NEWS AND LETTERS

ACKNOWLEDGEMENTS

In addition to our associate editors, the following have assisted the MAGAZINE *by refereeing papers during the past year. We appreciate the time and care they have given.*

Ayoub, Ayoub B., *The Pennsylvania State University, Abington, PA*
Barbeau, Edward J., *University of Toronto, Ontario, Canada*
Bivens, Irl C., *Davidson College, Davidson, NC*
Bloch, Norman J., *SUNY College at Brockport, Brockport, NY*
Boules, Raouf, *Towson State University, Towson, MD*
Casey, Stephen D., *The American University, Washington, DC*
Castellanos, Dario, *Valencia, Venezuela*
Chollet, John, *Towson State University, Towson, MD*
Fisher, J. Chris, *University of Regina, Saskatchewan, Canada*
Ghahramani, Saeed, *Towson State University, Towson, MD*
Godbole, Anant P., *Michigan Technological University, Houghton, MI*
Goldstein, Larry J., *Pheonixville, PA*
Green, Judy, *Marymount University, Arlington, VA*
Grimaldi, Ralph P., *Rose-Hulman Institute of Technology, Terre Haute, IN*
Guy, Richard, *University of Calgary, Alberta, Canada*
Johnson, Jerry A., *Oklahoma State University, Stillwater, OK*
Katz, Victor J., *University of District of Columbia, Washington, DC*
Krieger, Henry, *Harvey Mudd College, Claremont, CA*
Kung, Sidney J., *Jacksonville University, Jacksonville, FL*
Lay, David C., *University of Maryland, College Park, MD*
Maurer, Stephen B., *Swarthmore College, Swarthmore, PA*
May, Coy, *Towson State University, Towson, MD*
Mayfield, Elizabeth, *Hood College, Frederick, MD*
Michaels, John G., *State University of New York, Brockport, NY*
Morrison, John, *Towson State University, Towson, MD*
Nelsen, Roger B., *Lewis and Clark College, Portland, OR*
Penrice, Stephen, *Emory University, Atlanta, GA*
Richman, Dan, *Baltimore, MD*
Schattschneider, Doris M., *Moravian College, Bethlehem, PA*
Senechal, Marjorie, *Smith College, Northampton, MA*
Stern, Leonid, *Towson State University, Towson, MD*
Tattersall, James J., *Providence College, Providence, RI*
Tucker, Alan, *State University of New York at Stony Brook, Stony Brook, NY*
Wagon, Stan, *Macalester College, St. Paul, MN*
Zorn, Paul, *St. Olaf College, Northfield, MN*

INDEX TO VOLUME 68

AUTHORS

TITLES

PROBLEMS

P, S, Q, *and* C *refers to Proposals, Solutions, Quickies, and Comments, respectively. Page Numbers are given in parentheses. Thus,* P1338(56) *refers to proposal number* 1338, *which appears on page 56.*

February: P1464-1468; Q829-831; S1438-1440.
April: P1469-1473; Q832-834; S1443-1447.
June: P1474-1478; Q835-837; S1448-1453.
October: P1479-1483; Q838-840; S1454-1458.
December: P1484-1488; Q841-843; S1459-1463.

Dear Editor:

This is to point a slight error in "The Phi Number System Revisited," by Cecil Rousseau (October 1995). The Lemma on page 283 is incorrect. If one lets α equal ϕ, then the hypothesis of the Lemma is satisfied in that $\alpha \in \mathbb{Z}[\phi]$ and $1 \leq \alpha < \sqrt{5}$. However, the conclusion is not satisfied as

$$|N(\alpha)| = |N(\phi)| = 1 = |N(1/\phi - 1)| = |N(\alpha - 1)|.$$

This does not affect the main theorem as, in the proof of the theorem, we know $1 \leq \alpha < \phi$ and the Lemma can be proved for this more restrictive case.

Richard E. Stone
100 Birnamwood Drive
Burnsville, MN 55337

Dear Editor:

I certainly liked the argument in "A One-Sentence Proof That $\sqrt{2}$ Is Irrational," by David M. Bloom (October 1995). The author and other readers might be interested in what happened today, when I, for no good reason, was flipping through an old *American Mathematical Monthly* that was lying on my incredibly messy desk. To my amazement, I found basically the same argument in "A Simple Irrationality Proof for the Quadratic Surds," Vol. 75, pp. 772-773. There, the author gets slightly different mileage out of the last equation $\sqrt{k}$ be m/n where m and n are positive and one assumes that either (a) n is the smallest possible value; or (b) m is the smallest; or (c) $m+n$ is the smallest. Each one of these three is contradicted by the equation cited above, hence the author claims he has produced three different proofs of irrationality.

Rick Kreminski
East Texas State University
Commerce, TX 75429

Dear Editor:

There was a calculus manuscript I saw 20 years ago that had a beautiful bunch of algebraic methods for getting the integrals of elementary functions by methods such as those in "Limitless Integrals and a New Definition of the Logarithm," by David Shelupsky (October 1995). I had forgotten the author's name, but thought Vic Klee might remember it as he was the series advisor I was working with at the time. I wrote him, in part:

"The October *Mathematics Magazine* has a nice note on these sorts of formulas by the physicist David Shelupsky. The only problem with it is his notion that all of this is new. It may be original to him, or he may have amplified ideas he not recall hearing, but this is hardly new.

This is a case of no outlook on the past and light reviewing. I must have learned this general method from Burrows Hunt, one of my teachers at Reed. I see it is in his *Calculus and Linear Algebra*, W.H. Freeman and Co., 1967, p. 287, and it was certainly an old observation when I learned it from Hunt in the 1950s.

Those who do not know the past or fail to recall it will have the pleasure of rediscovering it and claiming it for their own as if it were new.

As for me, I am plagued with a memory that I can't put my finger on right now. Last night, after tossing and turning, I had the name I seek clear, after quite a few rounds of approximation. This morning it is gone. Can you refresh it?"

Klee supplied the answer: "Warren Stenberg."

Peter Renz
Academic Press
1300 Boylston St.
Chestnut Hill, MA 02167

Vol. 68, No. 5 December 1995

MATHEMATICS MAGAZINE

- Counting Quota Systems
- Generating Solutions to the N-Queens Problem
- Euclidean Constructions and the Geometry of Origami

An Official Publication of The MATHEMATICAL ASSOCIATION OF AMERICA

EDITORIAL POLICY

The aim of *Mathematics Magazine* is to provide lively and appealing mathematical exposition. This is not a research journal and, in general, the terse style appropriate for such a journal (lemma-theorem-proof-corollary) is not appropriate for an article for the *Magazine*. Articles should include examples, applications, historical background, and illustrations, where appropriate. They should be attractive and accessible to undergraduates and would, ideally, be helpful in supplementing undergraduate courses or in stimulating student investigations. Articles on pedagogy alone, unaccompanied by interesting mathematics, are not suitable. Neither are articles consisting mainly of computer programs unless these are essential to the presentation of some good mathematics. Manuscripts on history are especially welcome, as are those showing relationships between various branches of mathematics and between mathematics and other disciplines.

The full statement of editorial policy appears in this *Magazine*, Vol. 64, pp. 71–72, and is available from the Editor. Manuscripts to be submitted should not be concurrently submitted to, accepted for publication by, nor published by another journal or publisher.

Send new manuscripts to: Paul Zorn, Editor-Elect, Department of Mathematics, St. Olaf College, 1520 St. Olaf Avenue, Northfield, MN 55057-1098. Manuscripts should be typewritten and double spaced and prepared in a style consistent with the format of *Mathematics Magazine*. Authors should submit the original and two copies and keep one copy. In addition, authors should supply the full five-symbol Mathematics Subject Classification number, as described in *Mathematical Reviews*, 1980 and later. Illustrations should be carefully prepared on separate sheets in black ink, the original without lettering and two copies with lettering added. Do not use staples.

Cover art and design by Carolyn Westbrook.

AUTHORS

Stephen C. Young became interested in social choice theory as an undergraduate at Union College, where his honors thesis was written under the supervision of his co-authors. He received his M.A. in mathematics from the University of Wisconsin, and after spending a year teaching part-time at the College of DuPage and Morton College in Chicago, is working as an actuarial analyst for Aon Re Services, Inc.

Alan D. Taylor received his Ph.D. from Dartmouth College in 1975 and is presently the Marie Louise Bailey Professor of Mathematics at Union College. His research interests include logic and set theory, finite and infinitary combinatorics, and social choice theory. He is the author of *Mathematics and Politics* (Springer-Verlag, 1995) and, with Steven J. Brams, *Fair Division: From Cake-Cutting to Dispute Resolution* (Cambridge University Press, 1996).

William Zwicker attended Harvard College and M.I.T., where he earned his Ph.D. in 1975. He has been on the faculty at Union College since then. His research interests in combinatorial set theory and large cardinals have expanded recently to include combinatorial applications in such areas of political science as social choice theory, game theory, and fair divison. He is also the originator of the Hypergame Paradox. His other interests include mathematics education, music, and cooking.

Cengiz Erbas received his Ph.D. degree in Computer Science from Southern Methodist University (SMU) in 1994. Since then, he has been working for TRW Financial Systems. He developed interest in the *N*-Queens problem and its applications during his graduate studies. The present article is a part of his M.Sc. thesis, which he completed in 1992 under Murat M. Tanik.

Murat M. Tanik is the director of Software Systems Engineering Institute at the University of Texas, Austin. He is also Visiting Professor in the Department of Electrical and Computer Engineering. He has worked on software and systems-related projects since 1974 for NASA, Arthur A. Collins, and Raymond T. Yeh, and has been enjoying various aspects of *N*-Queens problem and its applications since 1974.

Robert Geretschläger teaches Mathematics and Descriptive Geometry (a separate subject in Austrian schools due to a fortunate quirk in the local school system) at Bunderealgymnasium Keplerstrasse in Graz, Austria. Although his doctoral thesis was in functional analysis, his main interests now are in geometry and inequalities. For the last five years, he has been a tutor for the final round of the Austrian Mathematical Olympiad.

Vol. 68, No. 5 December 1995

MATHEMATICS MAGAZINE

The *MATHEMATICS MAGAZINE* (ISSN 0025-570X) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, D.C. 20036 and Montpelier, VT, bimonthly except July/August.
The annual subscription price for the MATHEMATICS MAGAZINE to an individual member of the Association is $16 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $64. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% discount; and new members receive a 40% dues discount for the first two years of membership.) The nonmember/library subscription price is $68 per year.
Subscription correspondence and notice of change of address should be sent to the Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036. Microfilmed issues may be obtained from University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

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

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

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

ARTICLES

Counting Quota Systems[1]: A Combinatorial Question from Social Choice Theory

STEPHEN C. YOUNG
Aon Re Services, Inc.
Chicago, IL 60606

ALAN D. TAYLOR
WILLIAM S. ZWICKER
Union College
Schenectady, NY 12308

Introduction

We start with a brief introduction to social choice theory, a field where mathematics is applied to political science and economics. Our focus is on voting procedures for two alternatives. We consider a generalization of majority rule that we call a quota system, and we ask a counting problem—how many quota systems exist for a given number of voters? The solution reveals a connection between quota systems and Pascal's triangle, and leads to yet another appearance of the Catalan numbers.

1. Social Choice Procedures

Several voters wish to choose among a collection of alternatives. Any choice for the entire group should reflect the desires of the individual voters as much as possible. But it may not always be easy to do this. We need a procedure that will decide, based on the individual choices, what the collective choice should be.

We will consider only the situation where the number of alternatives to choose from is exactly two. This is a strong restriction, but we are still left with much to consider. We refer to the two alternatives as X and Y. Each of n voters independently chooses between X and Y (or neither, should they choose to abstain), and then our procedure will decide which of X and Y (or neither, in the case of a tie) has won.

To make this idea rigorous, we define a *social choice procedure* to be a function,

$$f: W^n \to W, \quad \text{where } W = \{X, Y, 0\},$$

with 0 representing an abstention or tie. The elements of W^n are called *profiles*. The procedure f chooses an outcome for every possible election among the n voters.

Let us now briefly consider some examples of social choice procedures. *Simple majority vote* is the procedure that chooses the alternative, X or Y, with the most votes, or declares a tie if both X and Y have the same number of votes. Though it

[1]This paper is a revised version of the first author's Undergraduate Honors Thesis at Union College, written under the supervision of the second and third authors.

may be the most familiar, simple majority is certainly not the only possible procedure; there are as many social choice procedures as there are functions from W^n to W. Some other examples of social choice procedures follow.

- Let A be the procedure that chooses the alternative that receives an even number of votes (and declares a tie if neither or both do). A corresponds to a parity function.
- Let B be the procedure that simply chooses Y for *any* profile of votes. B corresponds to a constant function.
- Choose one of the voters to be the *dictator* and let C be the procedure that chooses whichever alternative the dictator chooses (and declares a tie if neither is chosen). C corresponds to a projection function.

These procedures seem arbitrary and unjust; they are not equitable in their way of representing the voters. Are there equitable alternatives to majority rule?

2. Social Choice Properties

Since there are so many possible social choice procedures to choose from, we need a basis to compare procedures. For example, it may seem to you that the dictator procedure is less fair than simple majority vote, particularly if you yourself are not the dictator! For this reason we define three desirable properties for social choice procedures. Our properties should have some basis in what we believe to be fair or unfair. For example, we may want our procedure to be fair to the individual voters; that is, each vote should have equal weight. We define *anonymity* as invariance under permutation of the voters. It is clear that procedure C, the dictatorship procedure, is not anonymous. We may also want our procedure to be fair to the different alternatives; that is, a vote for X should benefit X as much as a vote for Y benefits Y. We define *duality* as invariance under permutation of the alternatives. Certainly procedure B, the constant Y procedure, is not dual.

By requiring that our procedure be dual and anonymous, we can avoid procedures such as B and C. But procedure A, the parity procedure, satisfies both duality and anonymity. In order to avoid such procedures, we define one additional property. First define a *change favorable to* X as any of the following changes in the profile of votes:

- changing a vote for Y to a tie vote,
- changing a tie vote to a vote for X, or
- changing a vote for Y to a vote for X.

For a given profile, suppose a procedure chooses X or declares a tie. If, after a change favorable to X is made, the procedure always then chooses X, we say that the procedure satisfies *monotonicity*. The intent is that a change favorable to X should not hurt X. Procedure A is not monotone: If X wins with an even number of votes, a change of a vote for Y to a vote for X will cause X to lose with an odd number of votes.

3. May's Theorem and Quota Systems

It is not difficult to see that simple majority vote is anonymous, dual, and monotone. In 1952, Kenneth May published a much stronger result, showing that simple majority vote is the *only* procedure that satisfies these three properties **[6]**. The theorem follows largely from the strength of the monotonicity condition; aside from saying that a change favorable to X should not hurt X, it also says that a change favorable to X will break a tie. This is a very strict property for a procedure to satisfy. May's Theorem can be generalized by asking, "How does May's result change when we weaken the monotonicity condition?" Define a procedure to be *weakly monotone* if:

- whenever the procedure chooses X, the procedure will still choose X after a change favorable to X is made, and
- whenever the procedure declares a tie, then the procedure will not choose Y after a change favorable to X is made (although it may again declare a tie).

Now we ask which social choice procedures are anonymous, dual, and weakly monotone. Simple majority vote certainly is, but are there any others? In fact there are; simple majority vote generalizes to a *quota system* **[10]**. A quota system consists of a list of quotas, $\{q_0, \dots, q_n\}$, which specify, according to the number of tie votes, how many votes an alternative needs to win. That is, the quota system declares a tie unless, for some number k, there are exactly k tie votes in the profile and one of the alternatives has at least q_k votes, in which case that alternative is the social choice.

In order for our quota system to be well defined, we must have

$$\frac{n-k}{2} < q_k, \quad \text{for } 0 \leq k \leq n, \tag{1a}$$

otherwise both X and Y could achieve the quota in some profile. In the other direction, if the quota q_k is higher than $n-k$, then it is too high for either alternative to attain, so every profile with k tie votes will result in a tie. We do not want to exclude this possibility, but we make the convention that

$$q_k \leq n-k+1, \quad \text{for } 0 \leq k \leq n, \tag{1b}$$

in order to avoid redundant quota systems. Finally, in order for our quota system to be weakly monotone, we require that

$$\text{either } q_{k+1} = q_k \text{ or } q_{k+1} = q_k - 1, \quad \text{for } 0 \leq k < n. \tag{2}$$

When we construct quota systems in the next section, condition (1) will give us bounds on each of the quotas, and condition (2) will restrict how the sequence progresses.

Just as in the case of simple majority vote, it is not difficult to see that quota systems satisfy anonymity, duality, and weak monotonicity. It is again less obvious that quota systems are the only such social choice procedures.

THEOREM 1. *Any social choice procedure that is anonymous, dual, and weakly monotone can be represented as a quota system.*

Proof. Let f be such a social choice procedure. Since f is anonymous, any two profiles that differ by a permutation of the votes will result in the same social choice. Thus it suffices to consider profiles of the form

$$(X, \dots, X, Y, \dots, Y, 0, \dots, 0).$$

For such a profile, let n_X be the number of votes for X, n_Y the number of votes for Y, and n_0 the number of votes for 0. It is clear that f depends only on n_X, n_Y, and n_0.

Now for each k with $0 \leq k \leq n-1$, consider the collection of profiles with $n_0 = k$. If each of these profiles results in a tie, define $q_k = n - k + 1$. Otherwise define q_k to be the smallest number of votes for X in any profile with k ties and a win for X, that is,

$$q_k = \min_{n_0 = k} \{n_X \mid X \text{ wins with } n_X \text{ votes}\}.$$

In particular, $q_k \leq n - k + 1$, and so q_k satisfies condition (1b).

Since f is weakly monotone, we see that X wins precisely when X gets q_k or more votes. Since f is dual, whenever X wins with j votes, Y must also win with j votes. Thus we must have $(n-k)/2 < q_k$, or else X and Y could both win for the same profile. So q_k satisfies condition (1a).

To see that our choice of q_k satisfies condition (2), note first that if q_{k+1} were greater than q_k, then for some profile, a change of a vote for Y to a tie vote causes X to lose. Further, note that if q_{k+1} were less than $q_k - 1$, then for some profile, a change of a tie vote to a vote for X causes X to lose. In either case we have a violation of weak monotonicity. The only other possibilities are $q_{k+1} = q_k$ or $q_{k+1} = q_k - 1$.

Hence, the sequence $\{q_0, \ldots, q_n\}$ forms a quota system identical to f.

Note that the quota systems do, in fact, generalize simple majority vote: Simple majority can be represented as the quota system where each quota is equal to the least integer greater than half the number of non-tie votes.

4. Building Quota Systems

Our goal now is to answer the following question. How many quota systems are there for n voters? A good place to start is to look at what happens for small values of n, and then search for a pattern.

Consider the case where there are two voters. First we calculate the bounds on q_k, $k = 0, 1, 2$, according to condition (1):

$$1 = \frac{2-0}{2} < q_0 \leq 2 - 0 + 1 = 3 \Rightarrow q_0 \in \{2, 3\}.$$

Similarly, $q_1 \in \{1, 2\}$ and $q_2 \in \{1\}$. Condition (2) imposes an additional requirement:

$$q_1 = q_0 \quad \text{or} \quad q_1 = q_0 - 1.$$

This gives three quota systems for two voters: $\langle 2, 1, 1 \rangle$, $\langle 2, 2, 1 \rangle$, and $\langle 3, 2, 1 \rangle$ (which always results in a tie). Note that $\langle 3, 1, 1 \rangle$ is not a possibility because of condition (2).

Table 2 will help the reader visualize the different possible systems. We have lined up the possibilities for the q_k so that each valid quota system for two voters corresponds to a path down Table 2, as shown in Table 2A. By reading the numbers off each of the paths (from the bottom of the path up), we get the three different quota systems.

As a further example, consider the case of three voters. For $n = 3$ we have:

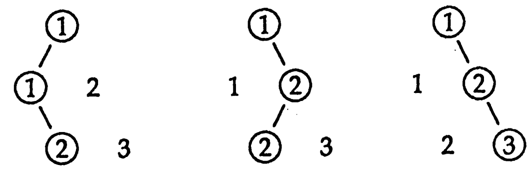
$$\frac{3}{2} = \frac{3-0}{2} < q_0 \leq 3 - 0 + 1 = 4 \Rightarrow q_0 \in \{2, 3, 4\}.$$

Also, $q_1 \in \{2, 3\}$, $q_2 \in \{1, 2\}$, and $q_3 \in \{1\}$. The picture for three voters is shown in Table 3.

TABLE 2. Quota possibilities for $n = 2$ voters.

Possibilities for q_2		1		
Possibilities for q_1	1		2	
Possibilities for q_0		2		3

TABLE 2A. There are three quota systems for two voters.



$\langle 2,1,1\rangle$ $\langle 2,2,1\rangle$ $\langle 3,2,1\rangle$

TABLE 3. Quota possibilities for $n = 3$ voters.

Possibilities for q_3		1			
Possibilities for q_2	1		2		
Possibilities for q_1		2		3	
Possibilities for q_0	2		3		4

The first thing to notice about Table 3 is that it extends Table 2: The top three rows agree in both diagrams. In general, Table l (for l voters) will extend Table k (for k voters), whenever $k \leq l$. To see this, note that by condition (1), for any number j with $0 \leq j \leq n$,

$$\frac{j}{2} = \frac{n-(n-j)}{2} < q_{n-j} \leq n-(n-j)+1 = j+1,$$

and that these bounds are independent of n. Thus, regardless of the number of voters, the jth row of our diagram (which lists the possibilities for q_{n-j}, starting with $j = 0$) will read

$$\left\lfloor \frac{j}{2} \right\rfloor + 1, \left\lfloor \frac{j}{2} \right\rfloor + 2, \ldots, j, j+1,$$

where $\lfloor r \rfloor$ represents the greatest integer less than or equal to r. So all of the diagrams merge into one infinite table, where the quota systems for n voters correspond to paths of length n. We shall call this Table ω. Table 2 and Table 3 are seen to be initial chunks of Table ω.

Returning from our digression, we next consider the number of paths in Table 3. It is clear by inspection that there are 6. See Table 3A.

Notice that each of these paths extends a path from Table 2A. For example, the quota system $\langle 3,2,2,1\rangle$ for three voters extends the quota system $\langle 2,2,1\rangle$ for two voters, and the corresponding paths are in the same relation. In fact, each of the quota systems for two voters has two possible extensions, as each of the paths can be extended to the right or to the left. This explains why the number of quota systems for $n = 3$ is double that for $n = 2$.

TABLE ω. Quota possibilities for any number of voters.

