed toward the midpoint of an adjacent side, then the billiard path is retraced periodically (after $n$ bounces) and the path traces another regular $n$-gon. The purpose of this note is to demonstrate that the regularity of plane polygons is characterized by this property.

THEOREM. *A closed convex polygon $P$ in the plane is regular if and only if $P$ contains a periodic billiard path $P'$ similar to $P$.*

The proof will rely on the following simple lemma.

LEMMA. *Let $x_1, \ldots, x_n$ be $n$ real numbers and define $y_i = \frac{1}{2}(x_{i-1} + x_i)$, $i = 1, \ldots, n$ (indices mod $n$). Then $y_1, \ldots, y_n$ is a permutation of $x_1, \ldots, x_n$ if and only if $x_1 = x_2 = \cdots = x_n$.*

*Proof of Lemma.* The if part is obvious.

Let $x = \min\{x_1, \ldots, x_n\}$. If $x_{i-1} > x$ or $x_i > x$, then $y_i = \frac{1}{2}(x_{i-1} + x_i) > x$. Thus the number of $y_j$ equal to $x$ is at least one less than the number of $x_j$ equal to $x$. This implies that the only possibility for $y_1, \ldots, y_n$ to be a permutation of $x_1, \ldots, x_n$ is for all the $x_j$'s to be equal.

*Proof of Theorem.* If $P$ is regular, the polygon joining successive midpoints of the sides of $P$ produces the required $P'$.

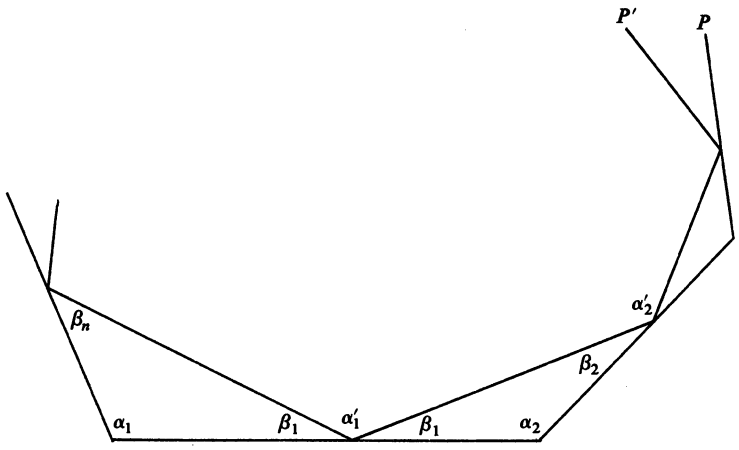



FIGURE 2

Now suppose $P'$ is periodic and similar to $P$. Then $P'$ must have its vertices on successive sides of $P$. Let $\alpha_1, \ldots, \alpha_n$ be the angles of $P$ in counterclockwise cyclic order. Let $\alpha_1', \ldots, \alpha_n'$ be the angles of $P'$ in counterclockwise cyclic order, where $\alpha_i'$ is located between $\alpha_i$ and $\alpha_{i+1}$. Since $P'$ is a billiard path, it makes equal angles $\beta_i$ with the side of $P$ as it traces out (i.e., bounces to produce) angle $\alpha_i'$ (FIGURE 2). Since $P$ and $P'$ are similar, there is a permutation $\sigma$ of $1, \ldots, n$ such that $\alpha_i = \alpha_{\sigma(i)}'$ and, for some $k$, $\sigma(i) \equiv i + k \bmod n$, $i = 1, \ldots, n$ or $\sigma(i) \equiv -i + k \bmod n$, $i = 1, \ldots, n$. Hence the following two systems of simultaneous equations result (see FIGURE 2):

$$\alpha_i + \beta_i + \beta_{i-1} = \pi, \; i = 1, \ldots, n \tag{1}$$

$$\alpha_i' + 2\beta_i = \pi, \; i = 1, \ldots, n \tag{2}$$

with $\alpha_1 + \cdots + \alpha_n = \alpha_1' + \cdots + \alpha_n' = (n-2)\pi$.

We now show this implies $\alpha_1 = \alpha_2 = \cdots = \alpha_n$. Solving for $\beta_i$ in (2), and substituting in (1), we get $\alpha_{\sigma(i)}' = \frac{1}{2}(\alpha_{i-1}' + \alpha_i')$, $i = 1, \ldots, n$. By the Lemma, $\alpha_1' = \alpha_2' = \cdots = \alpha_n'$, and so each $\alpha_i = (1 - 2/n)\pi$.

It remains to show that the sides are all equal in length. The triangles in FIGURE 3 are similar isoceles triangles having angles $(1 - 2/n)\pi$, $\pi/n$, $\pi/n$, and so for $\lambda = \cos(\pi/n)$ we have

$$l_i = \frac{l_{i-1}'}{2\lambda} + \frac{l_i'}{2\lambda}, \; i = 1, \ldots, n. \tag{3}$$

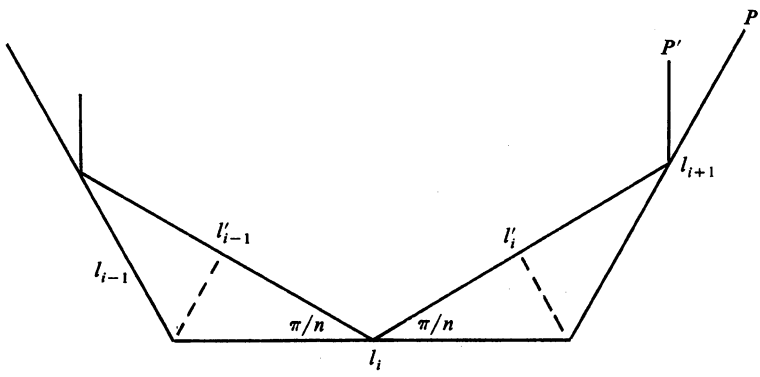



FIGURE 3

But $\mu l_i = l'_{\sigma(i)}$, where $\mu > 0$ is the coefficient of similarity between $P$ and $P'$, and so (3) may be written

$$l'_{\sigma(i)} = a(l'_{i-1} + l'_i),\ i = 1, \ldots, n, \tag{4}$$

where $a = \mu/2\lambda$. Summing these $n$ equations gives

$$\Sigma l'_i = 2a\Sigma l'_i.$$

Thus $a = 1/2$ and from (4) it follows, again by the Lemma, that $l'_1 = l'_2 = \cdots = l'_n$.

It is easy to see that within any regular polygon, other similar polygons can be inscribed which do not form a billiard path (FIGURE 4(a)). It might be interesting to try to find a condition alternative to "$Q$ is a billiard path" under which the following statement is true: if $P$ is a convex polygon containing an inscribed polygon $Q$ similar to $P$, then $P$ must be regular. For instance, the statement can be proved false for $P$ an arbitrary triangle (FIGURE 4(b)) and, true for $P$ an arbitrary rectangle.

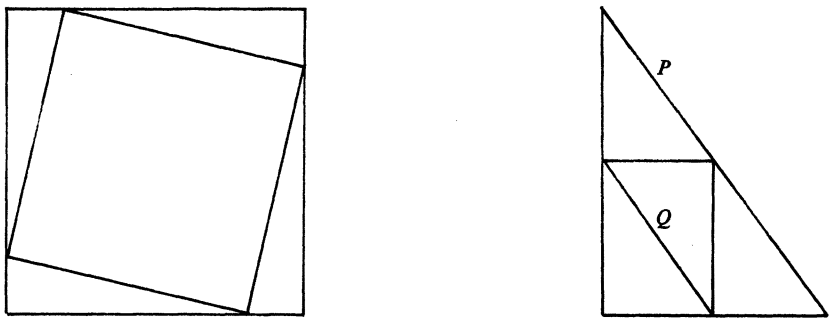



FIGURE 4

In view of our result in the plane, it is natural to ask if the five regular polyhedra in three space have interesting billiard path properties. In particular we can ask the following

*Question. Do all the regular polyhedra have a periodic billiard path of equal line segments hitting each face exactly once?*

Results in [1] show the answer is yes at least for the regular tetrahedron and the cube. On the other hand the existence of such a periodic billiard path does not guarantee that the corresponding polyhedron is regular. This was kindly pointed out to us by the referee, who offered the following argument. Consider a quadrilateral path of equal line segments which is close to the billiard path for the regular tetrahedron described by Martin Gardner in his book [1]. At each vertex take the line that bisects the angle of the two segments of the quadrilateral coming into the vertex. Construct the plane perpendicular to that line through the vertex. These four planes then determine a tetrahedron with the quadrilateral as a billiard path for it. If the quadrilateral is skewed, so will be the final tetrahedron.

**References**

[1] Martin Gardner, Bouncing Ball in Polygons and Polyhedrons, Martin Gardner's Sixth Book of Mathematical Games from Scientific American, Scribner, New York, 1971.

[2] Robert Sine and Vladislav Kreinovič, Remarks on billiards, Amer. Math. Monthly, 86(1979) 204–206.

# Nesting Behavior of Osculating Circles and the Fresnel Integrals

JOEL ZEITLIN
*California State University*
*Northridge, CA 91330*

The following elementary theorem from Stoker [1, p. 31] gives a remarkable and somewhat counterintuitive nesting property of the osculating circles of an arc with monotone curvature.