	1								
1		2							
	2		3						
2		3		4					
	3		4		5				
3		4		5		6			
	4		5		6		7		
4		5		6		7		8	
	.						.		
	.							.	
	.								.

TABLE 3A. There are six quota systems for three voters.

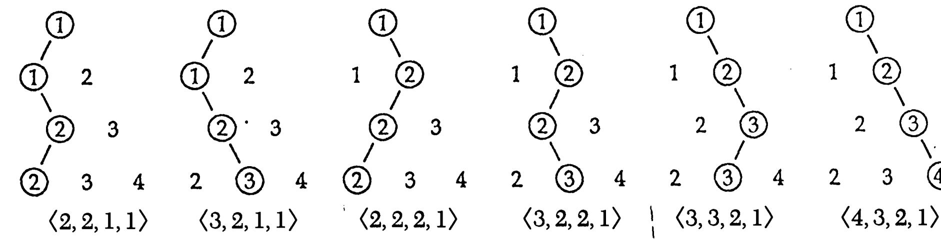


TABLE 4. Quota possibilities for $n = 4$ voters.

Possibilities for q_4		1				
Possibilities for q_3	1		2			
Possibilities for q_2		2		3		
Possibilities for q_1	2		3		4	
Possibilities for q_0		3		4		5

If this were always the case, we would have our counting done. But the situation is slightly more complicated. Consider Table 4.

Each of the paths from Table 3 with q_0 equal to 3 or 4 can be extended in two different ways. But a path with q_0 equal to 2, such as $\langle 2, 2, 2, 1 \rangle$, can only be extended by adding a 3 (to become $\langle 3, 2, 2, 2, 1 \rangle$, in our example).

So here the number of quota systems almost doubles, but not quite. This is an important observation, for if we can determine the number of paths that can be extended in only one way, then we will be able to find a recurrence relation for the number of quota systems.

5. Pascal's Corrupted Triangle

It turns out that there is an important connection between quota systems and Pascal's triangle, shown in FIGURE 1. Recall that the entries of Pascal's triangle are the binomial coefficients, $C(n, r)$. Here, $C(n, r)$ represents the number of ways of choosing r objects from among n objects, and is given by $C(n, r) = n!/[r!(n-r)!]$. If $r < 0$ or

$r > n$, there is no way to choose r objects from n, and $C(n, r) = 0$. The key to generating successive rows of the triangle is the *Pascal rule*,

$$C(n,r) = C(n-1, r-1) + C(n-1, r),$$

easily verified from the formula for $C(n, r)$ above.

The connection between Pascal's triangle and quota systems arises from the following observation. Each number in the triangle counts the paths through the triangle from the one at the top, down to that number. For example, there are four paths from $C(0,0)$ down to $C(4,3) = 4$, as shown in FIGURE 2. The justification for this is that a path from $C(0,0)$ to $C(n, r)$ consists of n moves, r of them going to the right. The number of ways to choose r right moves from n moves is $C(n, r)$.

Let us return to our task of counting quota systems. We begin with some notation. Let $Q(n, r)$ denote the number of quota systems for n voters such that $q_0 = r + 1$. (The reason for calling this number $Q(n, r)$ instead of $Q(n, r+1)$ will become clear soon.) So $Q(n, r)$ corresponds to the number of paths of length n through Table ω that end on the number $r + 1$. For example, Table 3A shows that $Q(3,1) = 2$, $Q(3,2) = 3$, and $Q(3,3) = 1$. Note that, in order for there to be any quota systems, condition (1) requires that

$$\frac{n}{2} - 1 < r \le n;$$

that is, $r + 1$ must be a valid possibility for q_0. Thus we have $Q(n, r) = 0$ for values of r outside this range. Now let $Q(n)$ denote the total number of quota systems for n voters, so that

$$Q(n) = \sum_{r=\left\lfloor \frac{n}{2} \right\rfloor}^{n} Q(n,r).$$

In our example with $n = 3$, we saw that

$$Q(3) = 6 = 2 + 3 + 1 = Q(3,1) + Q(3,2) + Q(3,3).$$

$C(0,0)$
$C(1,0)$ $C(1,1)$
$C(2,0)$ $C(2,1)$ $C(2,2)$
$C(3,0)$ $C(3,1)$ $C(3,2)$ $C(3,3)$
$C(4,0)$ $C(4,1)$ $C(4,2)$ $C(4,3)$ $C(4,4)$
$C(5,0)$ $C(5,1)$ $C(5,2)$ $C(5,3)$ $C(5,4)$ $C(5,5)$

1
1 1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1

FIGURE 1
Pascal's triangle.

1	1	1	1
1 1	1 1	1 1	1 1
1 2 1	1 2 1	1 2 1	1 2 1
1 3 3 1	1 3 3 1	1 3 3 1	1 3 3 1
1 4 6 4 1	1 4 6 4 1	1 4 6 4 1	1 4 4 4 1

FIGURE 2
There are $C(4,3)$ paths.

We have noted that every path of length $n \geq 1$ extends a path of length $n - 1$. Condition (2) tells us that there are at most two such extensions for each $(n - 1)$-path, and in fact there will be exactly two, so long as condition (1) does not interfere. This leads us to the following identity, which we shall call the *quota rule*,

$$Q(n, r) = Q(n-1, r-1) + Q(n-1, r),$$

which holds for all n and r. This equation is identical to the Pascal rule above. With this in mind we construct a diagram of quota numbers, which we call the *corrupted Pascal's triangle*. See FIGURE 3.

The corrupted Pascal's triangle works exactly like the ordinary Pascal's triangle. The only difference comes from the restriction imposed by condition (1), for when n is even, $Q(n, n/2 - 1) = 0$. These zero values form a 'barricade' absent in the ordinary Pascal's triangle, across which no quota path can pass. The corrupted Pascal's triangle can also be constructed by replacing each entry of Table ω by the number of paths that end on that number. Once we can find an expression for each entry in the corrupted triangle, we can sum across the nth row to get $Q(n)$.

$$\begin{array}{c}
Q(0,0) \\
Q(1,0) \quad Q(1,1) \\
Q(2,0) \quad Q(2,1) \quad Q(2,2) \\
\quad Q(3,1) \quad Q(3,2) \quad Q(3,3) \\
Q(4,1) \quad Q(4,2) \quad Q(4,3) \quad Q(4,4) \\
\quad Q(5,2) \quad Q(5,3) \quad Q(5,4) \quad Q(5,5)
\end{array}
\qquad
\begin{array}{cccccccc}
 & & 1 & & & & & \\
 & 1 & & 1 & & & & \\
\boxed{0} & & 2 & & 1 & & & \\
 & 2 & & 3 & & 1 & & \\
\boxed{0} & & 5 & & 4 & & 1 & \\
 & 5 & & 9 & & 5 & & 1
\end{array}$$

FIGURE 3
Corrupted Pascal's triangle.

6. Counting Quota Systems

We are now in a position to prove our main counting results. We will use the notation $\binom{n}{r}$ for $C(n, r)$.

THEOREM 2. *The number of quota systems for $n \geq 0$ voters with $q_0 = r + 1$ is given by*

$$Q(n, r) = \binom{n+1}{r+1} - \binom{n+1}{r+2}, \quad \textit{for } r > \frac{n}{2} - 1.$$

Proof. We induct on n. For $n = 0$ we have $Q(0,0) = 1 = 1 - 0 = \binom{1}{1} - \binom{1}{2}$. Now assume, as our inductive hypothesis, that

$$Q(n-1, r) = \binom{n}{r+1} - \binom{n}{r+2}, \quad \text{where } r > \frac{n-1}{2} - 1.$$

We consider two cases.

Case I. If $r > (n-1)/2$, then by the quota rule, the induction hypothesis and the Pascal rule, we have

$$\begin{aligned}Q(n,r) &= Q(n-1,r-1)+Q(n-1,r)\\ &= \left[\binom{n}{r}-\binom{n}{r+1}\right]+\left[\binom{n}{r+1}-\binom{n}{r+2}\right]\\ &= \left[\binom{n}{r}+\binom{n}{r+1}\right]-\left[\binom{n}{r+1}+\binom{n}{r+2}\right]\\ &= \binom{n+1}{r+1}-\binom{n+1}{r+2}.\end{aligned}$$

Case II. If n is odd and $r=(n-1)/2$, then the proof above breaks down because the induction hypothesis does not apply to $Q(n-1,(n-1)/2-1)$. But in fact, due to condition (1), $Q(n-1,(n-1)/2-1)=0$, so that

$$Q\left(n,\frac{n-1}{2}\right)=Q\left(n-1,\frac{n-1}{2}\right).$$

To complete this case, we use the Pascal rule in the following form:

$$\binom{n}{r}=\binom{n+1}{r}-\binom{n}{r-1}.$$

Now we have

$$\begin{aligned}Q\left(n,\frac{n-1}{2}\right)=Q\left(n-1,\frac{n-1}{2}\right)&=\binom{n}{\frac{n+1}{2}}-\binom{n}{\frac{n+3}{2}}\\ &=\left[\binom{n+1}{\frac{n+1}{2}}-\binom{n}{\frac{n-1}{2}}\right]-\left[\binom{n+1}{\frac{n+3}{2}}-\binom{n}{\frac{n+1}{2}}\right]\\ &=\binom{n+1}{\frac{n+1}{2}}-\binom{n+1}{\frac{n+3}{2}}.\end{aligned}$$

These two cases exhaust the possibilities for r, and we are done.

With this theorem, $Q(n)$ is easily calculated:

$$Q(n)=\sum_{r=\left\lfloor\frac{n}{2}\right\rfloor}^{n}Q(n,r)=Q(n,n)+\sum_{r=\left\lfloor\frac{n}{2}\right\rfloor}^{n-1}\binom{n+1}{r+1}-\binom{n+1}{r+2}.$$

This sum telescopes—every term cancels except for the first and last. So,

$$\begin{aligned}Q(n)&=Q(n,n)+\binom{n+1}{\left\lfloor\frac{n}{2}\right\rfloor+1}-\binom{n+1}{n+1}\\ &=1+\binom{n+1}{\left\lfloor\frac{n}{2}\right\rfloor+1}-1=\binom{n+1}{\left\lfloor\frac{n}{2}\right\rfloor+1}.\end{aligned}$$

Thus the sequence $\{Q(n)\}_{n=0}^{\infty}$ proceeds $\{1,2,3,6,10,20,35,70,126,\dots\}$ [8].

7. The Catalan Connection

We close with the Catalan numbers. Consider the following problem:

Ballot problem Suppose X and Y are candidates for office and there are $2n$ voters, n voting for X and n voting for Y. How many ways can the ballots be counted so that X is always ahead of or tied with Y?

We begin by examining solutions for small values of n. See FIGURE 4. We represent a vote for X by an arrow moving up, and one for Y by an arrow moving down. In order to satisfy the restriction that X be ahead of or tied with Y, we exclude paths that go below the starting point.

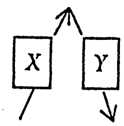
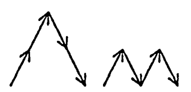
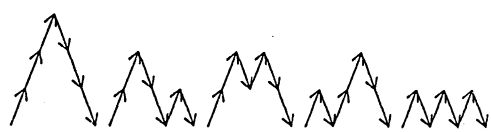
n	Ballot Paths	# of Paths
$n=1$		1
$n=2$		2
$n=3$		5

FIGURE 4
The Ballot Problem.

An elegant solution to the ballot problem involving Andre's reflection method is described in **[4]**, and the problem is generalized there. The numbers in the sequence $\{1, 2, 5, 14, 42, 132, 429, \dots\}$ **[8]**, found in the solution to the ballot problem, are known as the *Catalan numbers*. The equation

$$C_n = \frac{1}{n+1}\binom{2n}{n},$$

where C_n denotes the nth Catalan number, is derived in many combinatorics texts **[1, 9, 11]**. The Catalan sequence appears in the solutions to numerous combinatorial problems **[2, 7]**.

The appearance of the Catalan numbers down the left edge of the corrupted Pascal's triangle leads us to our final result.

THEOREM 3. *$Q(n)$ may be defined recursively as follows:*

$$Q(0) = 1,$$

$$Q(n+1) = \begin{cases} 2Q(n), & \text{if } n \text{ is even}, \\ 2Q(n) - C_{\frac{n+1}{2}}, & \text{if } n \text{ is odd}, \end{cases}$$

where C_k denotes the kth Catalan number. Also, for n odd, the number of quota systems for n voters with $q_0 = (n-1)/2 + 1$ is given by $Q(n, (n-1)/2) = C_{\frac{n+1}{2}}$.

Proof. We first prove the relationship between $Q(n)$ and $Q(n+1)$.

Case I. If n is even, then

$$Q(n+1) = \binom{n+2}{\left\lfloor \frac{n+1}{2} \right\rfloor + 1} = \frac{(n+2)!}{\left(\frac{n+2}{2}\right)!\left(\frac{n+2}{2}\right)!}$$

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

Case II. If n is odd, then

$$Q(n+1) + C_{\frac{n+1}{2}} = \binom{n+2}{\left\lfloor \frac{n+1}{2} \right\rfloor + 1} + \frac{2}{n+3}\binom{n+1}{\frac{n+1}{2}}$$

$$= \frac{(n+2)!}{\left(\frac{n+3}{2}\right)!\left(\frac{n+1}{2}\right)!} + \frac{2}{n+3}\frac{(n+1)!}{\left(\frac{n+1}{2}\right)!\left(\frac{n+1}{2}\right)!}$$

$$= \frac{2(n+2)+2}{n+3}\frac{(n+1)!}{\left(\frac{n+1}{2}\right)!\left(\frac{n+1}{2}\right)!}$$

$$= 2\binom{n+1}{\left\lfloor \frac{n}{2} \right\rfloor + 1} = 2Q(n).$$

Finally, note that when n is odd, we have

$$Q\left(n, \frac{n-1}{2}\right) = \binom{n+1}{\frac{n+1}{2}} - \binom{n+1}{\frac{n+3}{2}}$$

$$= \frac{(n+1)!}{\left(\frac{n+1}{2}\right)!\left(\frac{n+1}{2}\right)!} - \frac{(n+1)!}{\left(\frac{n+3}{2}\right)!\left(\frac{n-1}{2}\right)!}$$

$$= \left(1 - \frac{n+1}{n+3}\right)\frac{(n+1)!}{\left(\frac{n+1}{2}\right)!\left(\frac{n+1}{2}\right)!} = \frac{2}{n+3}\binom{n+1}{\frac{n+1}{2}} = C_{\frac{n+1}{2}}.$$

Consider this theorem in terms of our earlier comments about quota paths. When n is even, each n-path can be extended in two ways, and thus $Q(n+1) = 2Q(n)$. But when n is odd, any path with $r = (n-1)/2$ can be extended in only one way: Condition (1) eliminates one of the possible extensions.

We have shown that the number of such paths is $C_{\frac{n+1}{2}}$. After the previous discussion of the Catalan numbers, we expect a correspondence between these quota paths and ballot paths. This correspondence is illustrated in FIGURE 5.

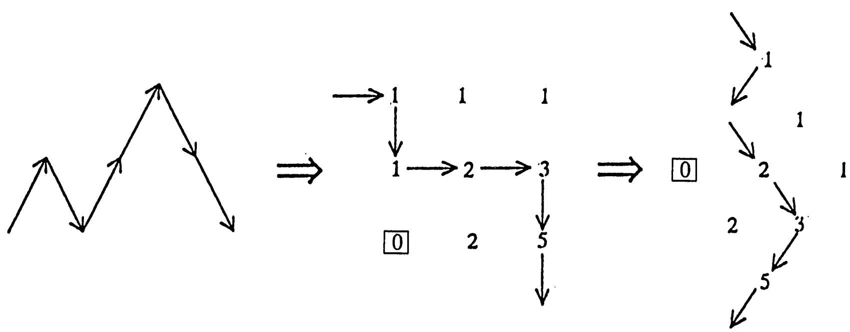


FIGURE 5
The Catalan connection between ballot paths and quota paths.

Acknowledgment. The authors would like to thank the editor and referees for their suggestions.

REFERENCES

1. R. Brualdi, *Introductory Combinatorics*, 2nd edition, Elsevier Science Publishing Co., Inc., New York, 1991.
2. M. Gardner, Mathematical Games, Catalan numbers: an integer sequence that materializes in unexpected places, *Scientific Amer.* 234, (June 1976), 120–125.
3. H. W. Gould, *Research Bibliography of Two Special Number Sequences*, Combinatorial Research Institute, 1977.
4. P. Hilton and J. Pederson, Catalan numbers, their generalization, and their uses, *Math. Int.* 13, (Spring 1991), 64–75.
5. J. Kelly, *Social Choice Theory*, Springer-Verlag, New York, 1988.
6. K. May, A set of independent necessary and sufficient conditions for simple majority decision, *Econometrica* 20 (1952), 680–684.
7. D. Singmaster, An elementary evaluation of the Catalan numbers, *Amer. Math. Monthly* 85 (1978), 366–368.
8. N. J. A. Sloane, *A Handbook of Integer Sequences*, Academic Press, 1973, sequences 294 and 577, and pp. 18–20.
9. D. Stanton and D. White, *Constructive Combinatorics*, Springer-Verlag, New York, 1986.
10. A. Taylor, *Mathematics and Politics: Strategy, Voting, Power and Proof*, Springer-Verlag, New York, 1995.
11. A. Tucker, *Applied Combinatorics*, 2nd edition, John Wiley & Sons, Inc., New York, 1984.

Generating Solutions to the N-Queens Problem Using 2-Circulants

CENGIZ ERBAS
TRW Financial Systems
Oakland, CA 95612

MURAT M. TANIK
Department of Electrical and Computer Engineering
University of Texas at Austin
Austin, TX 78712-1084

Introduction

The N-Queens problem (originally the 8-Queens problem) has been studied for more than a century [**3, 10, 11, 13**]. The problem attracted the attention of several famous mathematicians, including Gauss [**10, 13**] and Pólya [**14**]. During the last three decades, the problem has been discussed in the context of computer science and used as an example of backtracking algorithms, permutation generation, the divide and conquer paradigm, program development methodology, constraint satisfaction problems, integer programming, and specification [**7**]. Some practical applications, such as parallel memory storage schemes, VLSI testing, traffic control, and deadlock prevention are also mentioned in the literature [**7, 8**].

The N-Queens problem is to place N mutually nonattacking queens on an $N \times N$ chessboard. In other words, the objective is to place N queens on an $N \times N$ chessboard in such a way that no two are on the same row, column or diagonal. Example board configurations for 2-, 3-, and 4-Queens problems are given in Figure 1. As can be observed, there is no solution for the 2-Queens and the 3-Queens problem. However, there is a solution for the 4-Queens problem. In this configuration, none of the queens are on the same row, column, or diagonal.

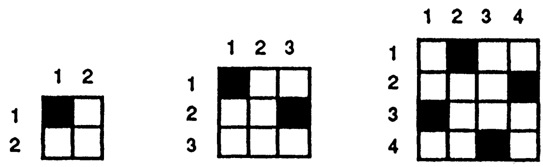


FIGURE 1
Example board configurations for 2-, 3-, and 4-Queens problem.

The literature describes methods for generating a solution for every N [**1, 6, 9, 12, 15**]. Each paper follows a different strategy for placing the queens on the board. In an earlier paper [**6**], we introduced a new method using linear congruence equations to generate a set of solutions for every N. While all the previous approaches generated a constant number of solutions for each N, our approach produces a progressively increasing number of solutions as N increases [**6**].

In this paper, we have developed an approach to the problem based on circulants. Earlier, Clark [**4**] represented Hoffman constructions [**12**] with circulant matrices. However, the method presented here generates a superset of Hoffman's solutions, and is more general than Clark's work.

2-Circulants, Latin Squares, and the N-Queens problem

Let A be an $M \times N$ matrix whose entries are chosen from among the integers $1, 2, 3, \ldots, N$. If each of these integers occurs at most once in each row and column of A, then A is called a *latin rectangle* of order $M \times N$ [2]. If A is a latin rectangle of order $N \times N$, it is called a *latin square* of order N.

A *circulant matrix* [5] of order N is an $N \times N$ matrix of the form given in FIGURE 2(a). Circulant matrices are examples of latin squares [5]. In a circulant matrix, all the rows (except the first one) are generated by rotating the previous row one position to the right. Therefore, the first row determines the whole matrix. The notation $C = \text{circ}(c_1, c_2, c_3, \ldots, c_N)$ represents a circulant matrix.

Following Davis [5], a *g-circulant* of order N is defined as an $N \times N$ matrix in which each row is generated by rotating the preceding row g places to the right. The notation $G = g - \text{circ}(c_1, c_2, c_3, \ldots, c_N)$ represents a g-circulant. A 2-circulant of order N is shown in FIGURE 2(b).

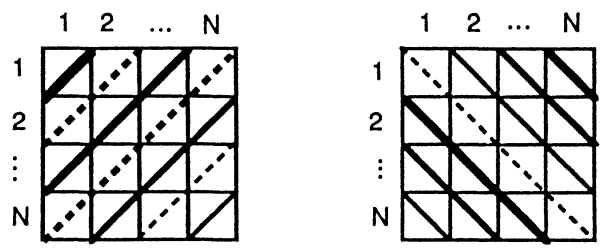
$$\begin{bmatrix} c_1 & c_2 & c_3 & \cdots & c_N \\ c_N & c_1 & c_2 & \cdots & c_{N-1} \\ c_{N-1} & c_N & c_1 & \cdots & c_{N-2} \\ & \cdots & & \cdots & \\ c_2 & c_3 & c_4 & \cdots & c_1 \end{bmatrix} \qquad \begin{bmatrix} c_1 & c_2 & c_3 & \cdots & c_N \\ c_{N-1} & c_N & c_1 & \cdots & c_{N-2} \\ c_{N-3} & c_{N-2} & c_{N-1} & \cdots & c_{N-4} \\ & \cdots & & \cdots & \\ c_3 & c_4 & c_5 & \cdots & c_2 \end{bmatrix}$$

(a) (b)

FIGURE 2

A circulant matrix, and a 2-circulant of order N.

An $N \times N$ square matrix has $2N - 1$ *negative diagonals* (FIGURE 3(a)). For any k, $1 - N \leq k \leq N - 1$, the elements in the k-th negative diagonal are the matrix elements x_{ij}, where $k = i - j$. The 0-th negative diagonal is called the *principal diagonal.* Similarly, an $N \times N$ square matrix has $2N - 1$ *positive diagonals* (FIGURE 3(b)). For any k, $2 \leq k \leq 2N$, the elements in the k-th positive diagonal are the matrix elements x_{ij} whose subscripts add to k, that is $k = i + j$.



(a) Positive diagonals (b) Negative diagonals

FIGURE 3

Negative and positive diagonals of an $N \times N$ chessboard.