THEOREM 1. *If $\alpha(s)$ is a regular curve with radius of curvature $\rho(s)$ and $\dot{\rho}(s) \neq 0$ for $s \in I$ (some interval), then no two of the osculating circles at $\alpha(s)$ for $s \in I$ intersect, i.e., the circles are nested within one another.*

In this note I will recall some definitions and background material leading to a proof of Theorem 1, give a computer plot of $r = \theta$ with its osculating circles, and apply Theorem 1 to prove the convergence of the Fresnel integrals $\int_0^\infty \sin t^2\, dt$ and $\int_0^\infty \cos t^2\, dt$.

For any regular curve in the plane there is a parametrization, $\alpha(s)$, by arc length, $s$, so that the **speed** $|\dot{\alpha}(s)|$ is 1, where the dot means differentiation with respect to $s$. The unit **tangent** vector is $T(s) = \dot{\alpha}(s)$ and the unit **normal** vector, $N(s)$, is defined by insisting that $T(s)$ and $N(s)$ form a right-handed system. The Frenet equations give the derivatives of this "moving" basis:

$$\dot{T}(s) = \kappa(s)N(s),$$

$$\dot{N}(s) = -\kappa(s)T(s).$$

The coefficient $\kappa(s)$ of $N(s)$ is defined by the first equation and is called the **signed curvature**. Note that $\kappa(s) = \pm|\ddot{\alpha}(s)|$. The sign of $\kappa(s)$ can be reversed by reparametrizing so that the curve is oriented in the opposite direction. We may thus consider $\kappa(s) \geq 0$ at any particular point, or we may assume that $|\kappa(s)|$ is decreasing if $\dot{\kappa}(s) \neq 0$.

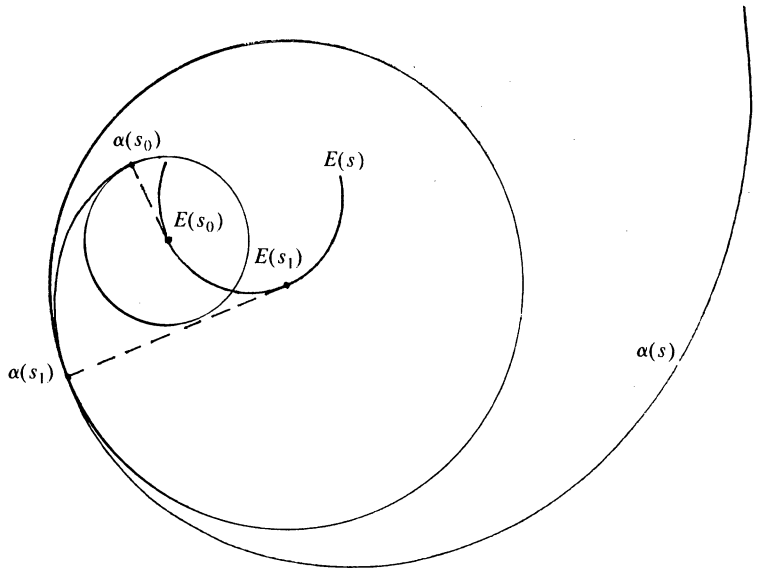



FIGURE 1. A curve with evolute and two nested osculating circles.

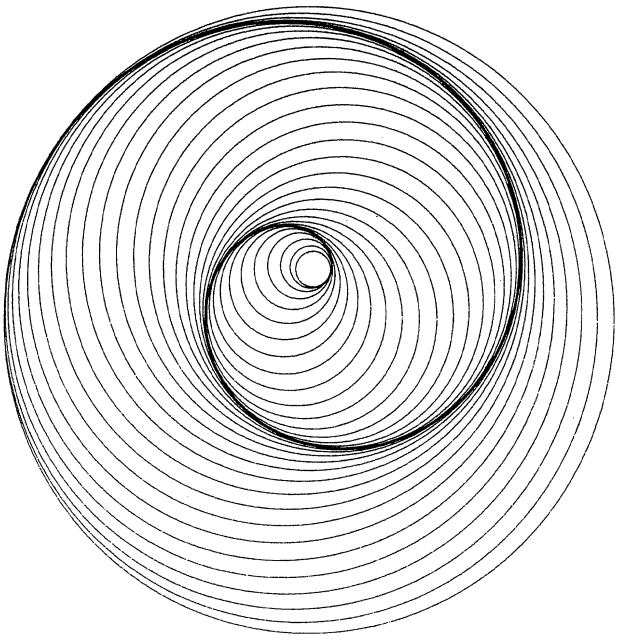


FIGURE 2. **The spiral $r = \theta$ with osculating circles.**

For $\kappa(s) \neq 0$ the **osculating circle** at $\alpha(s)$ is the circle of radius $\rho(s) = 1/|\kappa(s)|$ with center at $E(s) = \alpha(s) + (1/\kappa(s))N(s)$. The quantity $\rho(s)$ is called the **radius of curvature** and the curve, $E(s)$, (the locus of **centers of curvature**) is called the **evolute**. The osculating circle at a point, $\alpha(s)$, is the unique unit speed circle through $\alpha(s)$ which has the same first and second derivatives at the point. Note that $\dot{E}(s) = \pm\dot{\rho}(s)N(s)$ so that $E(s)$ is not automatically of unit speed, and the arc length along $E(s)$ from $E(s_0)$ to $E(s_1)$ is given by

$$L(s_0, s_1) = \int_{s_0}^{s_1} |\dot{E}(s)|\, ds = |\rho(s_1) - \rho(s_0)|.$$

Theorem 1 is now proved by observing that for $\rho(s)$ increasing on $[s_0, s_1]$ and $x$ any point on the osculating circle at $s_0$, the Euclidean distance from $x$ to $E(s_1)$ satisfies the following inequality:

$$d(x, E(s_1)) \leqslant \rho(s_0) + d(E(s_0), E(s_1)) < \rho(s_0) + L(s_0, s_1) \leqslant \rho(s_1).$$

Thus $x$ lies inside the osculating circle at $s_1$. See FIGURE 1.

An **involute** $I(s)$ of $E(s)$ is constructed by imagining a flexible string attached at some point $E(s_0)$ and then "unwrapped" so that $I(s) = E(s) - f(s)N(s)$, where $f(s)$ is a constant plus the length of string unwrapped along $E(s)$ from $s_0$ to $s$. The original curve $\alpha(s)$ is an involute of $E(s)$. (In fact, this is how FIGURE 1 was drawn.)

A plot of $r = \theta$ with osculating circles is shown in FIGURE 2. This plot is from a program written by David Gungner, an undergraduate math major at C.S.U.N. and was executed on a Cal Comp. 936. The family of osculating circles disjointly covers the plane except for (0, 0). This curve is an envelope of the family of osculating circles, i.e., at each point on the curve it is tangent to some member of the family, but it is not a member of the family. This example is particularly nice since the family of circles do not intersect one another. Note that in some books one finds envelopes defined in terms of "limiting positions of intersections of members of the family"—clearly a different notion.

Theorem 1 can be applied in an unusual way to show the convergence of the Fresnel integrals $\int_0^\infty \cos t^2\,dt$ and $\int_0^\infty \sin t^2\,dt$. To do this, we consider the spiral of Fresnel

$$\alpha(s) = \left( \int_0^s \cos t^2\,dt, \int_0^s \sin t^2\,dt \right)$$

for $s \in (0, \infty)$. Then $\dot{\alpha}(s) = (\cos s^2, \sin s^2)$ so $|\dot{\alpha}(s)| = 1$ and $s$ represents arc length from $\alpha(0)$ to $\alpha(s)$. As Stoker [**1**, p. 26] indicates, $\kappa = \dot{x}_1\ddot{x}_2 - \ddot{x}_1\dot{x}_2$, so $\kappa(s) = 2s$ and $\rho(s) = 1/2s$. Theorem 1 now tells us that the osculating circles are nested, so $\alpha(s)$ is on the osculating circle at $\alpha(s)$ and thus inside any osculating circle at $\alpha(t)$ for $t < s$. If we apply Cantor's nested set theorem [**2**, p. 27] to the closed discs $D_n$, obtained by filling in the osculating circles at $\alpha(n)$, we see there is a point $\mathbf{p}$ contained in all $D_n$. Then for all $s$, $\alpha(s)$ and $\mathbf{p}$ are both in $D_s$, so $|\alpha(s) - \mathbf{p}| \leqslant 2\rho(s) = 1/s \to 0$ as $s \to \infty$. The coordinates of $\alpha(s)$ which are the Fresnel integrals must then converge also.

It is an instructive exercise to plot just the first or the second coordinate of $\alpha(s)$, the curve $\alpha(s)$, and/or the osculating circles of $\alpha(s)$ for the interval $0 < s < 10$.

**References**


[1] J. J. Stoker, Differential Geometry, Wiley-Interscience, New York, 1969.

[2] R. Creighton Buck, Advanced Calculus, McGraw-Hill, New York, 1965.


**Laplace Transforms**

In solving for $x(t)$ we pass

through $L$ to $y(s)$

on a roundabout track.

It's like taking a trip through the Looking-Glass,

uncovering $x(t)$

on our way back.

Katharine O'Brien

# Counting the Integer Solutions of a Linear Equation with Unit Coefficients

V. N. Murty
*The Pennsylvania State University*
*Middletown, PA 17057*

Consider a linear equation in $k$ variables $x_1, x_2, \ldots, x_k$ with unit coefficients

$$x_1 + x_2 + \cdots + x_k = m \tag{1}$$

where $m$ is a given positive integer. We wish to count (i) the solutions of (1) in positive integers; (ii) the solutions of (1) in nonnegative integers; (iii) the solutions of (1) in integers satisfying the restriction

$$x_1 > c_1, x_2 > c_2, \ldots, x_k > c_k \text{ where } c_1, c_2, \ldots, c_k$$

are integers fixed in advance; and (iv) the solutions of (1) in positive integers not exceeding a given integer $c$. Niven [1] gave an interesting technique of counting the solutions of (1) in all the four cases. The object of this paper is to present an alternative method of counting which appears to be equally interesting and a little simpler, especially for the situation (iv). These counting techniques play an important role in computing probabilities in random experiments of throwing dice, or classical occupancy problems.