LEMMA 1. *Let* $G = g - \text{circ}(c_1, c_2, c_3, \ldots, c_N)$, *where* g *and* N *are positive integers, and* $M = N/[\gcd(N, g)]$.

(a) *If* $\gcd(N, g) = 1$, *then* G *is a latin square.*

(b) *If* $h = \gcd(N, g) \neq 1$, *then* G *is composed of* h *identical latin rectangles of order* $M \times N$.

Proof. Let c_j (where $1 \le j \le N$) be an entry of $G = g - \text{circ}(c_1, c_2, c_3, \dots, c_N)$. For each row $1 \le i \le N$, the column index of c_j is given by $g(i-1) + j \pmod N$.

(a) If $\gcd(N, g) = 1$, then all the column indices $c_j = g(i-1) + j \pmod N$, are distinct for $1 \le j \le N$.

(b) If $h = \gcd(N, g) \neq 1$, then for every row i,

$$i + \frac{N}{g},\ i + 2\frac{N}{g},\ \dots,\ i + (h-1)\frac{N}{g},$$

the column indices are identical. Furthermore, for every consecutive M rows, the column indices are distinct. Therefore, G can be partitioned into h latin rectangles.

In the rest of this paper, we introduce a method that generates a set of solutions to the N-Queens problem using latin squares. The latin squares used are constructed from two latin rectangles of order $(N/2) \times N$, each of which is obtained from a 2-circulant. For ease of comparison with Hoffman's solutions, we name our corresponding constructions Construction A, B, C_1, and C_2. We consider the class of (even) integers $N \ge 4$, $N = 6\alpha - 2$, $N = 6\alpha$, and $N = 6\alpha + 2$, where $\alpha = 1, 2, 3, \dots$. The solutions for odd integers $N \ge 5$, $N = 6\alpha - 1$, $N = 6\alpha + 3$, and $N = 6\alpha + 1$, are generated using the solutions of the even integers.

Generating Solutions for Even N, $N \ge 4$

In this section, we construct solutions to the N-Queens problem for $N = 6\alpha - 2$, $N = 6\alpha$, and $N = 6\alpha + 2$, where $\alpha = 1, 2, 3, \dots$. Construction A generates solutions for $N = 6\alpha - 2$ and $N = 6\alpha$, and Construction B for $N = 6\alpha - 2$ and $N = 6\alpha + 2$. Both Construction A and Construction B generate solutions for $N = 6\alpha - 2$.

Construction A ($N = 6\alpha - 2$, $N = 6\alpha$ for $\alpha = 1, 2, 3, \dots$) Construction A is described in four steps. The first three steps construct a latin square, and the fourth step (Theorem 1) generates a solution to the N-Queens problem from this latin square.

Step 1. Construct the 2-circulant $G_1(N) = 2 - \text{circ}(1, 2, 3, \dots, N)$.

Since $\gcd(N, 2) = 2 \neq 1$, $G_1(N)$ does not form a latin square (Lemma 1). It consists of two copies of a latin rectangle, $L_1(N)$, of order $(N/2) \times N$. See FIGURE 4 for an example.

$$G_1(N) = \left[\begin{array}{c} L_1(N) \\ \hdashline L_1(N) \end{array}\right] = \left[\begin{array}{cccccccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\
11 & 12 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
9 & 10 & 11 & 12 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
7 & 8 & 9 & 10 & 11 & 12 & 1 & 2 & 3 & 4 & 5 & 6 \\
5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 1 & 2 & 3 & 4 \\
3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 1 & 2 \\
\hdashline
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\
11 & 12 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
9 & 10 & 11 & 12 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
7 & 8 & 9 & 10 & 11 & 12 & 1 & 2 & 3 & 4 & 5 & 6 \\
5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 1 & 2 & 3 & 4 \\
3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 1 & 2
\end{array}\right]$$

FIGURE 4

$G_1(N)$, the latin rectangle $L_1(N)$, and an example for $N = 12$.

Step 2. Construct $G_2(N) = 2 - \mathrm{circ}(2, 3, 4, \ldots, N, 1)$.

Since N and 2 are not relatively prime, $G_2(N)$ does not provide a latin square either. Similar to Step 1, if $G_2(N)$ is partitioned into two submatrices of order $(N/2) \times N$, each of these submatrices corresponds to a latin rectangle, called $L_2(N)$. See FIGURE 5 for an example.

$$G_2(N) = \left[\begin{array}{c} L_2(N) \\ \hdashline L_2(N) \end{array}\right] = \left[\begin{array}{cccccccccccc} 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 1 \\ 12 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\ 10 & 11 & 12 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 8 & 9 & 10 & 11 & 12 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 6 & 7 & 8 & 9 & 10 & 11 & 12 & 1 & 2 & 3 & 4 & 5 \\ 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 1 & 2 & 3 \\ \hdashline 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 1 \\ 12 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\ 10 & 11 & 12 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 8 & 9 & 10 & 11 & 12 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 6 & 7 & 8 & 9 & 10 & 11 & 12 & 1 & 2 & 3 & 4 & 5 \\ 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 1 & 2 & 3 \end{array}\right]$$

FIGURE 5

$G_2(N)$, the latin rectangle $L_2(N)$, and an example for $N = 12$.

Step 3. Construct a latin square, $L(N)$, of order N by placing $L_2(N)$ under $L_1(N)$ as shown in FIGURE 6.

$$L(N) = \left[\begin{array}{c} L_1(N) \\ \hdashline L_2(N) \end{array}\right] = \left[\begin{array}{cccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ 11 & 12 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 9 & 10 & 11 & 12 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 7 & 8 & 9 & 10 & 11 & 12 & 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 1 & 2 & 3 & 4 \\ 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 1 & 2 \\ \hdashline 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 1 \\ 12 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\ 10 & 11 & 12 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 8 & 9 & 10 & 11 & 12 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 6 & 7 & 8 & 9 & 10 & 11 & 12 & 1 & 2 & 3 & 4 & 5 \\ 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 1 & 2 & 3 \end{array}\right]$$

FIGURE 6

Constructing the latin square $L(N)$ using the latin rectangles $L_1(N)$ and $L_2(N)$, and an example for $N = 12$.

In order to see that $L(N)$ is a latin square, it is sufficient to demonstrate that each integer $1, 2, 3, \ldots, N$ appears only once at each column of $L(N)$. Consider an integer x such that $1 \leq x \leq N$. Since $L_1(N)$ is a 2-circulant, if x is an even (odd) integer, it appears on the even (odd) numbered columns of $L_1(N)$: $x, x+2, x+4, \ldots, x-2$. $L_2(N)$ is obtained by shifting each row of $L_1(N)$ one position to the left. Thus, if x is an even (odd) integer, it appears on the odd (even) numbered columns of $L_2(N)$: $x-1, x+1, x+3, \ldots, x-3$.

THEOREM 1. *Let $N \geq 4$ be an integer of the form $N = 6\alpha$ or $N = 6\alpha - 2$, where $\alpha = 1, 2, 3, \ldots$. If the latin square $L(N)$ is superimposed onto an $N \times N$ chessboard, and queens are placed on the squares that contain the integer 2, then, the resultant board configuration corresponds to a solution for the N-Queens problem.*

Proof. The positions of the queens (the coordinates of 2's) on the board are

$$(i, 2i), \qquad \text{where } i = 1, 2, 3, \dots, \frac{N}{2} \quad (\text{upper half, } L_1(N)), \text{ and}$$

$$\left(\frac{N}{2} + i, 2i - 1\right) \quad \text{where } i = 1, 2, 3, \dots, \frac{N}{2} \quad (\text{lower half, } L_2(N)).$$

Since $L(N)$ is a latin square, it suffices to check that the positive and the negative diagonals contain at most one queen.

By definition, the positive diagonal number k of the square (i, j) is given by $k = i + j$. Hence, the positive diagonal number of each queen is given by:

$$k = 3i, \qquad \text{where } i = 1, 2, 3, \dots, \frac{N}{2} \quad (\text{upper half of the board}), \text{ and}$$

$$k = 3i + \frac{N}{2} - 1, \quad \text{where } i = 1, 2, 3, \dots, \frac{N}{2} \quad (\text{lower half of the board}).$$

For the upper half of the board $k = 3i \equiv 0 \pmod 3$. If $N = 6\alpha - 2$ then for the lower half of the board $k = 3(\alpha + i) - 2 \equiv 1 \pmod 3$. If $N = 6\alpha$ then for the lower half of the board $k = 3(\alpha + i) - 1 \equiv 2 \pmod 3$. In both cases, no two queens correspond to the same positive diagonal.

Similarly, the negative diagonal number k of the square (i, j) is given by $k = i - j$, and it can be verified that no two queens are placed on the same negative diagonal.

Note that if in Theorem 1 queens are placed on the squares containing the integer x, $1 \leq x \leq N$ and $x \neq 2$, the positions of the queens on the row $N/2$, and $(N/2) + 1$ are given by $(N/2, x - 2)$ and $((N/2) + 1, x - 1)$, respectively. Both of these queens lie on the same negative diagonal.

As an example, for the case $N = 6\alpha$, the solution generated with latin square $L(12)$ corresponds to the board configuration given in FIGURE 7.

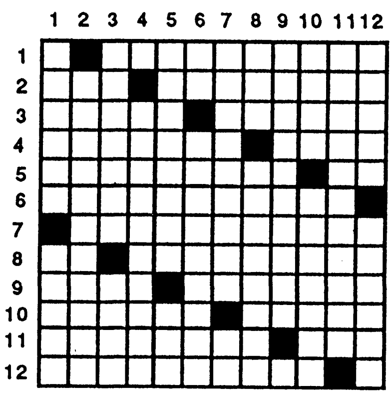


FIGURE 7
A solution for the 12-Queens problem generated by latin square $L(12)$.

Construction B ($N = 6\alpha - 2$, $N = 6\alpha + 2$ **for** $\alpha = 1, 2, 3, \dots$) Let the set $D(N)$ be defined as follows:

$$D(N) = \left\{ d \in Z \mid d = 6i + 3, \text{ where } 0 \leq i \leq \left\lfloor \frac{\alpha - 1}{3} \right\rfloor \right\},$$

where $\lfloor x \rfloor$ denotes the largest integer less than or equal to x. For some α, examples of $D(N)$ are given below:

For $\alpha = 2$, $\quad D(10) = D(14) = \{3\}$.

For $\alpha = 4$, $\quad D(22) = D(26) = \{3, 9\}$.

For $\alpha = 15$, $\quad D(88) = D(92) = \{3, 9, 15, 21\}$.

For every $d \in D(N)$, let us apply the following solution-generating process for the N-Queens problem. Similar to Construction A, the first three steps construct a latin square, and the fourth step (Theorem 2) generates a set of solutions to the N-Queens problem using this latin square.

Step 1. Construct the 2-circulant $G_1(N) = 2\text{-circ}(1, 2, 3, \dots, N)$.

Since N and 2 are not relatively prime, by Lemma 1, $G_1(N)$ consists of two identical latin rectangles $L_1(N)$ of order $(N/2) \times N$.

Step 2. Construct

$$G_3(N, d) = 2 - \text{circ}(N - d + 1, N - d + 2, \dots, N, 1, 2, 3, \dots, N - d).$$

Since N and 2 are not relatively prime, $G_3(N, d)$ does not provide a latin square either. Similar to Step 1, if $G_3(N, d)$ is partitioned into two submatrices of order $(N/2) \times N$, then each of these submatrices corresponds to a latin rectangle, called $L_3(N, d)$.

Step 3. Construct a square matrix $L(N, d)$ of order N by placing $L_3(N, d)$ under $L_1(N)$ as shown in FIGURE 8, for the case $N = 14$, $d = 3$.

$$L(N,d) = \left[\begin{array}{c} L_1(N) \\ \hline L_3(N,d) \end{array}\right]$$

$$= \left[\begin{array}{cccccccccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 \\
13 & 14 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\
11 & 12 & 13 & 14 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
9 & 10 & 11 & 12 & 13 & 14 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 1 & 2 & 3 & 4 & 5 & 6 \\
5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 1 & 2 & 3 & 4 \\
3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 1 & 2 \\
\hline
12 & 13 & 14 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\
10 & 11 & 12 & 13 & 14 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\
8 & 9 & 10 & 11 & 12 & 13 & 14 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\
6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 1 & 2 & 3 & 4 & 5 \\
4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 1 & 2 & 3 \\
2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 1 \\
14 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13
\end{array}\right]$$

FIGURE 8

Constructing the latin square $L(N, d)$ using the latin rectangles $L_1(N)$ and $L_3(N, d)$, and an example for $N = 14$ and $d = 3$.

Then $L(N, d)$ corresponds to a latin square. Consider an integer $1 \le x \le N$. Since $L_1(N)$ is a 2-circulant, if x is an even (odd) integer, it appears on the even (odd) numbered columns of $L_1(N)$: $x, x+2, x+4, \dots, x-2$. $L_3(N, d)$ can be obtained by shifting each row of $L_1(N)$ to the right by d positions. By definition d is an odd integer. Thus, if x is an even (odd) integer, it appears on the odd (even) numbered columns of $L_3(N, d)$: $x + d, x + d + 2, x + d + 4, \dots, x + d - 2$.

THEOREM 2. *Let $N \geq 4$ be an integer of the form $N = 6\alpha - 2$ or $N = 6\alpha + 2$. For every $d \in D(N)$, let the set $S(N, d)$ be defined as follows:*

$$S(N,d) = \{ s \in Z \mid d+1 \leq s \leq N-(2d-2) \}.$$

Let $L(N, d)$ be a latin square constructed by using the steps 1–3 of Construction B. If latin square $L(N, d)$ is superimposed onto an $N \times N$ chessboard, each square contains an integer from 1 to N. If queens are placed onto the squares that contain the integer s, where $s \in S(N, d)$, then, the resultant board configuration corresponds to a solution for the N-Queens problem.

Proof. Since $L(N, d)$ is a latin square, it suffices to check that the positive and the negative diagonals contain at most one queen. We consider these two cases separately.

Let k be a negative diagonal of $L(N, d)$, where $1 - N \leq k \leq N - 1$. Since the upper half of $L(N, d)$ is constructed using a 2-circulant, any two consecutive negative diagonal elements $x_{i,j}$ and $x_{i+1,j+1}$ belonging to the upper half of $L(N, d)$ satisfy the following:

$$x_{i+1,j+1} = x_{i,j} - 1 \pmod{N}.$$

The same argument is valid for the lower half of $L(N, d)$. Furthermore, the boundary elements (elements of the rows $N/2$ and $(N/2) + 1$) satisfy the following:

$$x_{(N/2)+1,j+1} = x_{(N/2),j} - 1 - d \pmod{N}.$$

Therefore, every negative diagonal of $L(N, d)$ provides $N - d$ consecutive distinct integers. The identical elements on negative diagonals form two identical matrices of order $d \times d$ on the upper left and lower right of $L(N, d)$ (FIGURE 9(a)). All the other negative diagonal elements are distinct.

Similarly, let k be a positive diagonal of $L(N, d)$, where $2 \leq k \leq 2N$. Since the upper half of $L(N, d)$ is constructed using a 2-circulant, any two consecutive positive diagonal elements $x_{i,j}$ and $x_{i+1,j-1}$ belonging to the upper half of $L(N, d)$ satisfy the following:

$$x_{i+1,j-1} = x_{i,j} - 3 \pmod{N}.$$

The same argument applies for the lower half of $L(N, d)$. The boundary elements (elements of the rows $N/2$ and $(N/2) + 1$) satisfy the following:

$$x_{(N/2)+1,j} = x_{(N/2),j+1} - 3 - d \pmod{N}.$$

Since N and 3 are relatively prime, every positive diagonal of $L(N, d)$ provides $N - d/3$ consecutive distinct integers. Hence, the identical elements on positive diagonals form two identical matrices of order $(d/3) \times (d/3)$ on the upper right and lower left of $L(N, d)$ (FIGURE 9(b)). All the other positive diagonal elements are distinct. These identical matrices consist of the integers,

$$S'(N,d) = \{ s \in Z \mid 1 \leq s \leq d \text{ and } N-(2d-3) \leq s \leq N \}.$$

The complement of $S'(N, d)$ with respect to the integer set $\{1, 2, \ldots, N\}$ gives the set $S(N, d)$. Thus, for every integer belonging to the set $S(N, d)$, $L(N, d)$ provides a solution to the N-Queens problem.

As an example, let $N = 14$. $D(14)$ contains only one element, namely, $D(14) = \{3\}$. The set $S(14, 3)$ is computed as $S(14, 3) = \{4, 5, 6, 7, 8, 9, 10\}$. The corresponding solutions are given in FIGURE 10.

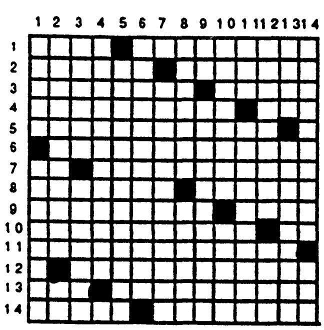
$$\begin{bmatrix}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 \\
13 & 14 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\
11 & 12 & 13 & 14 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
9 & 10 & 11 & 12 & 13 & 14 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 1 & 2 & 3 & 4 & 5 & 6 \\
5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 1 & 2 & 3 & 4 \\
3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 1 & 2 \\
12 & 13 & 14 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\
10 & 11 & 12 & 13 & 14 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\
8 & 9 & 10 & 11 & 12 & 13 & 14 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\
6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 1 & 2 & 3 & 4 & 5 \\
4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 1 & 2 & 3 \\
2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 1 \\
14 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13
\end{bmatrix}$$

(a)

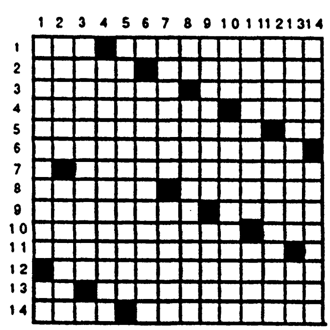
$$\begin{bmatrix}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & \boxed{14} \\
13 & 14 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\
11 & 12 & 13 & 14 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
9 & 10 & 11 & 12 & 13 & 14 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 1 & 2 & 3 & 4 & 5 & 6 \\
5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 1 & 2 & 3 & 4 \\
3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 1 & 2 \\
12 & 13 & 14 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\
10 & 11 & 12 & 13 & 14 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\
8 & 9 & 10 & 11 & 12 & 13 & 14 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\
6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 1 & 2 & 3 & 4 & 5 \\
4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 1 & 2 & 3 \\
2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 1 \\
\boxed{14} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13
\end{bmatrix}$$

(b)

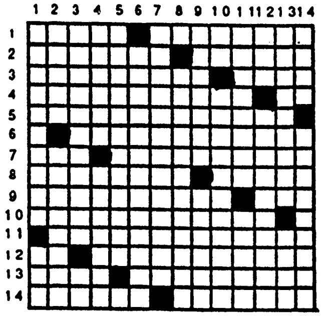
FIGURE 9
Conflicting negative and positive diagonal elements.



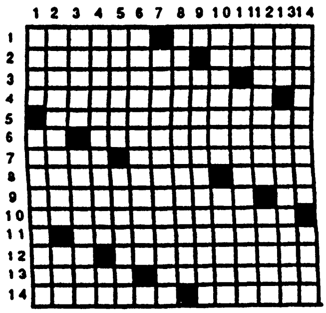
(a)



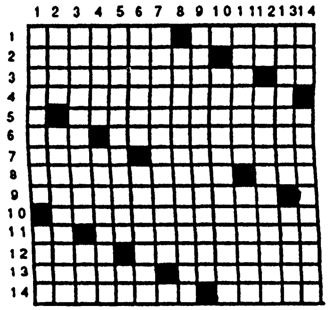
(b)



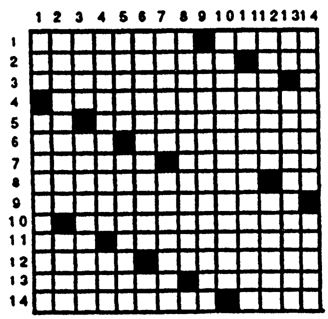
(c)



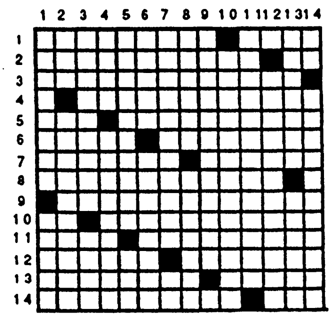
(d)



(e)



(f)



(g)

FIGURE 10
Solutions for the 14-Queens problem generated using latin square $L(14, 3)$.

Generating Solutions for Odd N, $N > 4$

In this section, we utilize the previous solutions to generate solutions for $N = 6\alpha - 1$, $N = 6\alpha + 1$, and $N = 6\alpha + 3$. We introduce two constructions, C_1 and C_2, that use the solutions generated by Construction A and B, respectively. The basic idea is to choose the solutions of the $(N - 1)$-Queens problem that do not place any queen on the principal diagonal, and place a queen on the square (N, N).

Construction C_1 ($N = 6\alpha - 1$, $N = 6\alpha + 1$ for $\alpha = 1, 2, 3, \ldots$) Construction C_1 generates solutions for $N = 6\alpha - 1$ and $N = 6\alpha + 1$ using the solutions generated for $N = 6\alpha - 2$ and $N = 6\alpha$, respectively.

Steps 1–3. Construct the latin square $L(N-1)$ by applying Steps 1–3 of Construction A. As an example, let $N = 13$, then the latin square $L(12)$ is given in FIGURE 6.

Step 4. Superimpose the matrix $L(N-1)$ onto an $(N-1) \times (N-1)$ chessboard. Observe that when the queens are placed onto the squares having the integer 2, the resultant board configuration corresponds to a solution for the $(N-1)$-Queens problem. An example board configuration for 12-Queens is given in FIGURE 7.

THEOREM 3. *Let $N \geq 4$ be an integer of the form $N = 6\alpha - 1$ or $N = 6\alpha + 1$. The solution obtained for the $(N-1)$-Queens problem by superimposing the matrix $L(N-1)$ onto an $(N-1) \times (N-1)$ chessboard, does not have any queen on the principal diagonal. Thus, placing a queen on the square (N, N) generates a solution for the N-Queens problem.*

Proof. Since the queens on the first half of the board correspond to the negative diagonals, $1, 2, 3, \ldots, N/2$, and, the queens on the lower half correspond to $-N/2, (-N/2) + 1, (-N/2) + 2, \ldots, -1$, Construction A does not place any queen on the negative diagonal 0. Thus, a solution for $N = 6\alpha - 1$ and $N = 6\alpha + 1$, is obtained by placing a queen to the square (N, N).