To count the solutions of (1) in positive integers we consider the infinite series expansion

$$\left(\sum_{i=1}^{\infty} x^i\right)^k \quad \text{for } |x| < 1$$

and examine the coefficient of $x^m$ in this expansion. The expansion that we are looking at is the product of $k$ factors. If we pick $x^{x_1}$ from the first factor, $x^{x_2}$ from the second, ..., and $x^{x_k}$ from the $k$th factor and multiply, we obtain $x^{x_1+x_2+\cdots+x_k}$ and hence the coefficient of $x^m$ in the above expansion is precisely the total number of solutions of (1) in positive integers since each $x_i$ is greater than or equal to 1. Noting that

$$\sum_{i=1}^{\infty} x^i = x(1-x)^{-1} \text{ for } |x| < 1$$

implies

$$\left(\sum_{i=1}^{\infty} x^i\right)^k = x^k(1-x)^{-k},$$

and

$$(1-x)^{-k} = \sum_{r=0}^{\infty} \binom{k+r-1}{r} x^r,$$

we have

$$\left(\sum_{i=1}^{\infty} x^i\right)^k = \sum_{r=0}^{\infty} \binom{k+r-1}{r} x^{r+k}.$$

This shows that the coefficient of $x^m$ is $\binom{k+r-1}{r}$ where $r$ is chosen such that $r+k=m$ or $r=m-k$. *Thus the total number of solutions of* (1) *in positive integers is* $\binom{m-1}{m-k}$. *Since* $\binom{n}{r} = \binom{n}{n-r}$, *this total number can also be expressed as* $\binom{m-1}{k-1}$. The above technique of counting is called the "generating function approach" since the infinite series expansion generates (as coefficients) the desired solution.

To count the solutions of (1) in nonnegative integers, all we have to do is to look at the coefficient of $x^m$ in the infinite series expansion

$$\left(\sum_{i=0}^{\infty} x^i\right)^k = (1+x)^{-k} = \sum_{r=0}^{\infty} \binom{k+r-1}{r} x^r$$

*and thus the total number of solutions of* (1) *in nonnegative integers is equal to*

$$\binom{m+k-1}{m} = \binom{m+k-1}{k-1}.$$

To count the solutions of (1) in integers subject to the restriction

$$x_1 > c_1, x_2 > c_2, \ldots, x_k > c_k$$

we look at the linear equation

$$y_1 + y_2 + \cdots + y_k = m - \sum_{i=1}^{k} c_i - k \tag{2}$$

where each $y_i$ is a nonnegative integer and is equal to $(x_i - c_i - 1)$; $i = 1, 2, \ldots, k$. There is a one-to-one correspondence between the solutions of (1) satisfying the given restriction and the solutions of (2) in nonnegative integers. But the number of solutions of (2) in nonnegative integers is known to us from our previous count and is equal to $\binom{r}{k-1}$, where $r = m - \sum_{i=1}^{k} c_i - 1$. *Thus the number of solutions of* (1) *in integers satisfying the restriction*

$$x_1 > c_1, x_2 > c_2, \ldots, x_k > c_k \quad \textit{is} \quad \binom{m-s-1}{k-1}, \textit{ where } s = \sum_{i=1}^{k} c_i.$$

Finally, to count the solutions of (1) in positive integers not exceeding $c$, we look at the coefficient of $x^m$ in the polynomial expansion

$$\left(\sum_{i=1}^{c} x^i\right)^k = x^k (1-x^c)^k (1-x)^{-k},$$

where this has the series expansion

$$x^k (1-x^c)^k (1-x)^{-k} = \left[\sum_{l=0}^{k} (-1)^l \binom{k}{l} x^{lc+k}\right]\left[\sum_{r=0}^{\infty} \binom{k+r-1}{r} x^r\right].$$

The coefficient of $x^m$ in the product on the right is

$$\Sigma (-1)^l \binom{k}{l} \binom{k+r-1}{r} x^{lc+k+r}$$

where the summation runs through values of $l$ and $r$ such that

$$lc + k + r = m;\ r \geqslant 0,\ 0 \leqslant l \leqslant k$$

or

$$k + r = m - lc;\ m - k \geqslant lc;\ 0 \leqslant l \leqslant k$$

or

$$k + r = m - lc;\ 0 \leqslant l \leqslant \frac{m-k}{c}.$$

*Hence the required coefficient is*

$$\sum_{l=0}^{\left[\frac{m-k}{c}\right]} (-1)^l \binom{k}{l} \binom{m-lc-1}{k-1}$$

*which is the total number of solutions of* (1) *in positive integers not exceeding* $c$.

**Reference**

[1] Ivan Niven, Mathematics of Choice: How to Count Without Counting, New Mathematical Library No. 15, Math. Assoc. of Amer. (1965), pp. 54–66.

# Calculating Sums of Powers as Sums of Products

KENNETH R. KUNDERT
*University of Wisconsin-Platteville*
*Platteville, WI 53818*

Noether [4], in an effort to motivate a discussion of statistical estimation, asks his readers to imagine that they are standing on a street corner waiting for a taxi. Observing the numbers on the shields of the next $p$ taxis which pass by, he then proposes several rather simple methods for estimating the total number $n$ of taxis in the fleet. Bailey [1] provides a rigorous investigation of a similar problem involving the estimation of the number $n$ of cars taking part in a race. Given a random sample $X_1, X_2, \ldots, X_p$ (with replacement) of serial numbers (for example, the numbers on the cars which cross the finish line), Bailey examines properties of $H = \max(X_1, X_2, \ldots, X_p)$ and, in particular, shows that the expected value of $H$ is

$$E(H) = n - n^{-p} \sum_{h=0}^{n-1} h^p.$$

Efforts to actually calculate the expected value are dependent upon one's ability to evaluate the summation at the far right.

Sums of powers, similar to the summation given above, become cumbersome and time-consuming to compute whenever $n$ gets even slightly large. It is possible, however, to evaluate a summation of the form $\sum_{x=1}^{n} x^p$ by evaluating an alternate summation of but $p$ terms (resulting in considerable savings of computer time, since $p$ is ordinarily much smaller than $n$). The technique, common to numerical analysis (see, for example, Froberg [2]), requires Stirling's numbers of the second kind. Janardan and Janardan [3] describe the generation of these values through a recursion relation. We will show how these same values can be obtained by synthetic division.

Begin by considering the factorial polynomial

$$x^{(k)} = \begin{cases} x(x-1)(x-2)\cdots(x-k+1) = \dfrac{x!}{(x-k)!} & \text{for } x \geqslant k \geqslant 1 \\ 0, & \text{elsewhere.} \end{cases} \tag{1}$$

Janardan and Janardan [3] show that powers $x^p$ of $x$ can be represented as sums of factorial polynomials in the following way:

$$x^p = \left|{p \atop 1}\right| x^{(1)} + \left|{p \atop 2}\right| x^{(2)} + \cdots + \left|{p \atop p}\right| x^{(p)} = \sum_{k=1}^{p} \left|{p \atop k}\right| x^{(k)}, \tag{2}$$

where the coefficients $\left|{p \atop 1}\right|, \left|{p \atop 2}\right|, \ldots, \left|{p \atop p}\right|$ are Stirling's numbers of the second kind. For example, the authors show that $x^5$ can be written as

$$x^5 = 1x^{(5)} + 10x^{(4)} + 25x^{(3)} + 15x^{(2)} + 1x^{(1)}.$$

If we synthetically divide $x^4$ by $x-1$, $x-2$, $x-3$ and $x-4$, we obtain the coefficients (Stirling's numbers) 1, 10, 25, 15 and 1 as remainders in our synthetic division process.

$$
\begin{array}{r|cccccl}
1 & 1 & 0 & 0 & 0 & 0 & \\
 & & 1 & 1 & 1 & 1 & \\
\hline
2 & 1 & 1 & 1 & 1 & \boxed{1} & \leftarrow \left|\begin{matrix}5\\1\end{matrix}\right| = r_1 \\
 & & 2 & 6 & 14 & & \\
\hline
3 & 1 & 3 & 7 & \boxed{15} & \leftarrow \left|\begin{matrix}5\\2\end{matrix}\right| & = r_2 \\
 & & 3 & 18 & & & \\
\hline
4 & 1 & 6 & \boxed{25} & \leftarrow \left|\begin{matrix}5\\3\end{matrix}\right| & & = r_3 \\
 & & 4 & & & & \\
\hline
 & \boxed{1} & \boxed{10} & \leftarrow \left|\begin{matrix}5\\4\end{matrix}\right| & & & = r_4 \\
 & \nwarrow \left|\begin{matrix}5\\5\end{matrix}\right| & & & & & = r_5
\end{array}
$$

In general, the coefficient $\left|\begin{matrix}p\\k\end{matrix}\right|$ in (2) is the remainder obtained through repeated synthetic division of $x^{p-1}$ by $x-1$, $x-2,\ldots,x-k$. Denoting the coefficient $\left|\begin{matrix}p\\k\end{matrix}\right|$ by $r_k$, equation (2) can be written

$$x^p = r_1 x^{(1)} + r_2 x^{(2)} + \cdots + r_p x^{(p)} = \sum_{k=1}^{p} r_k x^{(k)}. \tag{3}$$