As an example for the case $N = 6\alpha + 1$, a solution for the 13-Queens problem is given in FIGURE 11 (obtained by using FIGURE 7).

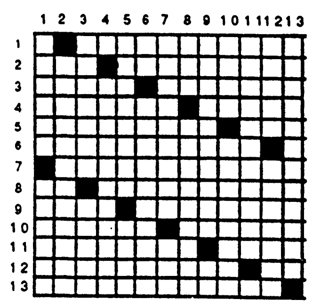


FIGURE 11
A solution for the 13-Queens problem generated from the 12-Queens problem.

An alternate way of obtaining the same solution is:

Step 1. Construct the 2-circulant $L(N) = 2 - \text{circ}(1, 2, 3, \ldots, N)$.

Since N and 2 are relatively prime, $L(N)$ forms a latin square of order N.

Step 2. Superimpose the latin square $L(N)$ onto an $N \times N$ chessboard. Place the queens onto the squares having the integer 2. The resultant board configuration generates the same solution to the N-Queens problem.

In this case, it can be observed that the integer "2" is not the only integer that provides a solution. All i, $1 \leq i \leq N$, can also be used to generate solutions, as demonstrated in Theorem 4.

THEOREM 4. *Let $N \geq 4$ be an integer of the form $N = 6\alpha - 1$ or $N = 6\alpha + 1$, and also let $L(N) = 2 - \text{circ}(1, 2, 3, \ldots, N)$. If the matrix $L(N)$ is superimposed onto an $N \times N$ chessboard, and the queens are placed onto the squares that contain the integer s, where $s \in \{1, 2, 3, \ldots, N\}$, the resultant board configuration provides a solution to the N-Queens problem.*

Proof. Since N and 1 are relatively prime, none of the negative diagonals of $L(N) = 2 - \text{circ}(1, 2, \ldots, N)$ contains a pair of identical integers. Since N and 3 are relatively prime, none of the positive diagonals of $L(N)$ contains a pair of identical integers. So, for all $s \in \{1, 2, 3, \ldots, N\}$, placing queens on squares labeled s in $L(N) = 2 - \text{circ}(1, 2, \ldots, N)$ provides a solution to the N-Queens problem. As an example, the solutions generated using the latin square $L(5) = 2 - \text{circ}(1, 2, 3, 4, 5)$ are illustrated in FIGURE 12.

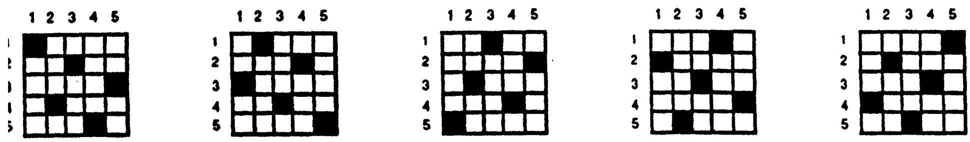


FIGURE 12
Solutions for the 5-Queens problem generated using $2 - \text{circ}(1, 2, 3, 4, 5)$.

Construction C_2 ($N = 6\alpha - 1$, $N = 6\alpha + 3$ for $\alpha = 1, 2, 3, \ldots$) Construction C_2 generates solutions building on Construction B. The solutions of interest are the ones that do not place any queen onto the principal diagonal.

This time, we divide the elements of $D(N-1)$, which is defined in Construction B, into two distinct subsets, and consider each subset separately. Then, we define the set $P(N, d)$ and $Q(N, d)$ for these two subsets, respectively. The sets $P(N, d)$ and $Q(N, d)$ are used to generate solutions to the $(N-1)$-Queens problem that do not place any queen on the principal diagonal. Then, a solution is generated by placing a queen on the square (N, N).

Steps 1–3. Construct the latin square $L(N-1, d)$ by applying the Steps 1–3 of Construction B. As an example, see FIGURE 10.

Step 4. If $N \geq 4d - 1$, then define the set $P(N, d)$, as follows:

$$P(N, d) = \left\{ p \in Z \,\middle|\, \left(\frac{N-1}{2}\right) - d + 2 \leq p \leq \left(\frac{N-1}{2}\right) + 1 \right\}.$$

If $N < 4d - 1$, define the set $Q(N, d)$, as follows:

$$Q(N, d) = \{ q \in Z \mid d + 1 \leq q \leq N - (2d - 1) \}.$$

For example, let $N = 1803$, an integer in the form of $N = 6\alpha + 3$. Then, $\alpha = 300$, and $k = \lfloor \alpha - 1/3 \rfloor = 99$. Therefore, the integer set $D(N) = D(1802) = \{3, 9, 15, \ldots, 596, 602\}$. For every $d \leq 1804/4$, the set $P(1803, d)$, and for every $d > 1804/4$, the set $Q(1803, d)$ are defined.

LEMMA 2. *If the matrix $L(N-1, d)$ is superimposed onto an $(N-1) \times (N-1)$ chessboard, and the queens are placed onto the squares containing the integer s, where*

$$s \in P(N, d), \quad \text{if } N \geq 4d - 1, \quad \text{and}$$
$$s \in Q(N, d), \quad \text{if } N < 4d - 1,$$

then the resultant board configuration corresponds to a solution for the $(N-1)$-Queens problem.

Proof. From Theorem 2, we know that every element of $S(N-1,d)$ provides a solution to the $(N-1)$-Queens problem. Hence, it is sufficient to demonstrate that both $P(N,d)$ and $Q(N,d)$ are subsets of $S(N-1,d)$. The set $S(N-1,d)$ is defined as follows:

$$S(N-1,d)=\{s\in Z\mid d+1\le s\le N-(2d-1)\}.$$

The set $S(N-1,d)$ is identical to the set $Q(N,d)$. In order to demonstrate that $S(N-1,d)\supseteq P(N,d)$, it is sufficient to show that $N-1/2-d+2\ge d+1$, and $N-1/2+1\le N-(2d-1)$. Since, $N\ge 4d-1$, then $N-1/2\ge 2d-1$, and $N-1/2-d+2\ge d+1$. Furthermore, $N-1/2\ge 2d-2+1$, and $N-1/2+1\le N-(2d-1)$. Therefore $S(N-1,d)\supseteq P(N,d)$.

Since both $P(N,d)$ and $Q(N,d)$ are subsets of $S(N-1,d)$, the elements of $P(N,d)$ and $Q(N,d)$ provide solutions to the $(N-1)$-Queens problem. FIGURE 10(c) is an example of one such board configuration.

THEOREM 5. *The solutions generated for the $(N-1)$-Queens problem by Lemma 2 do not place any queen on the principal diagonal. Thus, placing a queen on the square (N,N) generates a solution for the N-Queens problem.*

Proof. The principal diagonal of $L(N-1,d)$ follows the pattern:

$$1,N-1,N-2,N-3,\dots,\frac{N-1}{2}+2,\frac{N-1}{2}-d+1,\frac{N-1}{2}-d,\dots,3,2.$$

In other words, the integers i that appear on the principal diagonal are $1\le i\le (N-1)/2-d+1$, and $(N-1)/2+2\le i\le N-1$. Hence, the integers, p, that do not appear on the principal diagonal are given by $(N-1)/2-d+2\le p\le (N-1)/2+1$.

On the other hand, in order to be a solution to the $(N-1)$-Queens problem, the integer, q, should also be a member of the set $S(N-1,d)$; that is,

$$d+1\le q\le N-(2d-1).$$

Therefore, the intersection of the sets $P(N,d)$ and $S(N-1,d)$ provides solutions that do not place any queen on the principal diagonal. The intersection set is given by

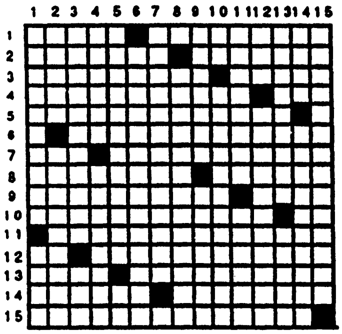
$$\begin{aligned}&P(N,d), && \text{if } N\ge 4d-1, \text{ and}\\ &Q(N,d), && \text{if } N<4d-1.\end{aligned}$$

As an example, the solutions given in FIGURES 10(c), 10(d), and 10(e) do not have any queen on the principal diagonal. The solutions for the 15-Queens problem are obtained by placing a queen on the square (15, 15) (see FIGURE 13).

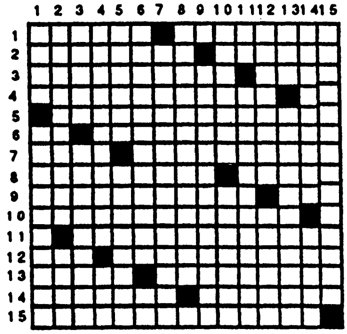
Summary of Results

We have introduced a new method that generates solutions for the N-Queens problem by using 2-circulants and latin squares. The latin squares used in this method are constructed from two latin rectangles each of which is obtained using 2-circulants.

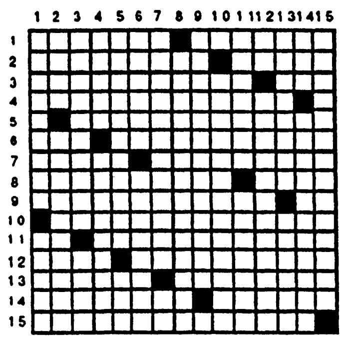
Construction A generates only one solution for a given N, where $N=6\alpha-2$ or $N=6\alpha$. *Construction B* constructs a latin square $L(N,d)$ for every $d\in D(N)$. $D(N)$ consists of $k=\lfloor(\alpha-1)/3\rfloor$ elements. For every $d\in D(N)$, we have defined the set $S(N,d)$ and demonstrated that each $s\in S(N,d)$ provides a distinct solution. The



(a)



(b)



(c)

FIGURE 13

Solutions for the 15-Queens problem generated from the 14-Queens solutions.

cardinality of the set $S(N,d)$ is given by $|S(N,d)| = N - 3d + 2$. Therefore, the number of solutions generated by Construction B can be calculated as follows:

$$\sum_{\forall d \in D(N)} (N - 3d + 2) = \sum_{i=0}^{k} (N - 3(6i+3) + 2)$$

$$= \sum_{i=0}^{k} (N - 7 - 18i) = (k+1)(N - 9k - 7).$$

Construction C_1 generates solutions for $N = 6\alpha - 1$ and $N = 6\alpha + 1$. It constructs a 2-circulant and provides a solution for each integer $1 \le i \le N$. Thus, Construction C_1 generates N distinct solutions for a given N. *Construction* C_2 applies for $N = 6\alpha - 1$ and $N = 6\alpha + 3$. First, it generates a solution for the $(N-1)$-Queens problem using Construction B. In order to generate a solution for the $(N-1)$-Queens problem, we constructed a latin square $L(N-1,d)$ for every $d \in D(N-1)$. Then, we defined the set $P(N,d)$ for every $d \le (N+1)/4$ and $Q(N,d)$ for every $d > (N+1)/4$. The cardinality of these sets is given by $|P(N,d)| = d$ and $|Q(N,d)| = N - 3d + 1$. Therefore, the number of elements generated by Construction C_2 can be calculated as follows:

$$\sum_{i=0}^{h}(6i+3)+\sum_{i=h+1}^{k}(N-3(6i+3)+1)$$
$$=\sum_{i=0}^{h}6i+3(h+1)-\sum_{i=h+1}^{k}18i+(N-8)(k-h)$$
$$=12h(h+1)-9k(k+1)+(11-N)h+(N-8)k+3.$$

Acknowledgment. Thanks to P. J. Campbell for bringing his paper [3] to our attention. Thanks to the editor and anonymous referees for their suggestions to improve the exposition of this paper.

REFERENCES

1. W. Ahrens, *Mathematische Unterhaltungen und Spiele*, Vol. 1, Teubner, 1921.
2. R. A. Brualdi, *Introductory Combinatorics*, North Holland, 1979, pp. 205–213.
3. P. J. Campbell, Gauss and the 8-queens problem: A study in miniature of the propagation of historical error, *Historia Mathematica*, 4 (1977) 397–404.
4. D. S. Clark, A combinatorial theorem on circulant matrices, *Amer. Math. Monthly* (1985) 725–729.
5. P. J. Davis, *Circulant Matrices*, John Wiley & Sons, Inc., New York, 1979.
6. C. Erbas, M. M. Tanik and Z. Aliyazicioglu, Linear congruence equations for the solutions of the *N*-Queens problem, *Inform. Process. Lett.* 41 (1992) 301–306.
7. C. Erbas, S. Sarkeshik, and M. M. Tanik, Different perspectives of the *N*-Queens problem, *ACM, Computer Science Conference '92*, Kansas City, MO, March 2–6, 1992, 99–108.
8. C. Erbas, M. M. Tanik, and V. S. S. Nair, A circulant matrix-based approach to storage schemes for parallel memory systems, *5th IEEE Symposium on Parallel and Distributed Processing*, Dallas, December 1–4, 1993.
9. B. J. Falkowski and L. Schmitz, A note on the queens' problem, *Inform. Process. Lett.* **23** (1986) 39–46.
10. J. Gingsburg, Gauss's arithmetization of the problem of n queens, *Scripta Math.* 5 (1939) 63–66.
11. S. Gunther, Zur mathematisches theorie des Schachbretts, *Archiv der Mathematik und Physik* 56 (1874) 281–292.
12. E. J. Hoffman, J. C. Loessi, and R. C. Moore, Constructions for the m-queens problem, this MAGAZINE, 42(2) (1969) 66–72.
13. *Briefwechsel zwischen C. F. Gauss und H. C. Schumacher*, herausgeg, von Peters, 6. Band, Altona (1865) 105–122.
14. G. Polya, Uber die "doppelt-periodischen" losungen des n-damen-problems, in W. Ahrens, *Mathematische Unterhaltungen und Spiele*, (1918) 364–374.
15. M. Reichling, A simplified solution of the N queens' problem, *Inform. Process. Lett.* 25 (1987) 253–255.

Euclidean Constructions and the Geometry of Origami

ROBERT GERETSCHLÄGER
Bundesrealgymnasium
A-8020 Graz, Austria

1. Introduction

Although the connection between geometry and origami is quite obvious, and has been researched in a limited way for centuries, very few people active in one field appear to be more than casually aware of the other. Occasionally an example in the problems section of some journal will concern itself with folding paper in some way, but origami is not generally considered a mathematical discipline. Some origamians (notably Kazuo Haga and Kunihiko Kashahara, see **[6]**, and Tomoko Fuse, see **[3]**) have done quite impressive work on the geometry of origami (especially on Platonic solids and related subjects), but most prefer to stick to the artistic side of the handicraft. In origami circles, excessive pondering of geometry is usually considered to detract in some way from the elegance and harmony of the art.

Nevertheless, the connection between geometry and origami is well established. Noted educators, such as the German, Friedrich Froebel, have suggested the use of origami as a tool for the teaching of elementary geometric forms. Some interesting work has been published on geometric aspects of origami, particularly as applied to specific models. Much is known about methods of folding regular polygons and polyhedra, for instance **[3, 5, 6, 9]**. It is also well known that the classic Delian problem (finding a cube twice the volume of a given cube) as well as the problem of angle trisection can be solved using methods of origami, despite being unsolvable by Euclidean methods. This article attempts to show the connections between origami folds and Euclidean constructions with straight-edge and compass. Somewhat surprisingly, we shall see that parabolas are elementary to origami construction, and play a role similar to that of circles in Euclidean constructions. There are also some aspects of projective geometry that sneak in unexpectedly.

Perhaps this paper will help in getting more mathematicians interested in origami from the scientific viewpoint. It is my firm belief that this can be fruitful for both the art of origami and the science of geometry.

2. Elementary Euclidean Procedures

When considering Euclidean constructions, it is assumed that specific points are known *a priori* in an infinite Euclidean plane. If needed, random points can be marked in addition to those already known. Using straight-edge and compass as tools, the following procedures are defined as being "allowed":

(E1) Given two non-identical points P and Q, one can draw the unique straight line $l = PQ$ containing both points, using the straight-edge.

(E2) Given a point M and the length of a line segment $r > 0$, one can draw the unique circle $c = \{M; r\}$ with M as center and r as radius, using the compass.

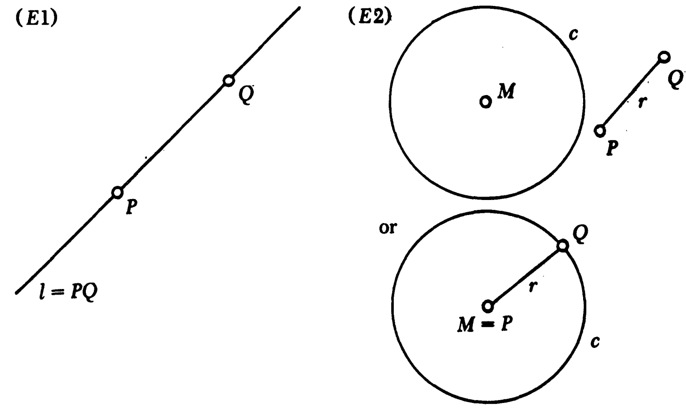


FIGURE 2.1

Specifically, the radius r must be given as the length of a line segment connecting two known points P and Q, one of which may be M, in which case the other is a point on the circle (FIGURE 2.1). Random lines or circles can be introduced, as their generation can always be understood as an application of (E1) and (E2) to random points.

Application of (E1) and (E2) to the points known *a priori* leads to specific straight lines and circles. Knowledge of these leads to further points by virtue of the following procedures of intersection, which are also defined as being "allowed":

(E3) Given two non-parallel straight lines l_1 and l_2, one can determine their unique point of intersection $P = l_1 \cap l_2$;

(E4) given a circle $c = \{M; r\}$ and a straight line l, such that the distance between M and l is not greater than r, one can determine the point(s) of intersection of c and l, and finally

(E5) given two circles $c_1 = \{M_1; r_1\}$ and $c_2 = \{M_2; r_2\}$, such that either
  i) neither contains the center of the other in its interior, and the distance between the centers is not greater than the sum of the radii, or
  ii) one contains the center of the other in its interior, and the distance between the centers is not less than the difference of the radii,
  one can determine the point(s) of intersection of c_1 and c_2.

(In practical applications, the locations of these points of intersection may not be known with sufficient precision. The angle of intersection may be very small, or the relative positions of intersecting straight lines or circles may be otherwise inconvenient, so that the points of intersection may not be practically accessible. They are nevertheless assumed to be known in theory.)

Iterated applications of (E1) to (E5) lead from *a priori* knowledge of specific points to specific straight lines and circles, then to further points, and so on. A geometric construction problem is said to be "solvable" by Euclidean methods, if it can be shown that iterated application exclusively of (E1) to (E5) leads from certain given points to those points and/or straight lines and/or circles that make up whatever geometric entity is sought.

It is the purpose of this paper to compare the solvability of problems in the Euclidean sense with that using elementary procedures of origami, which are defined in the next section.

3. Elementary Geometric Procedures of Origami

Origami is, of course, the art of paper folding. As anyone who has ever put a crease in a piece of paper knows, there are certain procedures in paper folding that seem natural and basic. It is natural to fold a straight line, for instance, whereas folding a curve is possible, but difficult to control. Although there are models in origami utilizing curved folds, for geometric purposes we shall exclude these as non-elementary.

As an origami model develops, its increasing complexity creates an increasingly complex analogous geometric pattern composed of straight lines (or line segments) on the paper used in producing the model. In most origami models, the result is in some way three-dimensional (although folding two-dimensional forms such as regular polygons or stars is also sometimes considered), but every model can be opened up, and what we will consider in this work is the geometry of the folds on the opened paper. Some abstraction will, of course, be necessary. For instance, although most origami is done with squares (or rectangles), and certainly all with finite pieces of paper, for theoretical purposes we will consider folding an infinite Euclidean plane. Also, folding multiple layers of paper leads to some interesting phenomena. For instance, the result of a multi-level fold may be a finite line segment. We will consider this line segment as automatically defining the infinite straight line of which it is a part. This is legitimized by the fact that every line segment has a unique pair of endpoints, which lead to a whole line by folding the infinite paper through these points (a procedure we will define as being allowed). Also, folding multiple layers allows points to be "marked" through the paper. That is, if a specific point comes to lie over another through folding, the multi-layered model can be folded (at least twice) along folds containing that specific point, thus "marking" the points on the other layers of paper immediately above or below that point. It therefore seems natural to assume that a known point, which is brought to lie over another through folding, defines the point on the other layer as equally well known. The same can be said to hold for a line that is brought into another position through folding; it too defines the overlying and underlying lines in the other layers of paper.

If we wish to compare Euclidean procedures to those used in origami, we must define "allowed" procedures, just as we did in the Euclidean case. The basic geometric entity of origami is the straight line. This differs from Euclidean geometry, where the basic entity is the point, knowledge of points then leading to straight lines and circles, and so on. It seems reasonable to assume that a straight line can be folded randomly anywhere on the plane, just as a point can be drawn randomly anywhere in the plane. (We shall use the short form verbs "draw" for Euclidean constructions and "fold" for origami constructions.) Given this, it seems reasonable to first define the following as "allowed":

(O1) Given two non-parallel straight lines l_1 and l_2, one can determine their unique point of intersection $P = l_1 \cap l_2$.

This is, of course, precisely the same as (E3), but in the origami case it defines how points can be considered to be known (which is not immediately obvious), whereas in the Euclidean sense it is *a posteriori* in the sense that points are considered well defined in the plane by simply marking them with a pencil or similar implement.

Two nonparallel straight lines are thus considered to define their point of intersection. This does not yet necessitate any folding other than that leading to the given lines. When two parallel straight lines are given, one can always fold one onto the other uniquely. The resulting fold is that straight line, which is parallel to both, and equidistant from them. When two intersecting straight lines are given, they can be folded onto each other in two ways, the resulting folds being the angle bisectors of the given lines. It therefore seems reasonable to assume the following as "allowed":

(O2) Given two parallel straight lines l_1 and l_2, one can fold the line m parallel to and equidistant from them ("mid-parallel"), and

(O3) Given two intersecting straight lines l_1 and l_2 one can fold their angle bisectors a and a'.

By virtue of the fact that folding a point or a line onto another spot implies knowledge of points and lines in the other layers immediately above or below, these two procedures include the transferring of a known angle to another line. (O2) is the transfer of a line to a parallel line and (O3) that to an intersecting line (FIGURE 3.1). Also, (O3) includes the rotation of a given line-segment, with one end-point in the point of intersection of the two lines, from one line to the other.