Referring back to the factorial polynomial in (1), let $f(x) = x^{(k+1)}/(k+1)$. Since $x^{(k+1)} = x!/(x-k-1)!$, we have

$$f(x+1) - f(x) = \frac{\dfrac{(x+1)!}{(x-k)!} - \dfrac{x!}{(x-k-1)!}}{k+1} = \frac{x!}{(x-k)!} = x^{(k)}$$

and

$$\sum_{x=1}^{n} x^{(k)} = \sum_{x=1}^{n} [f(x+1) - f(x)] = f(n+1) - f(1) = \frac{(n+1)^{(k+1)}}{k+1}. \tag{4}$$

Summing both sides of equation (3) with respect to $x$, it follows that

$$\begin{aligned}\sum_{x=1}^{n} x^p &= \sum_{x=1}^{n} \sum_{k=1}^{p} r_k x^{(k)} = \sum_{k=1}^{p} r_k \sum_{x=1}^{n} x^{(k)} \\ &= \sum_{k=1}^{p} r_k \frac{(n+1)^{(k+1)}}{k+1} = \sum_{k=1}^{p} r_k T_k \end{aligned} \tag{5}$$

where

$$T_1 = \frac{(n+1)^{(2)}}{2} = \frac{(n+1)n}{2} \text{ and } T_{k+1} = \frac{(k+1)(n-k)}{(k+2)} T_k \tag{6}$$

for $k = 1, 2, \ldots, p-1$.

Regardless of whether $n$ is 10 or $10^{10}$, the summation $\Sigma_{x=1}^{n} x^p$ can be evaluated by summing the $p$ products shown on the right side of (5). One of the factors in each of these products, $r_k$, is a Stirling number of the second kind and is obtained by synthetic division. The other factor, $T_k$, is determined from the recursion relation given in (6). Since computation involves only addition and multiplication, the procedure is extremely efficient when $n$ is large and $p$ is relatively small.

The summation $\Sigma_{x=1}^{n} x^p$ was computed directly (term by term) for $n = 50, 100, \ldots, 250$ and $p = 1, 3, \ldots, 15$. These same quantities were then computed as sums of products, $\Sigma_{k=1}^{p} r_k T_k$, using the function subprogram included at the end of this article. Each of these computations, run on a PDP 11/40, was repeated 100 times. Total CPU times for the 100 runs (rounded to the nearest 10th of a second) are shown in TABLE 1, where the first number for a given $n$ and $p$ is the CPU time for the sum $\Sigma_{k=1}^{p} r_k T_k$, and the number next to that in parentheses is the CPU time for the sum $(\Sigma_{x=1}^{n} x^p)$.

| $n$ \ $p$ | 1 | 3 | 5 | 7 | 9 | 11 | 13 | 15 |
|---|---|---|---|---|---|---|---|---|
| 50 | $.5_{(5.3)}$ | $1.9_{(6.1)}$ | $4.1_{(6.5)}$ | $6.8_{(6.9)}$ | $10.3_{(7.0)}$ | $14.5_{(7.3)}$ | $19.4_{(7.3)}$ | $24.9_{(7.6)}$ |
| 100 | $.5_{(10.4)}$ | $2.0_{(12.1)}$ | $4.1_{(12.9)}$ | $6.9_{(13.5)}$ | $10.3_{(13.8)}$ | $14.5_{(14.5)}$ | $19.4_{(14.5)}$ | $24.9_{(15.2)}$ |
| 150 | $.5_{(15.4)}$ | $2.0_{(18.0)}$ | $4.0_{(19.2)}$ | $6.9_{(20.3)}$ | $10.4_{(20.7)}$ | $14.5_{(21.6)}$ | $19.4_{(21.7)}$ | $24.9_{(22.7)}$ |
| 200 | $.5_{(20.5)}$ | $2.0_{(23.9)}$ | $4.1_{(25.7)}$ | $6.8_{(27.0)}$ | $10.4_{(27.5)}$ | $14.5_{(28.9)}$ | $19.3_{(28.9)}$ | $24.9_{(30.2)}$ |
| 250 | $.6_{(25.6)}$ | $2.0_{(29.8)}$ | $4.1_{(32.1)}$ | $6.9_{(33.8)}$ | $10.3_{(34.3)}$ | $14.5_{(36.1)}$ | $19.4_{(36.1)}$ | $25.0_{(37.9)}$ |

TABLE 1

The following function subprogram, written in BASIC and run on a PDP 11/40, provides a means to facilitate the calculation of $\Sigma_{k=1}^{p} r_k T_k$.

```
100 DEF FNS (P, N)
105 REM: P = POWER ON X *** N = UPPER LIMIT OF SUMMATION
110 DIM R(50)
115 P1 = P : N1 = N
120 REM: SYNTHETIC DIVISION ROUTINE TO OBTAIN STIRLING'S NUMBERS
125 R(1) = 1 : R(I) = 0 FOR I = 2 TO P1
130 FOR I = 1 TO P1 - 1
135 R(K) = R(K) + I*R(K - 1) FOR K = 2 TO P1 - I + 1
140 NEXT I
145 REM: SUM PRODUCTS OF THE FORM R(K)*T(K)
150 T = (N1 + 1)*N1/2 : S = T
155 FOR K = 2 TO P1
160 T = T*K*(N1 - K + 1)/(K + 1) : S = S + R(P1 - K + 1)*T
165 NEXT K
170 FNS = S
175 FNEND
```

**References**


[1] B. J. R. Bailey, The racing track revisited, Amer. Statist., 31 (1977) 30–33.

[2] C. E. Froberg, Introduction to Numerical Analysis, Addison-Wesley, Reading, MA, 1965.

[3] S. K. Janardan and K. G. Janardan, Stirling's numbers of the second kind—programming Pascal's and Stirling's triangles, Two-Year College Math. J., 9(1978) 243–248.

[4] G. Noether, Introduction to Statistics, 2nd ed., Houghton Mifflin, Boston, 1976.

# PROBLEMS

**Dan Eustice, Editor**
**Leroy F. Meyers, Associate Editor**
*The Ohio State University*

## Proposals

*To be considered for publication, solutions should be mailed before September 1, 1981.*

**1116.** In a Hamming $(2^r - 1, 2^r - r - 1)$ single error correcting code, $2^r - r - 1$ information bits and $r$ redundancy bits are transmitted. If at most one of the transmitted bits is in error, the receiver is able to correct the error and obtain the correct message. (If no redundancy bits were used, all $2^r - r - 1$ information bits would have to be correct in order to receive the correct message.) If $p$ is the probability of error in each bit, for what values of $p$ is the probability of the receiver obtaining the correct message greater with redundancy than without? [*G. A. Heuer, Concordia College, & Tom Stratton, student, Massachusetts Institute of Technology.*]

**1117.** Suppose that every pitch in a baseball game has fixed probabilities $p, q, r$, respectively, of resulting in a ball, a strike that is not a foul, or a foul. (The probability for all other events is then $1 - p - q - r$; these constitute the events that end a player's time at bat in any way other than a walk or a strikeout.) Let $P_K, P_B, P_E$, respectively, denote the probabilities that a given batter strikes out, gets a walk, or does anything else. Find closed-form formulas for $P_K$, $P_B$, and $P_E$ in terms of $p, q, r$. [*David A. Smith, Duke University.*]

**1118*.** Suppose $P$ is a nonempty set of prime numbers such that for all $p$ and $q$ in $P$, all the prime divisors of $pq + 1$ are in $P$. Is $P$ the set of all primes? [*F. David Hammer, University of California, Davis.*]

**1119.** Let the triangle $ABC$ be inscribed in a circle and let point $P$ be the centroid of the triangle. The line segments $AP$, $BP$, and $CP$ are extended to meet the circle in points $D$, $E$, and $F$, respectively. Prove that

$$\frac{|AP|}{|PD|} + \frac{|BP|}{|PE|} + \frac{|CP|}{|PF|} = 3.$$

[*K. R. S. Sastry, Addis Ababa, Ethiopia.*]

---

ASSISTANT EDITORS: Don Bonar, *Denison University;* William A. McWorter, Jr., *The Ohio State University. We invite readers to submit problems believed to be new. Proposals should be accompanied by solutions, when available, and by any information that will assist the editors. Solutions to published problems should be submitted on separate, signed sheets. An asterisk (*) will be placed by a problem to indicate that the proposer did not supply a solution. A problem submitted as a Quickie should be one that has an unexpected succinct solution. Readers desiring acknowledgment of their communications should include a self-addressed stamped card. Send all communications to this department to Dan Eustice, The Ohio State University, 231 W. 18th Ave., Columbus, Ohio 43210.*

**1120.** Let the triangle $ABC$ be inscribed in a circle and let $P$ be a point in the interior of the circle. The line segments $AP$, $BP$, and $CP$ are extended to meet the circle in points $D$, $E$, and $F$, respectively. Describe all such $P$ for which

$$\frac{|AP|}{|PD|}+\frac{|BP|}{|PE|}+\frac{|CP|}{|PF|}\leqslant 3.$$

[*Peter Ørno, The Ohio State University.*]

**1121.** After getting three hits in four times at bat, a baseball player's average changes from .233 to .252. Can one determine how many times the player has been at bat during the season? [*V. Frederick Rickey, Bowling Green State University.*]

# Quickies

*Solutions to Quickies appear at the conclusion of the Problems section.*

**Q667.** Let $A$ and $B$ be two points in $R^n$ ($n$-dimensional Euclidean space) and let $\pi$ be a linear subspace of $R^n$. Let the (orthogonal) projections of $A$ and $B$ into $\pi$ be $A_1$ and $B_1$, respectively. Let $P$ be a point in $\pi$ and let the projections of $A$ on $PB_1$ and $B$ on $PA_1$ be $A_2$ and $B_2$, respectively. Prove that $A_1$, $A_2$, $B_1$ and $B_2$ are concyclic. [*I. J. Good & D. R. Jensen, Virginia Polytechnic Institute.*]

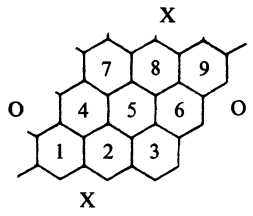
# Solutions

**Kriegspiel Hex** **November 1979**

**1084.** In the game of Kriegspiel Hex two players sit back to back, each with his own Hex board. (For a description and discussion of the game of Hex see Chapter 8 of *The First Book of Mathematical Puzzles and Diversions* by Martin Gardner, Simon and Schuster (1959).) An umpire with a master board directs the game as each player attempts to make a legal move without seeing his opponent's move. The umpire's duties are: (1) Advise each player of his turn, following a legal move by his opponent. (2) Declare an illegal move so that the offending player can try a different move. (3) State when a player has won.

(a) Show that there is a winning strategy for the first player in Kriegspiel Hex played on a $3\times 3$ board.

(b) Prove that there is no winning strategy for the first player in Kriegspiel Hex played on an $n\times n$ board, $n\geqslant 4$. [*William A. McWorter, Jr., The Ohio State University.*]




*Solution*: (a) Let the hexagons be numbered as shown. Player 1 is X, Player 2 is O. The winning strategy for X is indicated by the following diagrams. The given plays of O are known to X as she has previously attempted to play at these positions. Either

| X | 3 | 6 | 8 or 9 |
|---|---|---|---|
| O | ? | ? | |

or

| X | 3 | 7 | 5 |
|---|---|---|---|
| O | 6 | ? | |

or

| X | 3 | 7 | 4 | 1 or 2 |
|---|---|---|---|---|
| O | 6 | 5 | ? | |

or

| X | 3 | 6 | 7 | 5 | |
|---|---|---|---|---|---|
| O | 8 | 9 | ? | | |

or

| X | 3 | 6 | 7 | 4 | 1 or 2 |
|---|---|---|---|---|---|
| O | 8 | 9 | 5 | ? | |

.

(b) Let $n > 3$. To show there is no winning strategy for X it suffices to show that for every arrangement of $n-1$ X's on an $n \times n$ Hex board, there is an arrangement of $n-1$ O's such that the combination of these two is a losing position for X.

Suppose the board is situated so that X is attempting to form a chain from top to bottom using the columns and O a chain from right to left using the rows. Let an arrangement of $n-1$ X's be given. Since there are $n-1$ pairs of adjacent rows, either there is a pair of adjacent rows which contains at most one X, or $n$ is odd and each of the even-numbered rows contains exactly two X's. In the first case, we may assume, without loss of generality, that the upper row of a pair contains no X's. DIAGRAMS 1 and 2 show that there is a suitable arrangement of O's whether the hexagon labeled A is empty or contains an X. In the second case, $n \geqslant 5$ and DIAGRAM 1 can be used unless all the X's are placed at the ends of their rows, in which case DIAGRAM 3 shows a suitable arrangement of O's.

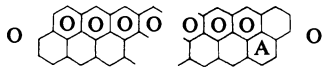



DIAGRAM 1

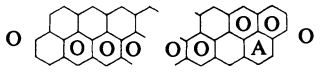



DIAGRAM 2

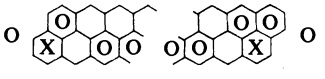



DIAGRAM 3

DUANE BROLINE
University of Evansville

*Also solved by Chico Problem Group of California State University and the proposer.*

## Another GCD Problem

**January 1980**

**1091.** Let $(x,y)$ denote the gcd of integers $x$ and $y$. If $a$ and $b$ are relatively prime integers with $a > b$, prove that for every pair of positive integers $m$ and $n$ we have

$$(a^m - b^m, a^n - b^n) = a^{(m,n)} - b^{(m,n)}.$$

[*Tom M. Apostol, California Institute of Technology.*]

*Solution*: Let $d=(m,n)$, $e=a^d-b^d$, $\delta=(a^m-b^m,a^n-b^n)$. We will show that $\delta=e$. Now $d|m$ so $e|(a^m-b^m)$. Similarly, $d|n$ so $e|(a^n-b^n)$, hence $e|\delta$. Now we show that $\delta|e$. Choose integers $x>0$ and $y>0$ such that $mx-ny=d$. Then

$$a^{mx}=a^{ny+d}=a^{ny}(b^d+e),$$

so

$$a^{mx}-b^{mx}=a^{ny}(b^d+e)-b^{ny+d}=b^d(a^{ny}-b^{ny})+ea^{ny}.$$

Now $\delta|(a^m-b^m)$ so $\delta|(a^{mx}-b^{mx})$. Also, $\delta|(a^n-b^n)$ so $\delta|(a^{ny}-b^{ny})$ and the last equation shows that $\delta|ea^{ny}$. But $(\delta,a)=1$ since any common divisor of $\delta$ and $a$ divides $a^m$, $a^m-b^m$, hence $b^m$, but $(a,b)=1$. Since $\delta|ea^{ny}$ and $(\delta,a)=1$ it follows that $\delta|e$, hence $\delta=e$.

TOM APOSTOL
California Institute of Technology

*Also solved by Gordon Fisher, L. Kuipers (Switzerland), Dan Shapiro, Lawrence Somer, B. Viswanathan (Canada), and Edward T. H. Wang (Canada). Shapiro provided a reference to R. D. Carmichael, On numerical factors of the arithmetic forms $\alpha^n \pm \beta^n$, Annals of Math., 15 (1913) 30–70.*

**An Elliptical Locus** **January 1980**

**1092.** Let $D$ be the disk $x^2+y^2<1$. Let the point $A$ have coordinates $(r,0)$ where $0<r<1$. Describe the set of points $P$ in $D$ such that the open disk whose center is the midpoint of $\overline{AP}$ and whose radius is $AP/2$ is a subset of $D$. [*Roger L. Creech, East Carolina University.*]

*Solution*: Denote the midpoint of $\overline{AP}$ by $C$, and let $Q$ denote the point of intersection of the ray from the origin through $C$ with the circle $x^2+y^2=1$. If $P$ has coordinates $(x,y)$, then

$$C \text{ is } \left(\frac{x+r}{2},\frac{y}{2}\right) \text{ and } Q \text{ is } \left(\frac{x+r}{k},\frac{y}{k}\right),$$

where $k=\sqrt{(x+r)^2+y^2}$. If the distance $CQ$ is greater than or equal to $AP/2$, then the open disk with center $C$ and radius $AP/2$ will be a subset of $D$. The inequality $CQ \geqslant AP/2$ reduces to $x^2+y^2/(1-r^2) \leqslant 1$. Hence $P$ must lie on or within an ellipse, with the exception of the two vertices $(\pm 1,0)$, which do not lie in $D$.

ROGER B. NELSEN
Lewis and Clark College

*Also solved by J. Binz (Switzerland), Walter Bluger (Canada), John Cruthirds & James Morrow, Jordi Dou (Spain), Frank Eccles, L. E. Eimers, Edilio Escalona (Venezuela), Howard Eves, Gordon Fisher, Nick Franceschine III, L. Kuipers (Switzerland), Mou-Liang Kung, Robert A. Leslie, Graham Lord (Canada), B. J. McCartin, P. J. Pedler (Australia), Mike Ratliff, Reuven R. Rottenberg (Israel), Santa Clara Problem Solving Group, Schülerprobleмgruppe (Switzerland), Yan-Loi Wong (Hong Kong), Ken Yocom, and the proposer.*

# Answers

## *Solutions to the Quickies which appear near the beginning of the Problems section.*

**Q667.** The plane $AA_1A_2$ is perpendicular to $B_1P$. Thus $A_1A_2B_1$ is a right angle. Therefore $A_2$ lies on the circle in the plane of $A_1PB_1$ having $A_1B_1$ as a diameter. Similarly, $B_2$ lies on the same circle.

# REVIEWS

**Paul J. Campbell, Editor**
*Beloit College*

**Pierre J. Malraison, Jr., Editor**
*MDSI, Ann Arbor*

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

Gardner, M., *Mathematical games*, Scientific American, 243:6 (December 1980) 18-28.

Discusses Richard Guy's "Strong Law of Small Numbers" which suggests theorems, conjectures and red herrings can arise from observing patterns for small primes.

Thomas, David E., *Mirror Images*, Scientific American 243:6 (December 1980) 206-230.

A look at the geometric transformations obtained with non-planar mirrors, including cylinders, tori and saddle points.

Requicha, A.A.G., *Representations for Rigid Solids: Theory, Methods and Systems*, ACM Comp. Surv. 12:4 (December 1980) 437-464.

This excellent article discusses the current state of research in the area of volumetric modelling: using a computer to develop and save models of solid objects. The main concern is not display or manipulation, but the internal data structures required by such a modeller. Some of the general theoretical problems and some of the approaches of existing volumetric systems are surveyed by the author who then presents the "Constructive Solid Geometry" method used by PADL at the University of Rochester.