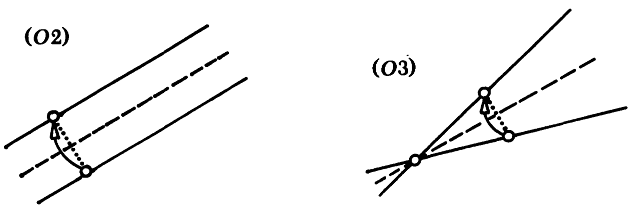


FIGURE 3.1

Given two points on a piece of paper, it is straightforward to fold the unique straight line joining the two points. Equally straightforward, however, is the folding of one given point onto the other. The resulting fold is obviously perpendicular to the line connecting the given points, and these points are equidistant from it. It is therefore the perpendicular bisector of the line segment defined by the two points. Further reasonable "allowed" procedures are therefore:

(O4) Given two non-identical points P and Q, one can fold the unique straight line $l = PQ$ connecting both points, and

(O5) Given two non-identical points P and Q, one can fold the unique perpendicular bisector b of the line segment PQ.

(O4) is, of course, identical to (E1).

As with the rotation of a line segment in (O3), (O5) includes transferring one end-point of a line segment of known length to another point (FIGURE 3.2). Together, (O5) and (O3) mean that one can transfer a line segment of known length to any spot on the paper, as any transfer can be achieved by combining reflections and rotations.

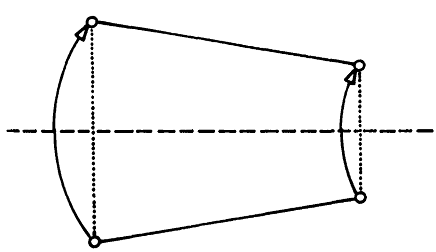


FIGURE 3.2

Given a point P and a straight line l, it is straightforward to fold the line onto itself, such that the given point lies on the fold. Such a fold is obviously unique. Since the given line is folded onto itself, the fold must be perpendicular to it. A further "allowed" procedure is thus:

(O6) Given a point P and a straight line l, one can fold the unique line l' perpendicular to l and containing P.

Finally, given a straight line l and a point P not on the line, it is straightforward to fold P onto any point on the line. The resulting (infinitely many) folds are the elements of the set of perpendicular bisectors of all line segments with the given point P at one end, and a point on the line l at the other. This is precisely the set of tangents of the parabola with P as its focus and l as its directrix (FIGURE 3.3). We see that parabolas, or rather the sets of tangents of parabolas, play an elementary role in the geometry of folding.

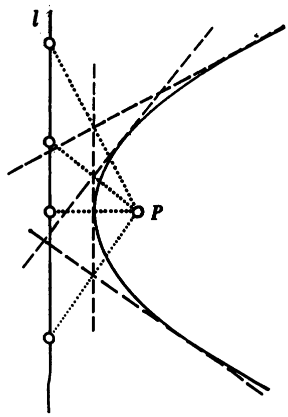


FIGURE 3.3

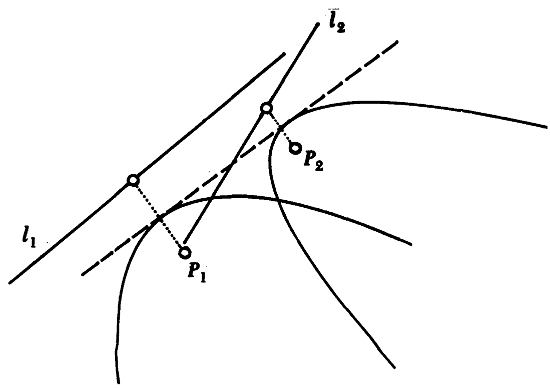


FIGURE 3.4

Given P and l and a farther point Q, it is also straightforward to fold P onto l such that Q lies on the fold. The fold is therefore a tangent of the parabola, containing Q. A further "allowed" procedure is therefore:

(O7) Given a point P and a straight line l, one can fold any tangent of the parabola with focus P and directrix l. Specifically, given a farther point Q, one can fold the tangents of the parabola containing Q.

As with (O3) and (O5), this includes the fact that one can find the points X and $\overline{X}$ on a given line l having the same distance from a given point Q as another given point P. This is slightly more specific than (O3). (O3) merely means that the points of a circle can be found by folding (see 4.2), but this implies that the points of intersection of a line and a circle can be found immediately (see 4.4 and FIGURE 4.2).

Given two points P_1 and P_2 and two lines l_1 and l_2, it is still straightforward to fold P_1 onto l_1 and P_2 onto l_2 with the same fold (FIGURE 3.4). The word "onto" must be understood in a wide sense here, as the points can come to lie above the lines, or the lines above the points, or one of each. In any case, the resulting fold is a common tangent of the two parabolas with foci P_1 and P_2 and directrices l_1 and l_2, respectively.

A final "allowed" procedure is thus:

(O7*) Given (possibly identical) points P_1 and P_2 and (possibly identical) lines l_1 and l_2, one can fold the common tangents of the parabolas p_1 and p_2 with foci P_1 and P_2 and directrices l_1 and l_2, respectively.

This procedure (O7*) is what makes the geometry of origami fundamentally different from Euclidean construction. As we shall see in the next two sections, Euclidean constructions are equivalent to that part of origami constructions utilizing (O1)–(O7). Procedure (O7*), however, allows constructions not accessible by Euclidean methods. The resulting constructions are similar (although not identical) to those utilizing a marked ruler, the theory of which is well established (see **[2]** pp. 74–78). It is not surprising that (O7*) goes beyond Euclidean constructions, if one considers what it means analytically. It is known that two conics in general have four common tangents. If both are parabolas, one of these common tangents is the line at infinity. It is thus a cubic problem to find the common tangents of two parabolas, and it is not to be expected that this cubic problem can be solved by Euclidean methods.

If the foci P_1 and P_2 are identical, the only common tangent is found by folding this point onto the point of intersection of l_1 and l_2. It is not surprising that there is only one further common tangent of p_1 and p_2 to be found, since it is known from projective geometry that the common focus is equivalent to a pair of common complex tangents.

If the directrices are identical, p_1 and p_2 not only have the line at infinity as a common tangent, but also a common point at infinity where the line at infinity is tangent. It therefore counts as a double common tangent, and there are only two further common tangents to be found. These are then, of course, the angle bisectors of the common directrix and the line joining the two foci. In these special cases, finding the common tangents can be solved by Euclidean methods, as the problem is reduced to a linear or quadratic one.

4. Reducing Euclidean Procedures to Origami

In this section, we shall show that each of the elementary Euclidean procedures (E1)–(E5) can be replaced by combinations of (O1)-(O7).

4.1. (E1) (E1) is identical to (O4).

4.2. (E2) A circle cannot be "drawn" by origami procedures. Nevertheless, a circle can still be considered to be well determined if one knows its center M and radius r, as one can determine any number of points and tangents of the circle. This can be achieved in the following manner (FIGURE 4.1):

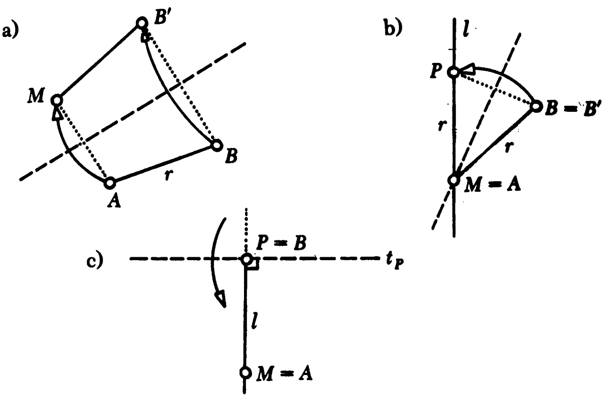


FIGURE 4.1

a) If the center M and radius $r = AB$ of a circle are known, it is possible to fold A to M by virtue of (O5) (folding the perpendicular bisector of MA). This brings B to a point B', and we have $r = MB'$.

b) If a specific line l through M is given, the radius $r = MB'$ can be folded onto it by virtue of (O3) (folding an angle bisector of $\angle MB', l$). This yields the point P as a point on the circle on the diameter l. (The other angle bisector yields the diametrically opposite point of the circle.)

c) Folding l onto itself through P by virtue of (O6) yields the line perpendicular to the diameter containing P, which is precisely the tangent of the circle in P.

4.3. (E3) (E3) is identical to (O1).

4.4. (E4) If a circle is known by its center M and a point P on its circumference, and a line l is given, the points of intersection of the circle and l can be found by folding P onto l such that the fold contains M. This is possible by virtue of (O7). In doing this, finding the points of intersection of a circle and a straight line is seen to be equivalent to finding the tangents of a specific parabola (with focus P and directrix l) containing a specific point (M).

4.5. (E5) Since circles are only accessible in origami through knowledge of specific points and tangents, it is not possible to find the common points of two circles

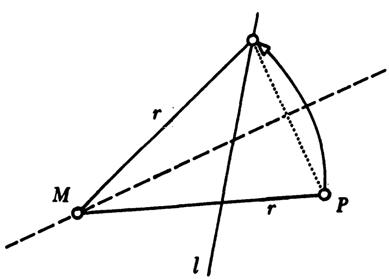


FIGURE 4.2

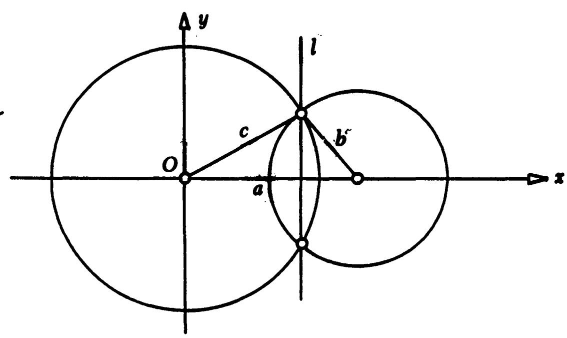


FIGURE 4.3

directly. It is, however, possible to find the common chord of intersecting circles, thus reducing (E5) to (E4). This can be achieved in the following manner.

We assume that two circles, whose points of intersection we wish to determine, are given. Let the distance between their centers be a, and let the radii be b and c, respectively. Assuming the center of one circle is the origin of a cartesian coordinate system, and the center of the other is on the x-axis, their equations are $x^2 + y^2 = c^2$ and $(x-a)^2 + y^2 = b^2$, respectively.

Their common chord is therefore the line represented by the equation

$$x^2 + y^2 - c^2 = x^2 - 2xa + a^2 + y^2 - b^2$$
$$\Leftrightarrow x = \frac{a^2 - b^2 + c^2}{2a}.$$

The common points of the two circles thus lie on the line that is perpendicular to the line connecting their centers, and whose distance to the center of the circle with radius c is $(a^2 - b^2 + c^2)/2a$ (or equivalently, whose distance to the center of the circle with radius b is $(a^2 + b^2 - c^2)/2a$). This line can be found by origami procedures. One way of doing this is described in the following four steps.

Step 1. Since the distances a and c are known, it is possible to fold a right triangle with sides a and c ((O3), (O5), (O6)). The length of the hypotenuse is then $\sqrt{a^2+c^2}$.

Step 2. The distance b and $\sqrt{a^2+c^2}$ are known. It is therefore possible to fold a right triangle with side b and hypotenuse $\sqrt{a^2+c^2}$ ((O3), (O5), (O6), (O7)). The length of the second side is then $\sqrt{a^2-b^2+c^2}$.

Step 3. A triangle can be folded with one side of unit length 1 and one side of length $\sqrt{a^2-b^2+c^2}$. A similar triangle can then be folded ((O2), (O3)) with a side of length $\sqrt{a^2-b^2+c^2}$ corresponding to that side of the first triangle with length 1. The side corresponding to that side of the first triangle with length $\sqrt{a^2-b^2+c^2}$ is then of length $a^2-b^2+c^2$.

Step 4. A triangle can be folded with one side of length $2a$ and another of length $a^2-b^2+c^2$ ((O2), (O3)). A similar triangle can be folded with a side of length 1 corresponding to the side of the first triangle with length $2a$ ((O2), (O3), (O5)). Then the side corresponding to the side of the first triangle of length $a^2-b^2+c^2$ has the length $a^2-b^2+c^2/(2a)$. This is precisely the length we had set out to produce.

(E5) is thus reduced to (E4), as we need only find the points of intersection of the common chord with either circle.

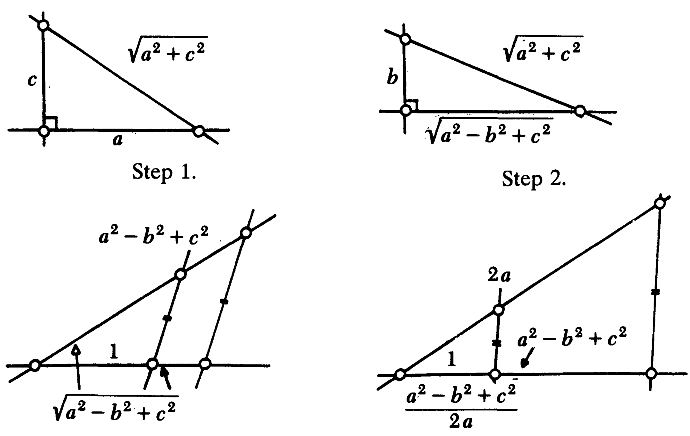


FIGURE 4.4

In summary, we have the following.

THEOREM 1. *Every construction that can be done by Euclidean methods can also be achieved by elementary methods of origami. Specifically, the Euclidean procedures* (E1)–(E5) *can all be replaced by combinations of the origami procedures* (O1)–(O7).

5. Reducing Origami Procedures to Euclidean Constructions

In this section we shall show that each of the elementary origami procedures (O1)–(O7) can be replaced by combinations of (E1)–(E5). That this is not to be expected for (O7*) has already been explained at the end of section 3.

5.1. Identical Procedures (O1) is identical to (E3), and (O4) is identical to (E1).

5.2. The Basic Procedures (O2), (O3), (O5), and (O6) The origami procedures (O2), (O3), (O5), and (O6) are the constructions of the mid-parallel of two parallel lines, the construction of the angle bisectors of two intersecting lines, the construction of the perpendicular bisector of a line segment, and the construction of a line through a given point and perpendicular to a given line, respectively. All of these constructions are known to be possible using Euclidean methods.

5.3. (O7) Determining the tangents of a parabola given by its focus F and directrix l through a given point P by Euclidean methods is not difficult, but the method is perhaps not as well known as those in 5.2. A review of some elementary properties of the parabola seems in order here (FIGURE 5.1). If a specific point T of a parabola with focus F and directrix l is known, we have $TF = Tl$ by definition of a parabola. The axis a of the parabola contains F and is perpendicular to l. The line parallel to a containing T intersects l in a point G. Obviously, $TG = Tl = TF$. If we determine the point X on a such that $FTGX$ is a rhombus, the diagonal TX of the rhombus is the tangent of the parabola in T.

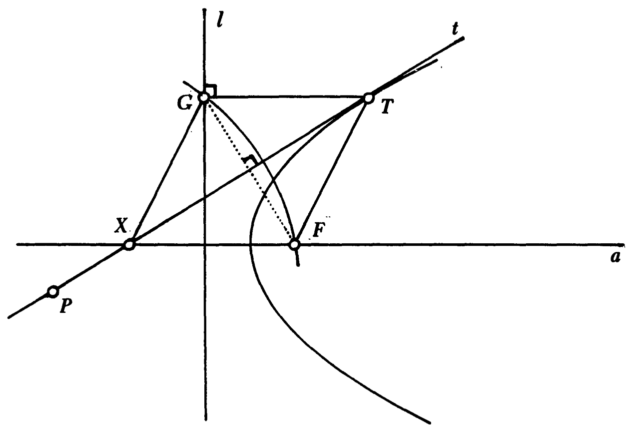


FIGURE 5.1

Knowing this, we see that we can find the tangents of a given parabola containing a given point P by the following method:

Since a tangent t containing P must be the perpendicular bisector of the line segment FG, where G is the point on l corresponding to that point T where t is tangent to the parabola, the distance from P to F must be equal to the distance from P to G. Using the compass, we can find the two possible points G, and can thus find the tangents t along with the points T by completing the rhombi defined by F and the two points G. The diagonals of these rhombi are then the tangents containing P.

We see that (O7) is also replaceable by Euclidean methods. In summary, we therefore have the following.

THEOREM 2. *Every construction that can be done by the origami methods* (O1)–(O7) *exclusively can also be achieved by Euclidean methods.*

Together with Theorem 1 this means that every geometric construction that can be achieved using the origami methods (O1)–(O7) can also be achieved by the Euclidean methods (E1)–(E5) and vice versa. The two sets of constructions are thus equivalent. As we have already seen, (O7*) adds additional geometric constructions to the set of possible constructions generated by these equivalent sets. We see that the set of constructions that can be generated by Euclidean methods is a true subset of the set *that can be generated by origami methods.*

6. Folding Cube Roots

As already mentioned in section 3, finding the common tangents of two parabolas is, in general, analytically a cubic problem. Since (O7*) makes it possible to fold the common tangents of two parabolas, it is to be expected to be possible to find cube roots utilizing (O7*), which is of course, not possible by Euclidean methods. In this section, we shall become acquainted with a simple method of folding the cube root of the quotient of the lengths of any two given line segments.

A method of folding $\sqrt[3]{2}$ is given in [8]. The more general method we shall use here is based on parabolas with a common vertex and perpendicular axes. That such parabolas intersect in points whose coordinates solve simple cubic equations was already known to Descartes, and that such parabolas have something to do with finding cube roots was even known in antiquity (see [4], p. 12, or [1], pp. 342–344). Since folding does not allow us to work with points of intersection, but rather with common tangents, we must deal with the dual problem, which works just as well, as we will see.

We consider the parabolas with the equations

$$p_1\colon y^2 = 2ax \quad \text{and} \quad p_2\colon x^2 = 2by.$$

Since these parabolas intersect in two points with real coordinates, they have only one real common tangent. We assume this tangent (which cannot be parallel to either axis) to be

$$t\colon y = cx + d.$$

We assume further that $P_1(x_1, y_1)$ is the point at which t is tangent to p_1. Then, t also has the equation

$$yy_1 = ax + ax_1 \quad \Leftrightarrow \quad y = \frac{a}{y_1}\cdot x + \frac{ax_1}{y_1}.$$

Therefore

$$\begin{aligned} c = \frac{a}{y_1} &\quad \text{and} \quad d = \frac{ax_1}{y_1} \\ \Rightarrow y_1 = \frac{a}{c} &\quad \text{and} \quad x_1 = \frac{d}{c} \\ \Rightarrow \frac{a^2}{c^2} &= 2a\cdot\frac{d}{c} \\ \Rightarrow a &= 2cd. \end{aligned}$$

We further assume that $P_2(x_2, y_2)$ is the point in which t is tangent to p_2. Then, t also has the equation

$$xx_2 = by + by_2 \quad \Leftrightarrow \quad y = \frac{x_2}{b}\cdot x - y_2.$$

Therefore

$$\begin{aligned} c = \frac{x_2}{b} &\quad \text{and} \quad d = -y_2 \\ \Rightarrow x_2 = bc &\quad \text{and} \quad y_2 = -d \\ \Rightarrow b^2c^2 &= -2bd \\ \Rightarrow d &= -\frac{bc^2}{2}. \end{aligned}$$

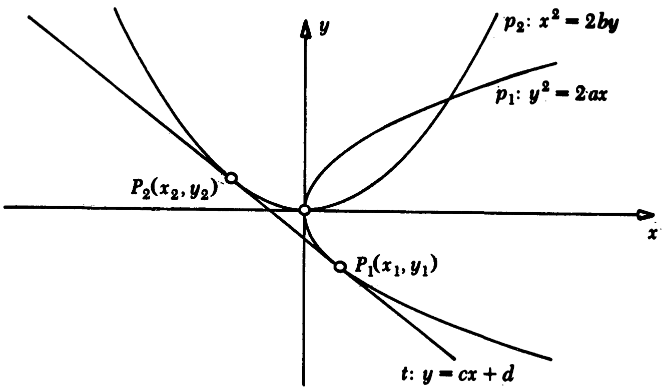


FIGURE 6.1

We therefore see that

$$a = 2cd \quad \text{and} \quad d = -\frac{bc^2}{2}$$

$$\Rightarrow a = -bc^3$$

$$\Rightarrow c = -\sqrt[3]{\frac{a}{b}}.$$

We see that the slope of the common tangent is the (negative) cube root of the quotient of the parameters of the parabolas.

If b is equal to the unit length, the slope of the common tangent is the cube root of the parameter of p_1.

A cube root can therefore be folded in the following manner. If a and b are given, we fold a right angle anywhere to represent the parabola axes. We fold $a/2$ to the left and right of the origin to obtain the directrix l_1 and the focus F_1 of p_1, respectively, and $b/2$ to the top and bottom to similarly obtain the directrix l_2 and the focus F_2 of p_2. (Since $\frac{a}{b} = \frac{2a}{2b}$, we can also use a and b rather than $a/2$ and $b/2$.) Folding F_1 onto l_1 and F_2 onto l_2 simultaneously by virtue of (O7*) gives us the common tangent t, whose slope is then $-\sqrt[3]{a/b}$. Folding the unit length from any point on t parallel to the x-axis and completing the right triangle with hypotenuse on t thus yields a line segment of length $\sqrt[3]{a/b}$ as the second side of the triangle.

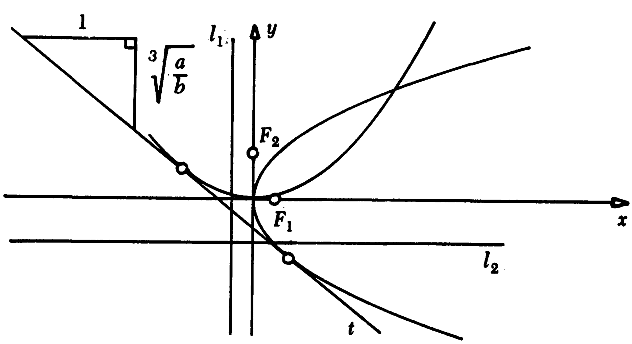


FIGURE 6.2

7. Solving General Cubic Equations

A slight generalization of the method presented in the preceding section allows us to solve general cubic equations. We can see this in the following manner.

We consider the parabolas with the equations

$$p_1\colon (y-n)^2 = 2a(x-m) \quad \text{and} \quad p_2\colon x^2 = 2by.$$

As before, we assume that the equation describing a common tangent of p_1 and p_2 (which need not be unique in this case), is

$$t\colon y = cx + d.$$

Again, such a common tangent cannot be parallel to either axis. We assume, as before, that $P_1(x_1, y_1)$ is the point in which t is tangent to p_1. Then, t is also represented by the equation

$$\begin{aligned} &(y-n)(y_1-n) = a(x-m) + a(x_1-m) \\ \Leftrightarrow\ &y = \frac{a}{y_1-n}\cdot x + n + \frac{ax_1 - 2am}{y_1-n}. \end{aligned}$$

Therefore

$$\begin{aligned} &c = \frac{a}{y_1-n} \quad \text{and} \quad d = n + \frac{ax_1-2am}{y_1-n} \\ \Rightarrow\ &y_1 = \frac{a+nc}{c} \quad \text{and} \quad x_1 = \frac{d-n}{c} + 2m, \end{aligned}$$

and

$$\begin{aligned} &(y_1-n)^2 = 2a(x_1-m) \\ \Rightarrow\ &\frac{a^2}{c^2} = 2a\left(\frac{d-n}{c} + m\right) \\ \Rightarrow\ &a = 2c(d-n+cm). \end{aligned}$$

As in the preceding section, assuming $P_2(x_2, y_2)$ to be the point in which t is tangent to p_2, we find that t is also represented by the equation

$$y = \frac{x_2}{b}\cdot x - y_2,$$

which again leads to

$$d = -\frac{bc^2}{2}.$$

Substituting for d, we obtain

$$\begin{aligned} &a = 2c\left(-\frac{bc^2}{2} - n + cm\right) \\ \Leftrightarrow\ &bc^3 - 2mc^2 + 2nc + a = 0 \\ \Leftrightarrow\ &c^3 - \frac{2m}{b}\cdot c^2 + \frac{2n}{b}\cdot c + \frac{a}{b} = 0. \end{aligned}$$

The slope of the common tangent is therefore a solution c of the cubic equation

$$c^3 - \frac{2m}{b}\cdot c^2 + \frac{2n}{b}\cdot c + \frac{a}{b} = 0.$$

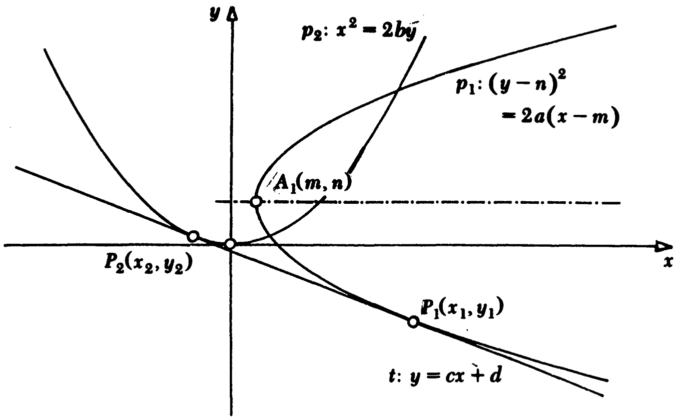


FIGURE 7.1

This equation can have either one real solution and a pair of complex solutions, or three real solutions, of which two or all three can be equal. This corresponds to parabolas that intersect, and parabolas that do not, respectively. Two solutions are equal if, and only if, two tangents are equal (that is, if the parabolas are tangent), and all three are equal if, and only if, the parabolas osculate (that is, if they have contact of third order).

Given a cubic equation, one can therefore fold the roots of the equation by the following method. Say the given equation is

$$x^3 + px^2 + qx + r = 0.$$

Assuming that the parameter b of p_2 is equal to the unit length 1, we have

$$p = -2m, q = 2n \quad \text{and} \quad r = a$$

or

$$m = -\frac{p}{2}, n = \frac{q}{2} \quad \text{and} \quad a = r.$$

We need therefore only find the point with coordinates

$$F_1\left(-\frac{p}{2} + \frac{r}{2}, \frac{q}{2}\right)$$

and the line l_1 represented by the equation

$$x = -\frac{p}{2} - \frac{r}{2}.$$

These are then focus and directrix of p_1, respectively. The focus F_2 of p_2 is $(0, \frac{1}{2})$, and its directrix l_2 is represented by the equation $y = -\frac{1}{2}$. Folding F_1 onto l_1 and F_2 onto l_2 simultaneously by virtue of (O7*) then yields the common tangent (or tangents) of p_1 and p_2, whose slope solves the given cubic equation.

8. Trisecting Angles

If we wish to trisect an angle using origami methods, we find that the preceding result gives us a straightforward method of doing so. It is known that the equation

$$\cos 3\alpha = 4\cos^3 \alpha - 3\cos \alpha$$

holds for any angle α. Assuming knowledge of $\cos 3\alpha$, finding $\cos\alpha$ is therefore merely a matter of solving the cubic equation

$$x^3 - \tfrac{3}{4}\cdot x - \tfrac{1}{4}\cdot\cos 3\alpha = 0.$$

As previously shown, this can be done by utilizing the parabola p_1 with the focus

$$F_1\left(-\tfrac{1}{8}\cos 3\alpha, -\tfrac{3}{8}\right)$$

and the directrix

$$l_1\colon\ x = \tfrac{1}{8}\cos 3\alpha$$

(since a is negative, the parabola is open to the left in this case), as well as the "unit parabola" p_2 with focus $F_2(0, \frac{1}{2})$ and directrix $l_2\colon\ y = -\frac{1}{2}$. The slope of the common tangent of these parabolas solves the cubic equation, yielding $\cos\alpha$, which immediately leads to α itself.

9. Conclusion

The method described in section 6 is by no means the only method of producing cube roots using origami methods, but it is effective, general, and acceptably easy to apply. Specifically, the Delian problem (doubling the volume of a cube) is reduced to a special case, and a fairly easy one at that. Perhaps this is what the oracle at Delos originally had in mind when it asked the Athenians to double the size of the cubic altar of Apollo in order to rid themselves of the plague. Maybe the oracle was really an origamian at heart.

Also, while the method of trisecting angles described in section 8 is not likely to be the most elegant for practical origami purposes, it is an immediate corollary of the solution of the general cubic equation using those origami methods defined as "allowed" in section 3, and therefore quite easy to grasp in this context. Perhaps angle trisectors should turn their attentions to origami in the future. In any case, we can hope that many more interesting results can be derived from the systematic study of the geometry of origami.

REFERENCES

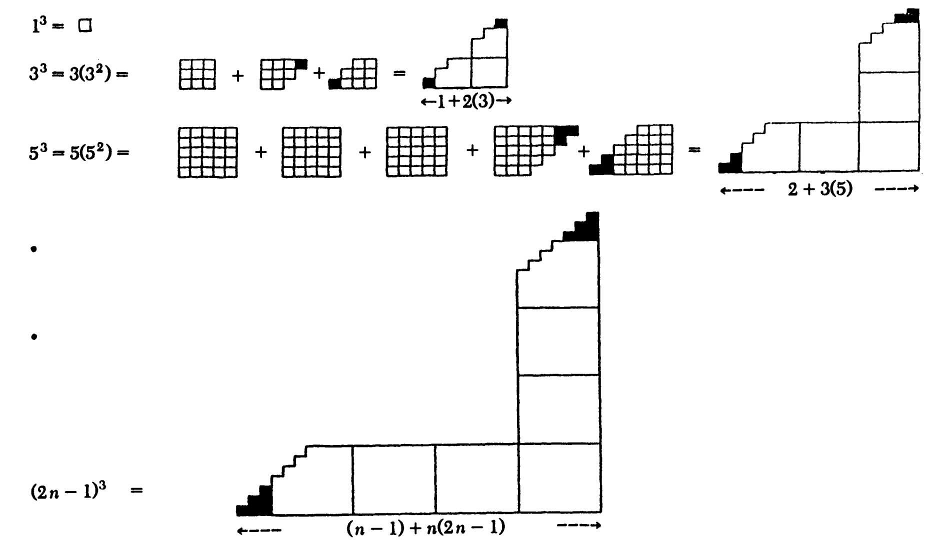
1. W. W. Rouse Ball and H. S. M. Coxeter, *Mathematical Recreations and Essays, 13th Edition*, Dover Publications, Inc., Mineola, NY (1987).
2. Ludwig Bieberbach, *Theorie der Geometrischen Konstruktionen*, Verlag Birkhäuser, Basel (1952).
3. Tomoko Fuse, *Unit Origami*, Japan Publications, Tokyo and New York (1990).
4. Johannes Hjelmslev, *Geometrische Experimente*, B. G. Teubner Verlag, Leipzig (1914).
5. Kunihiko Kasahara, *Origami Omnibus*, Japan Publications, Tokyo (1988).
6. Kunihiko Kasahara, Toshie Takahama, *Origami for the Connoisseur*, Japan Publications, Tokyo (1987).
7. H. Meschkowski, *Unsolved and Unsolvable Problems in Geometry*, Ungar Publications, New York (1966).
8. Peter Messner, *Crux Mathematicorum*, *Problem* 1054, Vol. 12, No. 10, 1986.
9. T. Sundara Row, *Geometric Exercises in Paper Folding*, Dover Publications, Inc., Mineola, NY (1966), reprint of 1905 edition.

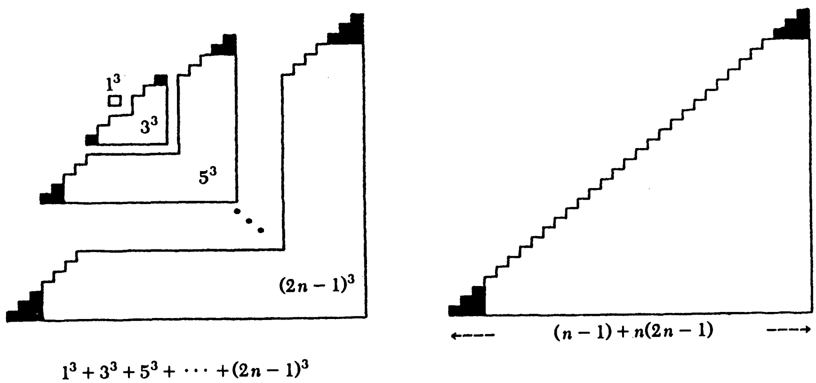
10. Postscript

After this paper was accepted for publication in this MAGAZINE, a paper titled "Reflections on a Mira" by John W. Emert, Kay I. Meeks, and Roger B. Nelsen appeared in the June-July 1994 edition of *The American Mathematical Monthly*. The

Mira is a semi-reflective geometric construction device, which allows constructions whose intrinsic geometry is essentially the same as that defined by origami methods. Some of the results of this paper are therefore quite similar to the results established there. Another interesting paper on the geometry of the mira is "Duplicating the Cube with a Mira" by George E. Martin, *Mathematics Teacher*, March 1979.

Proof without Words: The Sum of Consecutive Odd Cubes is a Triangular Number



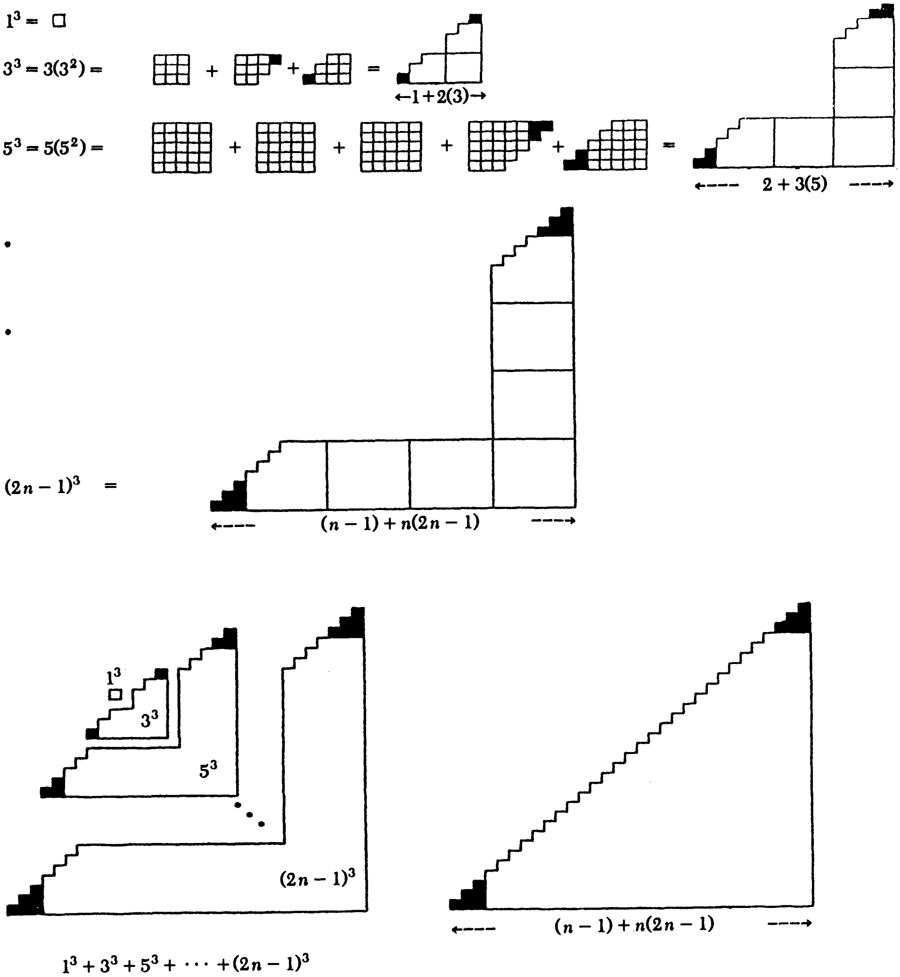


$$1^3 + 3^3 + 5^3 + \cdots + (2n-1)^3 = 1 + 2 + 3 + \cdots + (2n^2 - 1)$$
$$= \frac{(2n^2-1)(2n^2)}{2}$$
$$= n^2(2n^2 - 1)$$

—MONTE J. ZERGER
ADAMS STATE COLLEGE
ALAMOSA, CO 81102

Mira is a semi-reflective geometric construction device, which allows constructions whose intrinsic geometry is essentially the same as that defined by origami methods. Some of the results of this paper are therefore quite similar to the results established there. Another interesting paper on the geometry of the mira is "Duplicating the Cube with a Mira" by George E. Martin, *Mathematics Teacher*, March 1979.

Proof without Words: The Sum of Consecutive Odd Cubes is a Triangular Number



$$
\begin{aligned}
1^3 + 3^3 + 5^3 + \cdots + (2n-1)^3 &= 1 + 2 + 3 + \cdots + (2n^2 - 1) \\
&= \frac{(2n^2-1)(2n^2)}{2} \\
&= n^2(2n^2 - 1)
\end{aligned}
$$

—MONTE J. ZERGER
ADAMS STATE COLLEGE
ALAMOSA, CO 81102

NOTES

Sylow Fractals

BEN BREWSTER
Binghamton University
Binghamton, NY 13902

MICHAEL B. WARD
Bucknell University
Lewisburg, PA 17837

Professor Ben Brewster gave a talk for undergraduates at Bucknell University in which he showed how to think of the Sylow p-subgroups of S_{p^n}, the symmetric group on the numbers 1 through p^n, as groups of symmetries of wild Ferris wheels. While he talked, several in the audience noted a more *trendy* interpretation in terms of fractals. ("Oh, no!" some will groan.) Herein, we will show that *trendy* interpretation and use it as a vehicle for revealing the interesting structure of Sylow p-subgroups of S_{p^n} and for illustrating the group theoretic idea of a wreath product.

Before proceeding, we ought to mention that Brewster's talk was inspired by a remembrance of his colleague, Wolfgang Kappe. Professor Kappe told of watching L. Kaloujnine, who first published the structure of Sylow p-subgroups of S_{p^n} [3], deep in thought, making circles in the air with his fingers. Brewster, guessing what Kaloujnine might have been thinking, made those circles into his Ferris wheels.

For background, the reader need only know some elementary group theory as contained, for instance, in [1]. In particular, recall that if p^k is the highest power of a prime p dividing the order of a finite group G, then a subgroup of G of order p^k is called a Sylow p-subgroup of G. (The Sylow theorems assert, in part, that Sylow p-subgroups exist for every finite group and for every prime.)

Strictly speaking, the geometric figures whose symmetries we'll study are not fractals, but figures on their way to becoming fractals. Let's construct those figures. Start with any prime p and a regular p-gon (a regular 2-gon is a line segment). Call the p-gon P_1. Form P_2 from p smaller, congruent copies of P_1 by positioning one copy in each corner of P_1. "Smaller" means small enough that the p copies so arranged do not overlap, even at vertices. P_3 is constructed in a similar way from p smaller copies of P_2, P_4 from p smaller copies of P_3, and so on. See FIGURE 1 for examples with $p = 2, 3$.

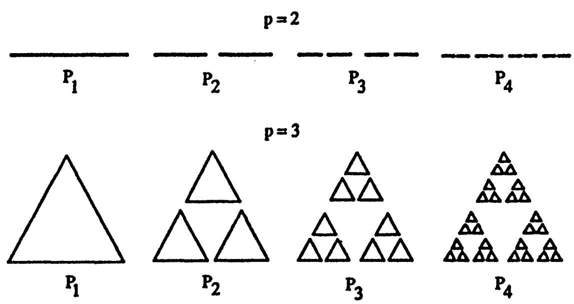


FIGURE 1

We see that in the case $p = 2$, the intersection of P_1, P_2, P_3, and so on is a Cantor set. When $p = 3$, the intersection is a (disconnected) version of the Sierspinski triangle. Both are familiar fractals.

For each P_n, the group of interest to us is the subgroup of the group of symmetries of P_n whose elements are made up of rotations of P_n itself along with rotations of one or more of the copies of P_i, $i < n$, that make up P_n. We'll name that subgroup $RotSym(P_n)$, the group of "rotational symmetries" of P_n. A sample symmetry from $RotSym(P_3)$ when $p = 3$ is illustrated in FIGURE 2. The symmetry results from rotating the topmost copy of P_2 clockwise 120°, rotating by 240° clockwise the lower-right copy of P_1 lying in the leftmost copy of P_2 and then rotating the whole thing clockwise 120°.

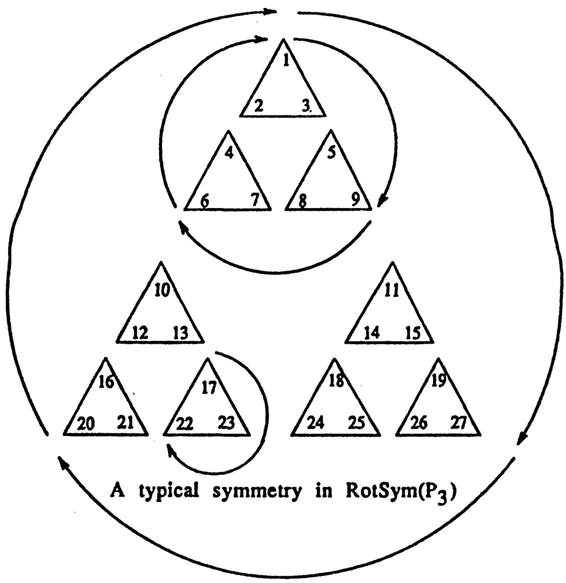
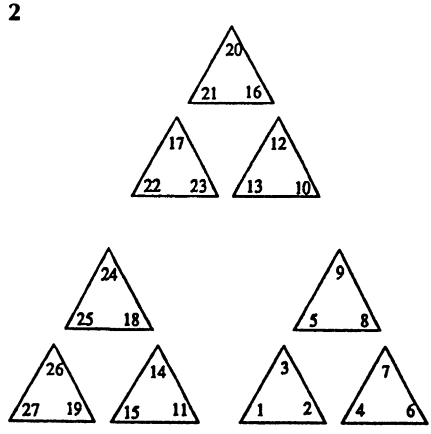


FIGURE 2



Resulting Orientation

Visually, imagine $RotSym(P_n)$ rotating any or all of the copies of P_j, $j \leq n$, contained within P_n, creating a pinwheel of spinning polygons.

By thinking recursively (i.e. relating $RotSym(P_n)$ to $RotSym(P_{n-1})$), the structure of $RotSym(P_n)$ is revealed. For example, consider $RotSym(P_2)$ when $p = 3$. Set R_1 equal to the set of symmetries in $RotSym(P_2)$ that move only the topmost small copy of P_1 inside P_2 while leaving the other two copies of P_1 fixed. Thus, R_1 consists of the symmetries of P_2 obtained by rotating the topmost copy of P_1 by 0°, 120° or 240° clockwise while leaving the bottom two copies of P_1 unchanged. Define R_2 and R_3 in a similar way for the lower-right and lower-left copies of P_1, respectively. R_1, R_2, and R_3 are each subgroups of $RotSym(P_2)$ and we can see that each of them is isomorphic to $RotSym(P_1)$. Moreover, since R_i and R_j move disjoint copies of P_1 when $i \neq j$, their elements commute and their intersection contains only the identity. Thus, the (internal) direct product $R_1 \times R_2 \times R_3$ is a subgroup of $RotSym(P_2)$ and is isomorphic to the direct product of three copies of $RotSym(P_1)$. (For more about the internal direct product of subgroups, see, for example, [1].) As is customary, we can denote the elements of $R_1 \times R_2 \times R_3$ by ordered triples, the entries coming from R_1, R_2, R_3, respectively. Exploiting the isomorphism between each R_i and $RotSym(P_1)$, we'll denote an element of R_i by the corresponding symmetry of P_1. Thus, for example, (240°, 120°, 0°) denotes rotating the topmost copy of P_1 by 240° clockwise, the lower-right copy by 120° clockwise and leaving the lower-left copy unchanged.

Continuing with the example, set t equal to a 120° clockwise rotation of P_2. We claim that t normalizes $R_1 \times R_2 \times R_3$. To establish the claim, we need to take any element (r_1, r_2, r_3) of $R_1 \times R_2 \times R_3$, $r_i \in R_i$, and show that $t^{-1}(r_1, r_2, r_3)t$ is back in $R_1 \times R_2 \times R_3$. Examine the symmetry $t^{-1}(r_1, r_2, r_3)t$ (reading right to left): t rotates P_2 120°, thus moving the topmost copy of P_1 to the lower right, the lower-right copy to the lower left and the lower left to the top. Next r_3 performs a symmetry transformation on the lower left, which used to be the lower right, while r_2 and r_1 operate on the lower-right and topmost copies of P_1, formerly the topmost and lower-left copies, respectively. Finally, t^{-1} rotates the three transformed copies of P_1 back to their original locations within P_2. The result, then, is that the symmetry r_3, which would have originally applied to the lower-left copy of P_1, instead applies to the lower-right copy; the symmetry r_2, which would have originally applied to the lower right, instead applies to the topmost; and the symmetry r_1, which would have originally applied to the topmost, instead applies to the lower left. Symbolically, $t^{-1}(r_1, r_2, r_3)t = (r_2, r_3, r_1)$, which is back in $R_1 \times R_2 \times R_3$. FIGURE 3, illustrates for example, how $t^{-1}(240°, 120°, 0°)t = (120°, 0°, 240°)$. At any rate, t not only normalizes $R_1 \times R_2 \times R_3$, but it does so in a very interesting way by simply permuting the symmetries in the ordered triples: second replacing first, third replacing second, first replacing third.