Wetherell, C.S., *Probabilistic Languages: a review and some open questions*, ACM Comp. Surv. 12:4 (December 1980) 361-380.

A look at building into a programming language a probability distribution to answer questions of the sort "How likely is a statement of the form $x$ to appear in a program?"

Schneider, D.I., An Annotated Bibliography of Films and Videotapes for College Mathematics, MAA, 1980; vi + 107 pp.

Five parts and an index by title. The first two parts list films by distributor and TV courses by subjects. A section on "related areas" covers other bibliographies and films in related fields (computers, physics). The fourth part is invaluable for getting films locally and fairly cheaply: it lists university film libraries and their holdings. The fifth part is a subject index. An excellent handbook for someone wanting to set up a mathematics film festival.

Hofstadter, D.R., *Metamagical Themas*, Scientific American 244:1 (January 1981) 22-32.

Announcing his arrival with an anagram of "Mathematical Games," Douglas Hofstadter, author of Gödel, Escher, Bach, begins a period of alternating his columns with those of Martin Gardner with an article on self-referential sentences. My own favorite:

"A ceux qui ne comprennent pas l'anglais, la phrase citee ci-dessous ne dit rien: 'For those who understand no French, the French sentence that introduced this quoted sentence has no meaning.'"

Harel, D., *et al.*, *Self Reference Referenced and Self Referenced*, CACM 23:12 (December 1980) 736-737.

An interesting exchange of correspondence on self-reference. (See review above.)

Schechter, Bruce, *A prime discovery*, Discover 2:1 (January 1981) 30-31.

Popular account of the recent accomplishment of Leonard Adleman (MIT/USC) and colleagues in devising a much more efficient method of determining whether a large integer is prime or not.

Andrews, George E., Partitions: Yesterday and Today, New Zealand Mathematical Society, 1979; 56 pp.

Written versions of 3 splendid lectures, in each of which the author endeavors "to show that significant ideas that arose 50 to 100 years ago are yet today the source of new mathematical discoveries." The underlying theme uniting the lectures is the concept of the Durfee square.

Crossley, John N., The Emergence of Number, Upside Down A Book Company (Box 226, Yarra Glen, Victoria, Australia 3775), 1980; 376 pp, A$14.50 + A$2.50 postage (P).

A scholarly but readable tracing of the origins and early development of the natural numbers, the complex numbers, and irrational numbers. The author offers in each case a genesis, investigating psychological, historical, and mathematical answers for how particular kinds of number emerged.

Hankins, Thomas L., Sir William Rowan Hamilton, Johns Hopkins U. Pr., 1980; xxi + 474 pp, $32.50.

In 1865 Hamilton was elected as the first Foreign Associate of the U.S. National Academy of Sciences; today his name is unfamiliar to most historians and scientists. Why was he so honored in his own time? "The answer is that Hamilton was one of the most imaginative mathematicians of the nineteenth century." In addition, he was an extraordinary human being. Both scientific talents and personal qualities emerge in this engaging biography.

Kline, Morris, The Loss of Certainty, Oxford U. Pr., 1980; 366 pp, $19.95.

Describes in non-technical language the "fall" of mathematics: the mathematical discoveries (non-Euclidean geometry, quaternions, paradoxes, Gödel's results) that led to the realization by mathematicians that mathematics does not possess the absolute and infallible truth for which the public at large admires and respects it. Though the overall iconoclastic tone tends toward negativism, the book ends on a positive note: despite its defects and limitations, mathematics is "man's supreme intellectual achievement and the most original creation of the human spirit," a jewel not be rejected for its imperfections, for it affords an unreasonably effective means of interpreting the natural world.

# NEWS & LETTERS

## CHAUVENET PRIZE

Kenneth I. Gross, Professor of Mathematics at the University of North Carolina, has been awarded the 1980 Chauvenet Prize by the MAA. The prize, which recognizes the excellence of his expository paper "On the Evolution of Noncommutative Harmonic Analysis," Am. Math. Monthly, 85(1978) 525-548, was presented at the Business Meeting of the Association on January 11, 1981. In 1979, Professor Gross received the Lester R. Ford Prize for his article. He has been an active participant in MAA activities, and has published numerous articles in his field of research interest, harmonic analysis.

## MATHEMATICS AND THE HUMANITIES

St. John's University, Collegeville, Minnesota, will be the site of "Mathematics and the Humanities--a Student Conference" on April 30 - May 1, 1981. Pi Mu Epsilon papers will be presented; lectures relating mathematics to music, literature and art will be given by Leonard Gillman (music), Donald Koehler (literature) and Doris Schattschneider (art). More information on the conference may be obtained by writing to Prof. Jerry Lenz, Dept. of Math., St. John's Univ., Collegeville, MN 56321.

## REQUEST FOR INFORMATION

Eric Temple Bell, mathematician and writer (best known for his Men of Mathematics), also wrote science fiction under the pen name John Taine. We are preparing a biography of Bell and would be grateful for additional background information in the form of anecdotes, photographs, correspondence, etc., that readers would be willing to share with us. Please write to Prof. G.L. Alexanderson, Dept. of Math., University of Santa Clara, Santa Clara, CA 95053, or to me.

Donald J. Albers
Dept. of Mathematics
Menlo College
Menlo Park, CA 94025

## MORE ON COIN TOSSING

The coin tossing game discussed by Frauenthal and Miller (this *Magazine*, Sept. 1980, pp. 239-243) was posed as a problem by Walter Penney in the October 1969 Journal of Recreational Mathematics. There is a beautiful discussion of it in Martin Gardner's column in the October 1974 Scientific American; Gardner gives an algorithm due to John H. Conway for calculating the odds of any n-tuplet beating any other n-tuplet. The algorithms can be verified by either the Markov chain approach or the directed graph approach of Frauenthal and Miller.

Philip D. Straffin, Jr.
Beloit College
Beloit, WI 53511

## MERSENNE NUMBERS AND BINOMIAL COEFFICIENTS

Two inequalities in "Mersenne Numbers and Binomial Coefficients" (this *Magazine*, Jan. 1981, p. 32) contain typographical errors. Below equation (1), change $0 \leq u \leq s$ to $0 \leq u < s$, and four lines later, change the inequality $0 \leq p \leq 2^{m-2} - 3$ to $0 \leq p < 2^{m-2} - 3$.

Michael Ecker
Luzerne 8, Viewmont Vil.
Scranton, PA 18508

An improved version of the paper by Dacić (this *Magazine*, Jan. 1981, p. 32) can be obtained from the fact that if $p$ is a prime and

$$n = n_0 + n_1 + \ldots + n_r p^r, \quad 0 \leq n_i < p$$

$$s = s_0 + s_1 + \ldots + s_r p^r, \quad 0 \leq s_i < p$$

then

$$\binom{n}{s} \equiv \binom{n_0}{s_0}\binom{n_1}{s_1}\cdots\binom{n_r}{s_r} \pmod{p}.$$

David Roselle
Virginia Tech
Blacksburg, VA 24061

# SOLUTIONS TO 1980 WILLIAM LOWELL PUTNAM EXAMINATION

*The following solutions to the 1980 Putnam Exam questions were prepared by Loren Larson of St. Olaf College, utilizing some of the results prepared by the Putnam Committee.*

A-1. Let $b$ and $c$ be fixed real numbers and let the ten points $(j,y_j)$, $j = 1,2,\ldots,10$, lie on the parabola $y = x^2 + bx + c$. For $j = 1,2,\ldots,9$, let $I_j$ be the point of intersection of the tangents to the given parabola at $(j,y_j)$ and $(j+1, y_{j+1})$. Determine the polynomial function $y = g(x)$ of least degree whose graph passes through all nine points of $I_j$.

*Sol.* The coordinates of $I_j$ are $(\frac{2j+1}{2}, j(j+1) + b(\frac{2j+1}{2}) + c)$. If we set $x_j = \frac{2j+1}{2}$, these coordinates take the form $(x_j, x_j^2 + bx_j + c - 1/4)$, and we see that $g(x) = x^2 + bx + c - 1/4$.

A-2. Let $r$ and $s$ be positive integers. Derive a formula for the number of ordered quadruples $(a,b,c,d)$ of positive integers such that $3^r7^s = lcm[a,b,c] = lcm[a,b,d] = lcm[a,c,d] = lcm[b,c,d]$. The answer should be a function of $r$ and $s$.

*Sol.* Each of $a,b,c,d$ must be of the form $3^m7^n$ with $m$ in $\{0,1,\ldots,r\}$ and $n$ in $\{0,1,\ldots,s\}$. Also, $m$ must be $r$ for at least two of the four numbers and $n$ must be $s$ for at least two of four numbers. Thus, there are $\binom{4}{4}+\binom{4}{3}r+\binom{4}{2}r^2$ choices of allowable $m$'s, and similarly $\binom{4}{4}+\binom{4}{3}s+\binom{4}{2}s^2$ choices of allowable $n$'s. The desired number is therefore $(1 + 4r + 6r^2)(1 + 4s + 6s^2)$.

A-3. Evaluate

$$\int_0^{\pi/2} \frac{1}{1 + (\tan x)^{\sqrt{2}}}\, dx.$$

*Sol.* The integrand (even when $\sqrt{2}$ is replaced by an arbitrary real number) is symmetric about the point $(\pi/4,1/2)$. It follows that the integral is equal to $\pi/4$.

*Alternate Sol.* Let $I$ be the given definite integral and $\sqrt{2} = r$. We show that $I = \pi/4$. Using $x = (\pi/2) - u$, one has

$$I = \int_{\pi/2}^{0} \frac{-du}{1 + \cot^r u} = \int_0^{\pi/2} \frac{\tan^r u\, du}{\tan^r u + 1}.$$

Hence $2I = \int_0^{\pi/2} \frac{1 + \tan^r x}{1 + \tan^r x}\, dx = \int_0^{\pi/2} dx = \pi/2$ and $I = \pi/4$.

A-4. (a) Prove that there exist integers $a,b,c$, not all zero and each of absolute value less than one million, such that $|a + b\sqrt{2} + c\sqrt{3}| < 10^{-11}$.

(b) Let $a,b,c$ be integers, not all zero and each of absolute value less than one million. Prove that

$$|a + b\sqrt{2} + c\sqrt{3}| > 10^{-21}.$$

*Sol.* (a) Let $S$ be the set of $10^{18}$ real numbers $r + s\sqrt{2} + t\sqrt{3}$ with each of $r,s,t$ in $\{0,1,\ldots,10^6 - 1\}$ and let $d = (1 + \sqrt{2} + \sqrt{3})10^6$. Then each $x$ in $S$ is in the interval $0 \le x < d$. Partition this interval into $10^{18} - 1$ equal subintervals, each of length $e = d/(10^{18} - 1)$. By the pigeonhole principle, two of the $10^{18}$ numbers of $S$ must be in the same subinterval, and their difference $a + b\sqrt{2} + c\sqrt{3}$ gives the desired $a,b,c$ since $e < 10^{-11}$.

(b) Let $F_1 = a + b\sqrt{2} + c\sqrt{3}$ and $F_2$, $F_3$, $F_4$ be the other numbers of the form $a \pm b\sqrt{2} \pm c\sqrt{3}$. The product $F_1F_2F_3F_4$ is a nonzero integer, and therefore $|F_1| \ge 1/|F_2F_3F_4| > (1/10^7)^3 = 10^{-21}$.

A-5. Let $P(t)$ be a nonconstant polynomial with real coefficients. Prove that the system of simultaneous equations

$$0 = \int_0^x P(t) \sin t\, dt$$

$$0 = \int_0^x P(t) \cos t\, dt$$

has only finitely many real solutions $x$.

*Sol.* Let $Q$ be the alternating sum of even derivatives of $P$, i.e., $Q = P - P'' + P^{(iv)} - \ldots$. Using repeated integrations by parts, the equations of the given system become

$$\int_0^x P(t) \sin t\, dt = -Q(x)\cos x + Q'(x)\sin x + Q(0) = 0,$$

$$\int_0^x P(t) \cos t\, dt = Q(x)\sin x + Q'(x)\cos x - Q'(0) = 0.$$

These imply that $x$ must satisfy

$$Q(x) = Q'(0)\sin x + Q(0)\cos x.$$

The right side of this equation is bounded, whereas $Q(x)$ is a polynomial of positive degree, and therefore all the solutions are in some interval $|x| \leq M$. In such an interval $P(x)\sin x$ has only finitely many zeros and $\int_0^x P(t)\sin t\, dt = 0$ has at most one more zero by Rolle's Theorem.

A-6. Let $C$ be the class of all real valued continuously differentiable functions $f$ on the interval $0 \leq x \leq 1$ with $f(0) = 0$ and $f(1) = 1$. Determine the largest real number $u$ such that

$$u \leq \int_0^1 |f'(x) - f(x)|\,dx$$

for all $f$ in $C$.

*Sol.*

$$\int_0^1 |f'-f|\,dx \geq \int_0^1 e^{-x}|f'-f|\,dx$$

$$= \int_0^1 |(fe^{-x})'|\,dx \geq \left|\int_0^1 (fe^{-x})'\,dx\right| = 1/e.$$

To see that $1/e$ is the largest lower bound, consider the functions $f_a(x)$ defined by

$$f_a(x) = \begin{cases} (e^{a-1}/a)x & 0 \leq x \leq a \\ e^{x-1} & a \leq x \leq 1 \end{cases}.$$

Then $\int_0^1 |f_a' - f_a|\,dx = e^{a-1}(1 - a/2)$. As $a \to 0$, this expression approaches $1/e$. The function $f_a$ does not have a continuous derivative, but one can smooth out the corner and thus show that no number greater than $1/e$ can be an upper bound.

B-1. For which real numbers $c$ is

$$(e^x + e^{-x})/2 \leq e^{cx^2}$$

for all real $x$?

*Sol.* For $c \geq 1/2$, $\dfrac{e^x + e^{-x}}{2} = \sum_{n=0}^{\infty} \dfrac{x^{2n}}{(2n)!} \leq \sum_{n=0}^{\infty} \dfrac{x^{2n}}{2^n n!} = e^{x^2/2} \leq e^{cx^2}$.

Conversely, if the inequality holds for all $x$, then

$$0 \leq \lim_{x\to 0} \frac{e^{cx^2} - (e^x + e^{-x})/2}{x^2} =$$

$$\lim_{x\to 0} \frac{(1 + cx^2 + \ldots) - (1 + x^2/2 + \ldots)}{x^2}$$

$= c - 1/2$, or equivalently, $c \geq 1/2$.

B-2. Let $S$ be the solid in three dimensional space consisting of all points $(x,y,z)$ satisfying the following system of six simultaneous conditions:

$$x \geq 0,\quad y \geq 0,\quad z \geq 0,$$

$$x + y + z \leq 11,$$

$$2x + 4y + 3z \leq 36,$$

$$2x + 3z \leq 24.$$

(a) Determine the number $v$ of vertices of $S$.
(b) Determine the number $e$ of edges of $S$.
(c) Sketch in the $bc$-plane the set of points $(b,c)$ such that $(2,5,4)$ is one of the points $(x,y,z)$ at which the linear function $bx + cy + z$ assumes its maximum value on $S$.

*Sol.* (a) $v = 7$. The seven vertices are $V_0 = (0,0,0)$, $V_1 = (11,0,0)$, $V_2 = (0,9,0)$, $V_3 = (0,0,8)$, $V_4 = (0,3,8)$, $V_5 = (9,0,2)$, and $V_6 = (4,7,0)$.

(b) $e = 11$. The eleven edges are $V_0V_1$, $V_0V_2$, $V_0V_3$, $V_1V_5$, $V_1V_6$, $V_2V_4$, $V_2V_6$, $V_3V_4$, $V_3V_5$, $V_4V_5$, and $V_4V_6$.

(c) Let $L(x,y,z) = bx + cy + z$. Since $L$ is linear and $(2,5,4)$ is on edge $V_4V_6$, the maximum of $L$ on $S$ must be assumed at $V_4$ and $V_6$ and the conditions on $b$ and $c$ are obtained from $L(0,3,8) = L(4,7,0) \geq L(x,y,z)$, with $(x,y,z)$ ranging over the other five vertices. These imply that $b + c = 2$ with $2/3 \leq b \leq 1$.

B-3. For which real numbers $a$ does the sequence defined by the initial condition $u_0 = a$ and the recursion $u_{n+1} = 2u_n - n^2$ have $u_n > 0$ for all $n \geq 0$? (Express the answer in the simplest form.)

*Sol.* For $n \geq 0$, set $b_n = u_n/2^n$. Then

$$b_n = b_0 + \sum_{k=0}^{n-1} (b_{k+1} - b_k)$$

$$= a + \sum_{k=0}^{n-1} (-k^2/2^{k+1})$$

$= a - \frac{1}{2}\sum_{k=0}^{n-1} k^2/2^k$. To evaluate the last sum, begin with the identity $\sum_{k=0}^{n-1} x^k = \frac{1-x^n}{1-x}$. Differentiate each side and multiply each side of the result by $x$. Then differentiate a second time, multiply by $x$, and set $x = 1/2$. This yields

$$\sum_{k=1}^{n-1} k^2/2^k = 6 - (n^2 + 2n + 3)/2^{n-1}.$$

It follows that $u_n = 2^n b_n = 2^n(a - 3) + (n^2 + 2n + 3)$, and from this we see that $u_n > 0$ for all $n \geq 0$ if and only if $a \geq 3$.

B-4. Let $A_1, A_2, \ldots, A_{1066}$ be subsets of a finite set $X$ such that $|A_i| > \frac{1}{2}|X|$ for $1 \leq i \leq 1066$. Prove there exist ten elements $x_1, \ldots, x_{10}$ of $X$ such that every $A_i$ contains at least one of $x_1, \ldots, x_{10}$. (Here $|S|$ means the number of elements in the set $S$.)

*Sol.* Suppose that $X = \{x_1, x_2, \ldots, x_m\}$, $m \geq 10$ (necessarily). Let $n_i$ be the number of $A_j$, $1 \leq j \leq 1066$, which contain $x_i$. Then $n_1 + n_2 + \ldots + n_m = |A_1| + \ldots + |A_{1066}| > 1066(m/2) = 533m$. Hence one of the $n_i$ exceeds 533; we may assume it is $n_1$. Let $B_1, \ldots, B_s$ be those sets $A_j$ not containing $x_1$ and $Y = \{x_2, x_3, \ldots, x_m\}$. Then $s = 1066 - n_1 \leq 532$ and each $|B_j| > |Y|/2$. Note that if $s = 0$, then $x_1$ is in all $A_j$ and we are done. If $s \neq 0$, we can assume that $x_2$ is in at least as many $B_j$ as any other $x_i$. Repeating the process, let $C_1, \ldots, C_t$ be the $B_j$ not containing $x_2$; as before one can show that $t \leq 265$. We continue in this way. The numbers of sets in the 4th through 10th sequences will number no more than 132, 65, 32, 15, 7, 3, and 1, respectively. Thus we obtain the desired elements $x_1, \ldots, x_{10}$.