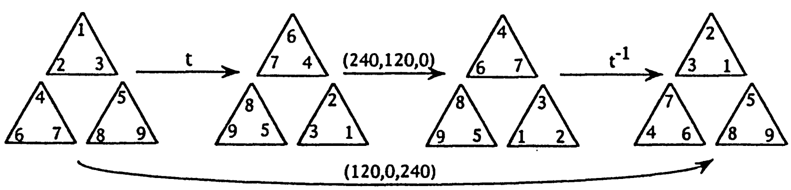


FIGURE 3

Generalizing the reasoning of the example, we see how the structure of $RotSym(P_n)$ relates to that of $RotSym(P_{n-1})$. P_n is made up of p disjoint, small copies of P_{n-1}. Setting R_i, $1 \leq i \leq p$, equal to the subgroup of $RotSym(P_n)$ consisting of symmetries that fix all except the ith copy of P_{n-1}, we obtain the subgroup $R_1 \times R_2 \ldots \times R_p$ of $RotSym(P_n)$, whose elements we can think of as ordered p-tuples of symmetries from $RotSym(P_{n-1})$. Moreover, if t is a clockwise rotation of P_n by $(180/p)°$ and $(r_1, r_2, \ldots, r_p)$ is any element of $R_1 \times R_2 \ldots \times R_p$, then $t^{-1}(r_1, r_2, \ldots, r_p)t = (r_2, r_3, \ldots, r_p, r_1)$. Thus, t normalizes $R_1 \times R_2 \ldots \times R_p$ by permuting components. $RotSym(P_n)$ is generated by $R_1, \ldots R_p$ and t and is seen to be equal to $(R_1 \times R_2 \ldots \times R_p)\langle t \rangle$ since $R_1 \times R_2 \ldots \times R_p$ is normalized by t (where $\langle t \rangle$ denotes the cyclic subgroup generated by t). We also note that $(R_1 \times R_2 \ldots \times R_p) \cap \langle t \rangle$ contains only the identity.

The structure we've just uncovered, a direct product of a number of copies of the same group with an element (in our case, t) normalizing it so as to permute the components of the ordered tuples of the direct product, is an important one in group theory. A group with that structure is called a *wreath product*. To read more about wreath products, see any advanced group theory text, for example, **[2]**.

Using the customary (regular) wreath product notation, what we've shown is that $RotSym(P_n)$ is $RotSym(P_{n-1}) \wr \langle t \rangle$, the wreath product of $RotSym(P_{n-1})$ with $\langle t \rangle$. In other words, using the obvious fact that $\langle t \rangle \cong Z_p$, a cyclic group of order p, we've shown $RotSym(P_n) \cong RotSym(P_{n-1}) \wr Z_p$. Thus, $RotSym(P_1) \cong Z_p$ (clearly), $RotSym(P_2) \cong RotSym(P_1) \wr Z_p \cong Z_p \wr Z_p$, $RotSym(P_3) \cong RotSym(P_2) \wr Z_p \cong (Z_p \wr Z_p) \wr Z_p$, etc. Group theoretically, then, $RotSym(P_n)$ is an iterated wreath product of n copies of Z_p. As advertised in the introduction, we've shown a way to visualize that particular wreath product as a group of symmetries.

Next, we'll calculate $|RotSym(P_n)|$, the order of $RotSym(P_n)$. We've already noted $RotSym(P_n) = (R_1 \times R_2 \times \cdots \times R_p)\langle t \rangle$ with $(R_1 \times R_2 \times \cdots \times R_p) \cap \langle t \rangle$ containing only the identity. From those two facts, it is a standard exercise to show that the index of $R_1 \times R_2 \times \cdots \times R_p$ in $RotSym(P_n)$ is p, the order of $\langle t \rangle$. (Show $(R_1 \times R_2 \times \cdots \times R_p)t^i$, $1 \le i \le p$, is a complete set of right cosets.) Therefore, $|RotSym(P_n)| = |R_1 \times R_2 \times \cdots \times R_p|p = |RotSym(P_{n-1})|^p p$ since $R_1 \times R_2 \times \cdots \times R_p$ is the direct product of p copies of $RotSym(P_{n-1})$. Noting that $|RotSym(P_1)| = p$, we calculate, for example, $|RotSym(P_2)| = p^p p$ and $|RotSym(P_3)| = (p^p p)^p p = p^{p^2} p^p p$. By induction, $|RotSym(P_n)| = p^{p^{n-1}} p^{p^{n-2}} \ldots p^p p = p^{p^{n-1} + p^{n-2} + \cdots + p + 1}$ or $p^{(p^n - 1)/(p-1)}$.

Following a standard procedure, we'll now show how to think of $RotSym(P_n)$ as a group of permutations. By construction, P_n has p^n vertices. Numbering the vertices, we can think of a symmetry of P_n as a permutation of the numbers 1 through p^n simply by recording how the symmetry permutes the numbered vertices. For instance, the symmetry in FIGURE 2 can be represented by the permutation

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 & 18 & 19 & 20 & 21 & 22 & 23 & 24 & 25 & 26 & 27 \\ 24 & 25 & 18 & 26 & 14 & 27 & 19 & 15 & 11 & 9 & 23 & 5 & 8 & 17 & 22 & 3 & 4 & 13 & 21 & 1 & 2 & 6 & 7 & 10 & 12 & 16 & 20 \end{pmatrix}.$$

In that way, we regard $RotSym(P_n)$ as a subgroup of S_{p^n}, the symmetric group on the numbers 1 through p^n.

To complete our intended program, we want to show that $RotSym(P_n)$, regarded as a subgroup of S_{p^n}, is actually a Sylow p-subgroup of S_{p^n}. To do so, we simply need to show that $|RotSym(P_n)|$, which we know from above is $p^{(p^n-1)/(p-1)}$, is the highest power of p dividing $|S_{p^n}| = (p^n)!$. To see that's the case, we'll think recursively again and look at what new powers of p appear between $(p^{n-1})!$ and $(p^n)!$. For convenience, $|m|_p$ will denote the highest power of p dividing the number m.

Consider

$$\frac{(p^n)!}{(p^{n-1})!} = (p^{n-1}+1)(p^{n-1}+2)\ldots p^n.$$

The only factors contributing factors of p are those of the form $p^{n-1} + kp$, $1 \le k \le p^{n-1} - p^{n-2}$. Thus, we have

$$\left|\frac{(p^n)!}{(p^{n-1})!}\right|_p = \left|\prod_{k=1}^{p^{n-1}-p^{n-2}} (p^{n-1}+kp)\right|_p$$

$$= p^{(p^{n-1}-p^{n-2})}\left|\prod_{k=1}^{p^{n-1}-p^{n-2}} (p^{n-2}+k)\right|_p = p^{(p^{n-1}-p^{n-2})}\left|\frac{(p^{n-1})!}{(p^{n-2})!}\right|_p.$$

Repeating the same calculation a total of $n-1$ times, we get

$$\left|\frac{(p^n)!}{(p^{n-1})!}\right|_p = p^{(p^{n-1}-p^{n-2})}p^{(p^{n-2}-p^{n-3})}p^{(p^{n-3}-p^{n-4})}\cdots p^{(p-1)}p = p^{p^{n-1}}.$$

Thus, $|(p^n)!|_p = p^{p^{n-1}}|(p^{n-2})!|_p$. Inductively, we now get $|(p^n)!|_p = p^{p^{n-1}}p^{p^{n-2}}\cdots p^p p = p^{p^{n-1}+p^{n-2}+\cdots+p+1}$ or $p^{(p^n-1)/(p-1)}$, the order of $RotSym(P_n)$.

In summary, we've kept all our promises by showing that a Sylow p-subgroup of S_{p^n} is isomorphic to the group of rotational symmetries of P_n, a sort of "pre-fractal," and, furthermore, it is the wreath product of p copies of a cyclic group of order p.

REFERENCES

1. I. N. Herstein, *Topics in Algebra*, 2nd edition, John Wiley & Sons, New York, 1975.
2. W. R. Scott, *Group Theory*, Dover Publications, Inc., Mineola, NY, 1987.
3. L. Kaloujnine, La structure des p-groups de Sylow des groups symétriques finis, *Ann. Sci. École Norm. Super.* 65 (1948), 239–276.

My Dance is Mathematics

Down, down, down into the darkness of the grave
Gently they go, the beautiful, the tender, the kind;
Quietly they go, the intelligent, the witty, the brave.
I know. But I do not approve. And I am not resigned.

From "Dirge without Music" by Edna St. Vincent Millary; offered by Hermann Weyl in a Memorial Address for Amalie Emmy Noether on April 26, 1935 at Bryn Mawr College.

They called you *der* Noether, as if mathematics
was for only men. In 1964, nearly thirty years
past your death, at last I saw you in a spotlight,
in a New York World's Fair mural
titled "Men of Modern Mathematics."

Colleagues praised your brilliance, after they had said
that you were fat and plain and rough and loud.
Some mentioned your kindness and good humor,
but only in the end admitted your key role
in creation of axiomatic algebra.

With laughter a tale is told from 1890,
when you were eight years old. At a birthday party,
you spoke up to solve a hard math puzzle.
That day you set yourself apart to be
someone that I would follow.

As I followed you, I saw you have to choose
between mathematics and other romance.
Though men could embrace both,
for you, the different standard.

Repeating the same calculation a total of $n-1$ times, we get

$$\left|\frac{(p^n)!}{(p^{n-1})!}\right|_p = p^{(p^{n-1}-p^{n-2})}p^{(p^{n-2}-p^{n-3})}p^{(p^{n-3}-p^{n-4})}\dots p^{(p-1)}p = p^{p^{n-1}}.$$

Thus, $|(p^n)!|_p = p^{p^{n-1}}|(p^{n-2})!|_p$. Inductively, we now get $|(p^n)!|_p = p^{p^{n-1}}p^{p^{n-2}}\dots p^p p = p^{p^{n-1}+p^{n-2}+\dots+p+1}$ or $p^{(p^n-1)/(p-1)}$, the order of $RotSym(P_n)$.

In summary, we've kept all our promises by showing that a Sylow p-subgroup of S_{p^n} is isomorphic to the group of rotational symmetries of P_n, a sort of "pre-fractal," and, furthermore, it is the wreath product of p copies of a cyclic group of order p.

REFERENCES

1. I. N. Herstein, *Topics in Algebra*, 2nd edition, John Wiley & Sons, New York, 1975.
2. W. R. Scott, *Group Theory*, Dover Publications, Inc., Mineola, NY, 1987.
3. L. Kaloujnine, La structure des p-groups de Sylow des groups symétriques finis, *Ann. Sci. École Norm. Super.* 65 (1948), 239–276.

My Dance is Mathematics

Down, down, down into the darkness of the grave
Gently they go, the beautiful, the tender, the kind;
Quietly they go, the intelligent, the witty, the brave.
I know. But I do not approve. And I am not resigned.

From "Dirge without Music" by Edna St. Vincent Millary; offered by Hermann Weyl in a Memorial Address for Amalie Emmy Noether on April 26, 1935 at Bryn Mawr College.

They called you *der* Noether, as if mathematics
was for only men. In 1964, nearly thirty years
past your death, at last I saw you in a spotlight,
in a New York World's Fair mural
titled "Men of Modern Mathematics."

Colleagues praised your brilliance, after they had said
that you were fat and plain and rough and loud.
Some mentioned your kindness and good humor,
but only in the end admitted your key role
in creation of axiomatic algebra.

With laughter a tale is told from 1890,
when you were eight years old. At a birthday party,
you spoke up to solve a hard math puzzle.
That day you set yourself apart to be
someone that I would follow.

As I followed you, I saw you have to choose
between mathematics and other romance.
Though men could embrace both,
for you, the different standard.

I heard fathers say, dance with Emmy. Do it early
in the evening, and she won't expect you
to stay with her. Max is kind and a good friend,
and his daughter likes to dance.

If a woman's dance
is mathematics,
must she dance alone?

I heard mothers say, don't tease.
Although Emmy's strange, her heart is kind.
She helps her mother clean the house,
and she can't help having a mathematical mind.

Teachers said, she's smart
but rather stubborn, contentious and loud,
an abstract thinker not like us,
and not conditioned to favor our ideas.

Students said, she's hard
to follow, bores me. A few in front row seats
saw her shape a vital research program—
they built while standing on her shoulders.

Emmy Noether's abstract axiomatic view
altered the face of algebra.
She helped us think in simple terms
that flowered in their generality.

In spite of Emmy's capabilities,
always there were reasons
not to give her rank
or permanent employment.
She's a pacifist, a woman.
She's a woman and a Jew.
She doesn't think as we do.

History books now say that Noether
is the greatest mathematician
that her sex has produced. They say
that—for a woman—she was very good.

Direct and courageous,
lacking self-concern,
elegant of mind,
nurturing and kind,
a moral solace
in an evil time,
a poet of logical ideas.

If a woman's dance
is mathematics,
must she dance alone?

Honor Emmy Noether.
Invite a mathematician
to dance.

—JoANNE GROWNEY
DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
BLOOMSBURG UNIVERSITY, BLOOMSBURG, PA 17815

Rubik's Clock and Its Solution

JÓZSEF DÉNES
Csaba Utca 10 V 42
Budapest 1122, Hungary

GARY L. MULLEN*
Pennsylvania State University
University Park, PA 16802

In this note we describe Rubik's latest puzzle, Rubik's Clock, which is sold by Matchbox. Since the solution provides a nice application of linear algebra to a practical problem, we also provide a solution (which incidentally, is not provided at the time of purchase) to the puzzle. The "clock" consists of 18 small clocks, nine on each of two sides of a plastic device that also contains four buttons and four wheels, as illustrated in FIGURE 1 for the front face of the clock (the back is similar). Note that the four wheels are on the perimeter and the four buttons are centered about the central clock.

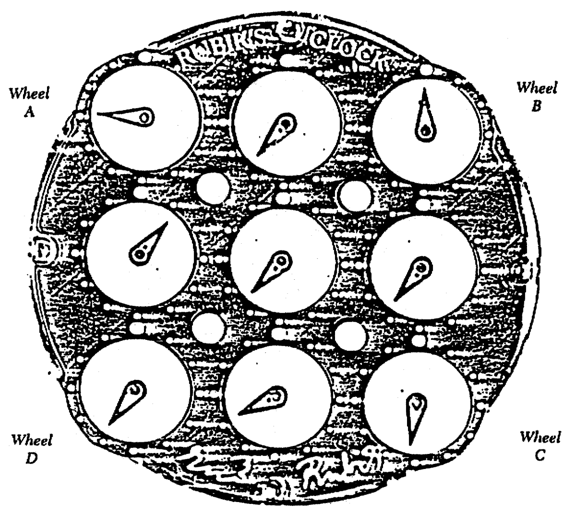


FIGURE 1
Rubik's Clock.

Each of the four buttons can be either up or down independently of the others and each of the wheels can be turned independently of the remaining wheels. As a wheel is turned one notch forward or clockwise (backward or counterclockwise), hands on some of the clocks on the front face move one hour forward (backward) while on the back face, some clocks move one hour backward (forward). Thus if one wheel is moved 12 notches in the same direction, then all 18 clocks will be in their original positions. If the four buttons are then changed to a different setting and a wheel is again moved, in general a different collection of the clocks move. The goal is, through a sequence of operations involving the four wheels and four buttons, to convert each of the 18 clocks from its original setting so as to read 12 o'clock. As will be illustrated below, this is equivalent to solving modulo 12 the system of linear equations represented by the matrix equation (*).

*This author would like to thank the National Security Agency for partial support under grant agreement #MDA904-87-H-2023. Thanks are also due to the Mathematical Sciences Department at Clemson University for its hospitality when this paper was written.

There are 12^{18} possible configurations of the 18 clocks, although as noted below, not all of these are realizable. For each of the 16 settings of the four buttons, the four wheels can be moved in 12^4 ways. Hence there are $(12^4)^{16} = 12^{64}$ possible moves or actions that could be applied. Although these commute with each other, one might still expect the solution to Rubik's Clock to be very complex. However, as we show below, the solution is mathematically very simple, in fact much simpler than the group theoretical solution of Rubik's Cube. See [1] and [2] for a discussion of various Rubik's puzzles, and the references in [1] for papers dealing with Rubik's Cube.

We first establish some notation. Let the side of the clock whose faces are light blue be side I and the side whose faces are dark blue be side II. Using side I with the clock positioned so that each of the nine clocks has 12 at the top, we label the four wheels at top left, top right, bottom left, and bottom right as A, B, C, and D, respectively. We will denote the 16 possible settings of the four buttons as $\begin{smallmatrix}00\\00\end{smallmatrix}, \begin{smallmatrix}10\\00\end{smallmatrix}, \begin{smallmatrix}01\\00\end{smallmatrix}, \dots, \begin{smallmatrix}11\\11\end{smallmatrix}$ where a 0 indicates that the button is up and 1 indicates that the button is down. Finally, number the clocks on side I as $\begin{smallmatrix}1 & 2 & 3\\4 & 5 & 6\\7 & 8 & 9\end{smallmatrix}$ and the clocks on side II as $\begin{smallmatrix}10 & 11 & 12\\13 & 14 & 15\\16 & 17 & 18\end{smallmatrix}$. For $i = 1, 2, \dots, 18$, let c_i be the i-th clock and b_i be the setting on the i-th clock; thus b_i is an integer between 1 and 12.

Given a fixed setting of the buttons it is not true that each of the wheels A, B, C, and D has a different effect on the 18 clocks. For example with the setting $\begin{smallmatrix}01\\01\end{smallmatrix}$ wheels A and D have the same effect moving hands on the clocks $c_1, c_2, c_4, c_5, c_7, c_8, c_{12}, c_{18}$ while wheels B and C have the same effect of moving hands on the clocks $c_3, c_9, c_{10}, c_{11}, c_{13}, c_{14}, c_{16}, c_{17}$. This is indicated from positions 25 and 26 in Table 1 and from columns 25 and 26 (indicated by $*$'s) of the matrix M defined below by the movement of the 18 clocks relative to the 30 unit operations. As another illustration we note that in the $\begin{smallmatrix}00\\00\end{smallmatrix}$ setting, each of the wheels A, B, C, and D has the same effect, (see position 1 in the table and column 1 of matrix M). In this case, hands on all clocks except $c_{11}, c_{13}, c_{14}, c_{15}, c_{17}$ move.

Of the 64 potentially different positions, one can check quite quickly that there are really only 30 truly different positions that need to be considered. These have been summarized in Table 1. We define the j-th unit as turning any one of the designated wheels from position j one notch in a clockwise direction. These unit operations then span the solution space for Rubik's Clock.

TABLE 1

POS.	*BUT.*	*WHEELS*	*POS.*	*BUT.*	*WHEELS*	*POS.*	*BUT.*	*WHEELS*
1	$\begin{smallmatrix}00\\00\end{smallmatrix}$	*ABCD*	11	$\begin{smallmatrix}01\\11\end{smallmatrix}$	*A*	21	$\begin{smallmatrix}00\\11\end{smallmatrix}$	*AB*
2	$\begin{smallmatrix}10\\00\end{smallmatrix}$	*BCD*	12	$\begin{smallmatrix}01\\11\end{smallmatrix}$	*BCD*	22	$\begin{smallmatrix}00\\11\end{smallmatrix}$	*CD*
3	$\begin{smallmatrix}10\\00\end{smallmatrix}$	*A*	13	$\begin{smallmatrix}10\\11\end{smallmatrix}$	*A CD*	23	$\begin{smallmatrix}10\\10\end{smallmatrix}$	*A D*
4	$\begin{smallmatrix}01\\00\end{smallmatrix}$	*A CD*	14	$\begin{smallmatrix}10\\11\end{smallmatrix}$	*B*	24	$\begin{smallmatrix}10\\10\end{smallmatrix}$	*BC*
5	$\begin{smallmatrix}01\\00\end{smallmatrix}$	*B*	15	$\begin{smallmatrix}11\\10\end{smallmatrix}$	*AB D*	25	$\begin{smallmatrix}01\\01\end{smallmatrix}$	*A D*
6	$\begin{smallmatrix}00\\01\end{smallmatrix}$	*AB D*	16	$\begin{smallmatrix}11\\10\end{smallmatrix}$	*C*	26	$\begin{smallmatrix}01\\01\end{smallmatrix}$	*BC*
7	$\begin{smallmatrix}00\\01\end{smallmatrix}$	*C*	17	$\begin{smallmatrix}11\\01\end{smallmatrix}$	*ABC*	27	$\begin{smallmatrix}10\\01\end{smallmatrix}$	*A C*
8	$\begin{smallmatrix}00\\10\end{smallmatrix}$	*ABC*	18	$\begin{smallmatrix}11\\01\end{smallmatrix}$	*D*	28	$\begin{smallmatrix}10\\01\end{smallmatrix}$	*B D*
9	$\begin{smallmatrix}00\\10\end{smallmatrix}$	*D*	19	$\begin{smallmatrix}11\\00\end{smallmatrix}$	*AB*	29	$\begin{smallmatrix}01\\10\end{smallmatrix}$	*A C*
10	$\begin{smallmatrix}11\\11\end{smallmatrix}$	*ABCD*	20	$\begin{smallmatrix}11\\00\end{smallmatrix}$	*CD*	30	$\begin{smallmatrix}01\\10\end{smallmatrix}$	*B D*

For $i = 1, \ldots, 18$ and $j = 1, \ldots, 30$, let $m_{ij} = 1, -1$, or 0 depending upon whether the hand of i-th clock moves 1 hour forward, 1 hour backward, or not at all under the j-th unit operation. Then the 18×30 matrix $M = (m_{ij})$ below indicates the movement of the 18 clocks relative to the 30 unit operations.

$$M = \begin{array}{c}
\begin{matrix} & * & * & & & & \end{matrix} \\
\begin{pmatrix}
1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\
1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\
1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\
1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\
1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \\
1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\
1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\
1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\
-1 & -1 & 0 & 0 & -1 & -1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & -1 & -1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & -1 \\
0 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & -1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & -1 & 0 & 0 & -1 \\
-1 & 0 & -1 & -1 & 0 & -1 & 0 & -1 & 0 & -1 & -1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & -1 & 0 & -1 & -1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & -1 \\
0 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & -1 & -1 & 0 & -1 & -1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & 0 & -1 & -1 & 0 & 0 & -1 & -1 & 0 & 0 & -1 \\
0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & -1 & -1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & -1 \\
-1 & -1 & 0 & -1 & 0 & 0 & -1 & -1 & 0 & -1 & 0 & -1 & -1 & 0 & 0 & -1 & -1 & 0 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & -1 & -1 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & -1 & 0 & -1 & -1 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & -1 & -1 & 0 & 0 & -1 \\
-1 & -1 & 0 & -1 & 0 & -1 & 0 & 0 & -1 & -1 & 0 & -1 & -1 & 0 & -1 & 0 & 0 & -1 & 0 & -1 & 0 & -1 & -1 & 0 & -1 & 0 & 0 & -1 & 0 & -1
\end{pmatrix}
\end{array}$$

Let X denote the column vector $[x_1 \cdots x_{30}]^T$ where x_j denotes the number of times the j-th unit operation is applied. For a given column vector $S = [b_1 \cdots b_{18}]^T$ where b_i denotes the setting on the i-th clock, the solution to the puzzle is now simply to solve, modulo 12, the system of linear equations represented by the matrix equation

$$MX = 12J - S, \tag{*}$$

where $J = [1, 1, \ldots, 1]^T$.