B-5. For each $t \geq 0$, let $S_t$ be the set of all nonnegative, increasing, convex, continuous, real valued functions $f(x)$ defined on the closed interval $[0,1]$ for which $f(1) - 2f(2/3) + f(1/3) \geq t[f(2/3) - 2f(1/3) + f(0)]$. Develop necessary and sufficient conditions on $t$ for $S_t$ to be closed under multiplication. (This closure means that, if the functions $f(x)$ and $g(x)$ are in $S_t$, so is their product $f(x)g(x)$. A function $f(x)$ is convex if and only if $f(su + (1-s)v) \leq sf(u) + (1-s)f(v)$ whenever $0 \leq s \leq 1$.)

*Sol.* The product of two nonnegative increasing continuous real valued functions has the same properties. Using the fact that $a \leq c$ and $b \leq d$ imply $ad + bc \leq ab + cd$, it is easy to show that $fg$ is convex when $f$ and $g$ are convex. The function $f(g) = x$ is in $S_t$ for all $t$, whereas $x^2$ is in $S_t$ only for $t \leq 1$. Therefore $t \leq 1$ is a necessary condition for $S_t$ to be closed under multiplication. The argument below shows that it is also a sufficient condition. Let $t \in [0,1]$. For a real valued function $h$ defined on $[0,1]$, let $E(h) = [h(1) - 2h(2/3) + h(1/3)] - t[h(2/3) - 2h(1/3) + h(0)]$. Suppose that $f$ and $g$ are in $S_t$, so $E(f) \geq 0$ and $E(g) \geq 0$. Then $E(fg) = g(2/3)E(f) + f(1/3)E(g) + [f(1)-f(1/3)][g(1) - g(2/3)] - t[f(1/3) - f(0)][g(2/3) - g(0)]$. By convexity, $f(1) - f(1/3) \geq 2[f(1/3) - f(0)]$, and $g(1) - g(2/3) \geq \frac{1}{2}[g(2/3) - g(0)]$. If $t \leq 1$, this implies $E(fg) \geq 0$, so $fg$ is in $S_t$.

B-6. An infinite array of rational numbers $G(d,n)$ is defined for integers $d$ and $n$ with $1 \leq d \leq n$ as follows:

$$G(1,n) = 1/n,$$

$$G(d,n) = \frac{d}{n}\sum_{i=d}^{n} G(d-1, i-1) \text{ for } d > 1.$$

For $1 < d \leq p$ and $p$ prime, prove that $G(d,p)$ is expressible as a quotient $s/t$ of integers $s$ and $t$ with $t$ not an integral multiple of $p$. (For example, $G(3,5) = 7/4$ with the denominator 4 not a multiple of 5.)

*Sol.* Let $F_d(x) = \sum_{n=d}^{\infty} G(d,n)x^n$. One sees that $F_d'(x) = d\,F_{d-1}(x)\,F_1'(x)$ by examining the coefficients of $x^{n-1}$ on each side. Then an induction gives $F_d(x) = [F_1(x)]^d$. Now, for $1 < d \leq p$, the coefficient $G(d,p)$ of $x^p$ in $F_d(x)$ is the coefficient of $x^p$ in

$$\left[\sum_{n=1}^{p-d+1} x^n/n\right]^d$$

and hence $G(d,p) = s/t$ with $s$ and $t$ integers and $t$ a product of primes less than $p$.

# The RAYMOND W. BRINK SELECTED MATHEMATICAL PAPERS Series.

This series is a collection of papers on single topics selected and reprinted from the journals of the Mathematical Association of America. The papers are grouped by subject within the topic of the volume and are indexed by author. Each volume is a rich source of mathematical stimulation for students and teachers.

Four volumes are available in this series:

Volume 1.
**SELECTED PAPERS ON PRECALCULUS**

Edited by Tom M. Apostol, Gulbank D. Chakerian, Geraldine C. Darden, and John D. Neff. xvii + 469 pages. Hardbound.

List: $17.50; MAA Members *(personal use)* $12.50

Volume 2.
**SELECTED PAPERS ON CALCULUS**

Edited by Tom M. Apostol, Hubert E. Chrestenson, C. Stanley Ogilvy, Donald E. Richmond, and N. James Schoonmaker. xv + 397 pages. Hardbound.

List: $17.50; MAA Members *(personal use)* $12.50

Volume 3.
**SELECTED PAPERS ON ALGEBRA**

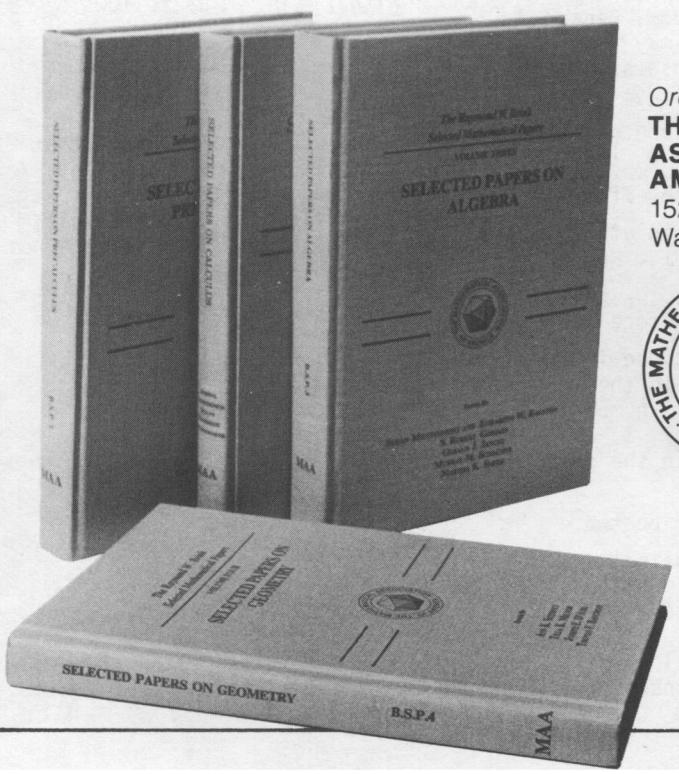
Edited by Susan Montgomery, Elizabeth W. Ralston, S. Robert Gordon, Gerald J. Janusz, Murry M. Schacher, and Martha K. Smith. xx + 537 pages. Hardbound.

List: $17.50; MAA Members *(personal use)* $12.50.

Volume 4.
**SELECTED PAPERS ON GEOMETRY**

Edited by Ann K. Stehney, Tilla K. Milnor, Joseph D'Atri, and Thomas F. Banchoff. ix + 338 pages. Hardbound.

List: $20.00; MAA Members *(personal use)* $14.00




*Order From:*
**THE MATHEMATICAL ASSOCIATION OF AMERICA**
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

# THE DOLCIANI MATHEMATICAL EXPOSITION SERIES.

All four of the books in this widely acclaimed series contain fascinating problems that have ingenious solutions and/or involve intriguing results. Ross Honsberger, author of three, and editor of one of the volumes is a master problem solver. These books will challenge the advanced mathematician as well as stimulate the mathematically talented high school student.

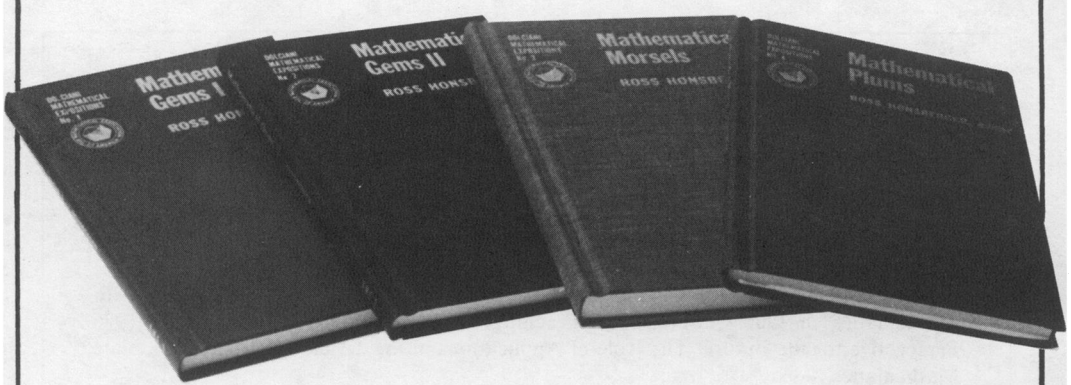
The four volumes available in this series are:

**MATHEMATICAL GEMS I,** by Ross Honsberger. xi + 176 pages. Hardbound.
List: $12.50. MAA Member: $9.00.

**MATHEMATICAL GEMS II,** by Ross Honsberger. ix + 182 pages. Hardbound.
List: $12.50. MAA Member: $9.00.

**MATHEMATICAL MORSELS,** by Ross Honsberger. xii + 249 pages. Hardbound.
List: $16.00. MAA Member: $12.00.

**MATHEMATICAL PLUMS,** edited by Ross Honsberger. Articles by R. P. Boas, G. D. Chakerian, H. L. Dorwart, D. T. Finkbeiner, Ross Honsberger, K. R. Rebman, and S. K. Stein. ix + 182 pages. Hardbound.
List: $14.00. MAA Member: $10.00.




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

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

MATHEMATICS MAGAZINE VOL. 54, NO. 2, MARCH 1981