The set of all achievable positions of the clocks is simply the modulo 12 row space of the matrix M. Consequently one could use the same method to arrive at any preassigned collection $G = [d_1 \cdots d_8]^T$ of times for the 18 clocks by simply solving the system $MX = G - S$, provided of course that $G - S$ is in the row space of M.

Clearly if (*) has a solution X, then it represents a solution to Rubik's Clock. Conversely given a solution to Rubik's Clock, it is easily seen that the operations satisfy (*). We remind the reader that a system of linear equations over the integers modulo 12 has a solution if, and only if, it has a solution modulo 3 and modulo 4. Of course given solutions modulo 3 and 4, one could, by the Chinese Remainder Theorem, construct a solution modulo 12.

Upon closer examination of Rubik's Clock one can see that clocks c_1 and c_{12}, c_3 and c_{10}, c_7 and c_{18}, and c_9 and c_{16} are dependent, so that if the first clock of one of the pairs advances an hour, the second clock of the pair goes back an hour. Consequently we only need to consider a system of size 14×30. The reader may want to write a program to solve (*) on his or her own personal computer to take the drudgery out of solving the system (*) by hand. For example on an IBM PS/2, a solution to (*) is found in less than 30 seconds.

From a solution X to (*) one can mechanically implement the solution on Rubik's Clock by simply applying the unit operations $x_1, \ldots, x_{30}$ according to Table 1. For instance if $x_8 = 4$, one consults Table 1 for position 8, then sets the buttons in position $\substack{00\\10}$ and uses any one wheel from A or B or C and moves that wheel 4 notches ahead in the clockwise direction. If $x_j = 0$ then position j from Table 1 need not be applied. Thus only the nonzero x_j's need to be implemented and so to make life easier one could look for a solution X of (*) with as many $x_j = 0$ as possible.

As an illustration consider the initial setting of the clocks given by:

$$\begin{array}{llll} b_1 = 8 & b_6 = 11 & b_{11} = 2 & b_{16} = 8 \\ b_2 = 11 & b_7 = 7 & b_{12} = 4 & b_{17} = 5 \\ b_3 = 11 & b_8 = 11 & b_{13} = 12 & b_{18} = 5 \\ b_4 = 11 & b_9 = 4 & b_{14} = 10 & \\ b_5 = 11 & b_{10} = 1 & b_{15} = 8 & \end{array} \qquad (**)$$

A solution to (*) is $x_1 = 1, x_2 = 5, x_3 = 10, x_4 = 4, x_5 = 6, x_6 = 10, x_7 = 8, x_8 = 5, x_{10} = 9$, and $x_{19} = 1$, with the remaining $x_j = 0$. We now indicate the settings on the various clocks as, for each j, the j-th unit operation is applied x_j times.

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}	c_{13}	c_{14}	c_{15}	c_{16}	c_{17}	c_{18}
initial	8	11	11	11	11	11	7	11	4	1	2	4	12	10	8	8	5	5
x_1	9	12	12	12	12	12	8	12	5	12	2	3	12	10	8	7	5	4
x_2	9	5	5	5	5	5	1	5	10	7	2	3	12	10	8	2	5	11
x_3	7	5	5	5	5	5	1	5	10	7	4	5	12	12	10	2	5	11
x_4	11	9	5	9	9	9	5	9	2	7	4	1	12	12	10	10	5	7
x_5	11	9	11	9	9	9	5	9	2	1	10	1	6	6	10	10	5	7
x_6	9	7	9	7	7	7	3	7	2	3	10	3	6	6	10	10	5	9
x_7	9	7	9	7	7	7	3	7	10	3	10	3	10	10	10	2	9	9
x_8	2	12	2	12	12	12	3	12	3	10	10	10	10	10	10	9	9	9
x_{10}	11	12	11	12	12	12	12	12	12	1	1	1	1	1	1	12	12	12
x_{19}	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12

We note that as a check, one can derive the settings of the clocks as each j-th unit operation is applied x_j times, the result can be derived by adding the vector $M_j x_j$ to the previously determined settings of the clock, i.e. $MX = M_1 x_1 + M_2 x_2 + \cdots + M_{30} x_{30}$, where M_j is the j-th column of M. We also note that the operations are commutative so that the order of implementation is not important. The commutativity is of course also apparent since the additive group of the integers modulo 12 is commutative.

Acknowledgement. We would like to thank Joel V. Brawley for a number of helpful suggestions. Thanks are also due the referees for a number of suggestions that helped to improve an earlier version of the paper.

REFERENCES

1. M. E. Larson, Rubik's revenge: the group theoretical solution, *Amer. Math. Monthly*, 92 (1985), No. 6, pp. 381–390.
2. E. C. Turner and K. F. Gold, Rubik's groups, *Amer. Math. Monthly*, 92 (1985), No. 9, pp. 617–629.

Position Graphs for Pong Hau K'i and Mu Torere

PHILIP D. STRAFFIN, JR.
Beloit College
Beloit, WI 53511

For the past several years at Beloit College, we have taught a course in Ethnomathematics, using Marcia Ascher's [2] as the primary text. The course has been extremely popular with students, who enjoy the exposure to mathematical ideas in other cultures, and broadening for both students and faculty.

For our students, one of the hardest sections of [2] is Ascher's graph theoretic analysis of the Maori game Mu Torere, which first appeared in this MAGAZINE [1]. I will present a simpler graph for Mu Torere, which makes strategy for playing the game transparent. I will begin by illustrating the graph theoretic technique on a very simple, but still interesting game—the Chinese children's game Pong Hau K'i.

Pong Hau K'i Pong Hau K'i, as described in [3] and [4], is played on the board in FIGURE 1. Black has two black stones, White has two white stones, which are initially placed as in FIGURE 1. The players alternately move one of their stones along an edge to the single empty position on the board; a player loses when this is not possible. I will assume that Black moves first.

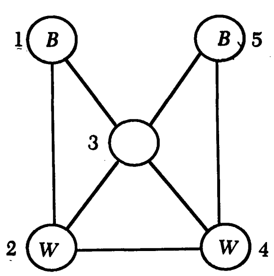
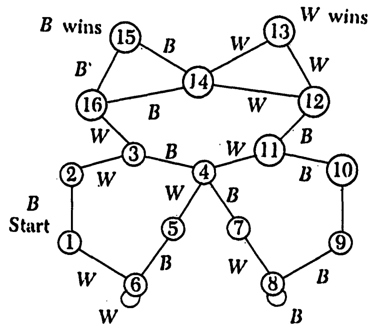


FIGURE 1
Starting position for Pong Hau K'i.

We begin by listing the possible positions in Pong Hau K'i. There are $5!/2!2!1! = 30$ of them. However, we can reduce this number by identifying two positions if they are carried into each other by the vertical symmetry of the board. There are two positions that are self-symmetric and the other 28 positions fall into 14 pairs, so we need to consider just 16 positions. They are numbered and listed in the table in FIGURE 2. For example, *WOBWB* represents the position in which black stones occupy nodes 3 and 5, white stones occupy nodes 1 and 4, and node 2 is empty. By symmetry, this position is equivalent to *BWBOW*; hence *BWBOW* does not appear in the list of positions. Notice that the game starts in position ①.

We can now represent Pong Hau K'i by a *position graph* in which the vertices are positions in the game. Two positions are joined by an edge labeled W if White can move from one position to the other, by an edge labeled B if Black can move from one position to the other. Notice that moves in Pong Hau K'i are reversible, so this is an undirected graph. The graph is shown in FIGURE 2. The loop at vertex ⑥ means that White can move from position ⑥ to itself. You should check your understanding of the notation by verifying that this is indeed possible! Notice that there are no edges labeled W at vertex ⑮. Hence if the game ever reaches this vertex, Black has won



① *BWOWB*	⑤ *OBWWB*	⑨ *WBOBW*	⑬ *OWWBB*
② *OWBWB*	⑥ *BOWWB*	⑩ *OBWBW*	⑭ *BBOWW*
③ *WOBWB*	⑦ *OWBBW*	⑪ *BOWBW*	⑮ *OBBWW*
④ *WBOWB*	⑧ *WOBBW*	⑫ *WOWBB*	⑯ *BOBWW*

FIGURE 2
The position graph for Pong Hau K'i.

(since Black must have moved to the vertex, so it is now White's move, and White cannot move). Similarly, White wins if the game reaches vertex ⑬.

We can use the graph for Pong Hau K'i to tell us how to play the game. The game starts at vertex ① with Black to move. Hence play goes

① $\xrightarrow{B}$ ② $\xrightarrow{W}$ ③ $\xrightarrow{B}$ ④,

at which point White has a choice of moving to ⑤ or to ⑪. As long as the game stays in the lower part of the graph, no one wins. However, if Black should move from ⑪ to ⑫, White immediately wins by moving to ⑬. Similarly, Black can win if White ever moves from ③ to ⑯. Hence neither player in this game can force a win, but both players can avoid losing by avoiding the one "trap." Black, for instance, in the position of Figure 3, must be sure to move her upper stone rather than her lower stone.

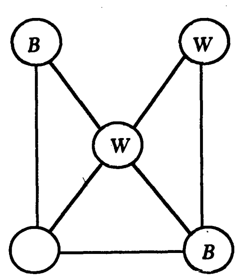


FIGURE 3
A dangerous position for Black in Pong Hau K'i.

Like the draw strategy in Western tic-tac-toe, this draw strategy in Pong Hau K'i is easily discovered by repeated play. A slightly more complicated version of it was given in **[5]**. If you play with children, it might be worth pointing out that after many moves Black might *want* to move to ⑫, either to end the tedium, or in hopes that White will miss the win at ⑬ and move to ⑭, giving Black the win.

Mu Torere The Maori game of Mu Torere (**[1]**–**[4]**) is played on a board that we can model as the graph with nine vertices as shown on the left of Figure 4. The right side of Figure 4 shows an actual modern Maori board design in the form of an octopus. Players *A* and *B* alternately move one of their stones along an edge to the single empty position; a player loses when this is not possible. For each player's first two moves, a stone can be moved from an outer node to the center only if it is adjacent to an opponent's stone. (This restriction is necessary to avoid trivial wins.)

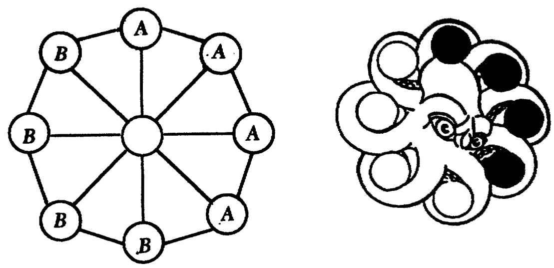
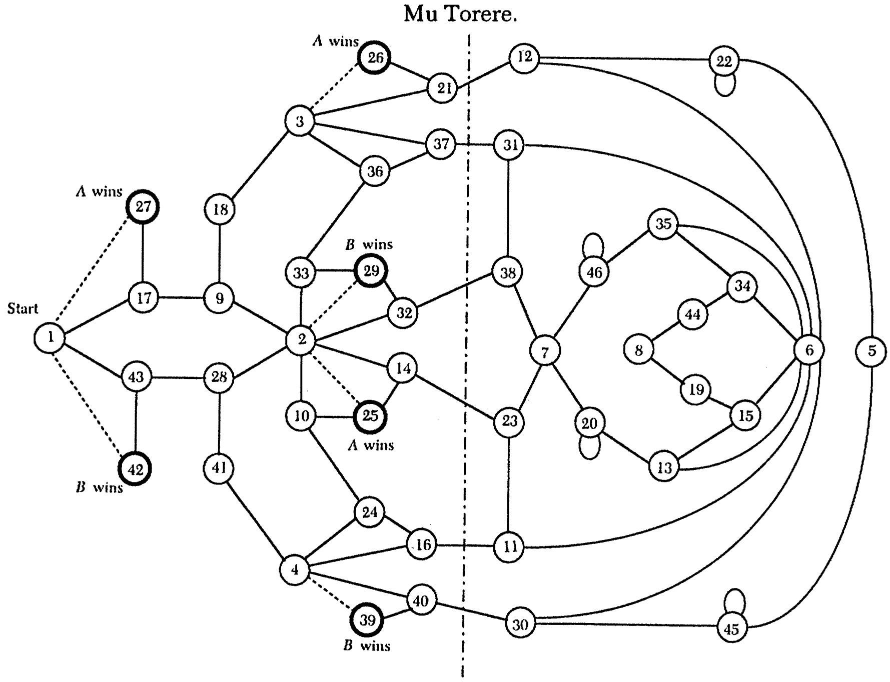


FIGURE 4
Mu Torere.



(1) OAAAABBBB	(13) AOBABABBA	(24) AOABBBAAB	(36) BOBAAABBA
(2) OAAABABBB	(14) AOAABABBB	(25) AOABABBBA	(37) BOAAABBAB
(3) OAAABBABB	(15) AOABABABB	(26) AOABBABBA	(38) BOAABABBA
(4) OAABAABBB	(16) AOBBBAABA	(27) AOABBBBAA	(39) BOBAABAAB
(5) OAABBAABB	(17) AOBBBBAAA	(28) BOAAABBBA	(40) BOAABAABB
(6) OAABABABB	(18) AOBBAAABB	(29) BOBABAAAB	(41) BOAABBBAA
(7) OAABABBAB	(19) AOBABABAB	(30) BOABBAABA	(42) BOBAAAABB
(8) OABABABAB	(20) AOABABBAB	(31) BOAABBABA	(43) BOAAAABBB
(9) AOBBBAAAB	(21) AOBBABBAA	(32) BOBBABAAA	(44) BOABABABA
(10) AOAABBBAB	(22) AOABBAABB	(33) BOABAAABB	(45) BOAABBAAB
(11) AOBABAABB	(23) AOBAABABB	(34) BOBABABAA	(46) BOBABAABA
(12) AOBAABBAB		(35) BOABABAAB	

FIGURE 5
The position graph for Mu Torere.

Mu Torere has $9!/4!4!1! = 630$ positions. However, we will identify two positions if they are carried into each other by one of the 16 symmetries (eight rotations and eight reflections) of the board. Either by direct count, or by applying Pólya's counting theory, we find that this reduces the number of positions to 46, which are listed in the table of FIGURE 5. In the notation, the first letter tells who occupies the center and the remaining eight letters read around the outside of the board; it is not necessary to specify where the reading starts and whether it is clockwise or counter-clockwise. Thus the starting position is ① *OAAAABBBB*. This is equivalent, for example, to *OAAABBBBA*, so the latter does not appear. The position graph for Mu Torere is given in FIGURE 5. The dotted lines are the moves that players cannot make in their first two moves. I have listed Mu Torere positions in the same order that Ascher does in [1] and [2], to facilitate comparison.

FIGURE 5 tells us how to play Mu Torere. Assume A moves first. The game starts

$$① \xrightarrow{A} ⑰ \xrightarrow{B} ⑨ \xrightarrow{A} ②,$$

at which point B can move to ㉝, ㉜, or ㉘. The last leads to a non-obvious win for A:

$$② \xrightarrow{B} ㉘ \xrightarrow{A} ㊶ \xrightarrow{B} ④ \xrightarrow{A} ㉔ \xrightarrow{B} ⑩ \xrightarrow{A} ㉕.$$

If B moves to ㉝ or ㉜ and avoids obvious mistakes, the play moves to the right of the dotted line in FIGURE 5. As long as it stays in the right half, no one wins. However, any move back across the dotted line loses within a couple of moves. Hence neither player can force a win, but both players can avoid a loss by learning to recognize and avoid three "traps." FIGURE 6 shows the losing sequences for A. The common characteristic is that a wedge of B stones (shown in dotted lines) is allowed to "open." The way to play Mu Torere is to avoid opening pitfalls and then *never allow your opponent to open a wedge*. A version of this draw strategy appears in [3].

The draw strategy for Mu Torere can be discovered in repeated play, but it is not obvious to beginning or even lightly experienced players. In fact, when we played Mu

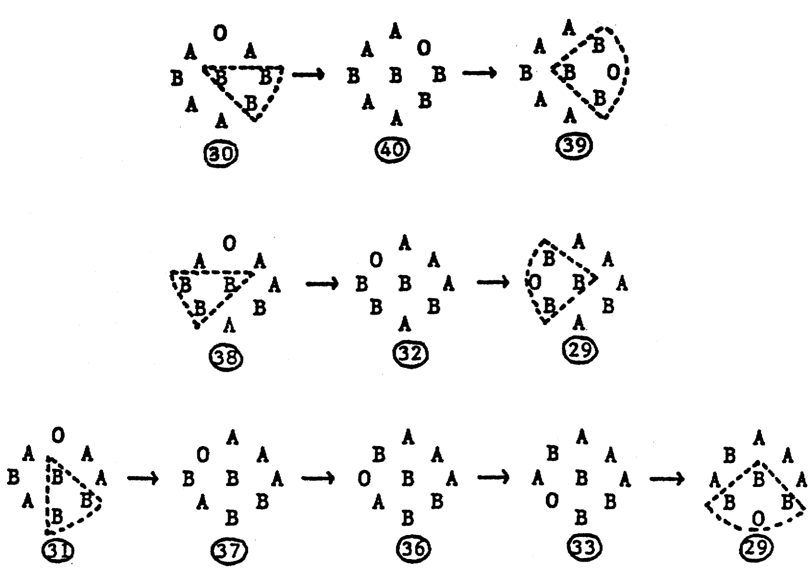


FIGURE 6
Losing sequences for A in Mu Torere.

Torere in my Ethnomathematics class, 23 out of 24 games were won within 20 moves. My students enjoyed the cultural background of the game and they enjoyed playing it, but they also enjoyed seeing how graph theory can analyze a positional game.

REFERENCES

1. M. Ascher, Mu Torere: an analysis of a Maori game, this MAGAZINE, 60 (1987), 90–100.
2. M. Ascher, *Ethnomathematics: A Multicultural View of Mathematical Ideas*, Brooks/Cole Publishing Co., Pacific Grove, CA, 1991.
3. R. C. Bell, *Board and Table Games from Many Civilizations*, Dover Publications, Mineola, NY, 1979.
4. R. C. Bell and M. Cornelius, *Board Games Round the World*, Cambridge University Press, Cambridge, 1988.
5. R. Evans, Pong Hau K'i, *Games and Puzzles*, 53 (1976), 19.

Math Bite: A Quick Counting Proof of the Irrationality of $\sqrt[n]{k}$

VINCENT P. SCHIELACK, JR.
Texas A and M University
College Station, TX 77843

Assume the positive integer k is not the nth power of an integer ($\sqrt[n]{k}$ is obviously rational in the contrary case). If $\sqrt[n]{k}$ is rational, then there exist integers a and b such that $\sqrt[n]{k} = a/b$, or $kb^n = a^n$. Since k is not an nth power of an integer, it must have some prime factor p whose multiplicity is not congruent to $0 (\mathrm{mod}\, n)$. Now the multiplicity of p (and every other prime factor) in b^n is congruent to $0(\mathrm{mod}\, n)$, so the multiplicity of p in kb^n is not congruent to $0(\mathrm{mod}\, n)$, while the multiplicity of p in a^n obviously is congruent to $0(\mathrm{mod}\, n)$, forcing the contradiction $kb^n \neq a^n$. Thus, $\sqrt[n]{k}$ is rational only when k is the nth power of an integer. (Note that, unlike the standard proofs of special cases of this result, such as the irrationality of $\sqrt{2}$, this counting argument does not require the assumption that a and b are relatively prime.)

Torere in my Ethnomathematics class, 23 out of 24 games were won within 20 moves. My students enjoyed the cultural background of the game and they enjoyed playing it, but they also enjoyed seeing how graph theory can analyze a positional game.

REFERENCES

1. M. Ascher, Mu Torere: an analysis of a Maori game, this MAGAZINE, 60 (1987), 90–100.
2. M. Ascher, *Ethnomathematics: A Multicultural View of Mathematical Ideas*, Brooks/Cole Publishing Co., Pacific Grove, CA, 1991.
3. R. C. Bell, *Board and Table Games from Many Civilizations*, Dover Publications, Mineola, NY, 1979.
4. R. C. Bell and M. Cornelius, *Board Games Round the World*, Cambridge University Press, Cambridge, 1988.
5. R. Evans, Pong Hau K'i, *Games and Puzzles*, 53 (1976), 19.

Math Bite: A Quick Counting Proof of the Irrationality of $\sqrt[n]{k}$

VINCENT P. SCHIELACK, JR.
Texas A and M University
College Station, TX 77843

Assume the positive integer k is not the nth power of an integer ($\sqrt[n]{k}$ is obviously rational in the contrary case). If $\sqrt[n]{k}$ is rational, then there exist integers a and b such that $\sqrt[n]{k} = a/b$, or $kb^n = a^n$. Since k is not an nth power of an integer, it must have some prime factor p whose multiplicity is not congruent to $0 \pmod n$. Now the multiplicity of p (and every other prime factor) in b^n is congruent to $0 \pmod n$, so the multiplicity of p in kb^n is not congruent to $0 \pmod n$, while the multiplicity of p in a^n obviously is congruent to $0 \pmod n$, forcing the contradiction $kb^n \neq a^n$. Thus, $\sqrt[n]{k}$ is rational only when k is the nth power of an integer. (Note that, unlike the standard proofs of special cases of this result, such as the irrationality of $\sqrt{2}$, this counting argument does not require the assumption that a and b are relatively prime.)

Continued Powers and a Sufficient Condition for Their Convergence

DIXON J. JONES
5112 Fairchild Ave.
Fairbanks, AK 99709-4523

Background Continued powers are the conceptual ancestors and historical descendants of *continued square roots*, which are expressions of the form

$$\lim_{k\to\infty} x_0 + \sqrt{x_1 + \sqrt{x_2 + \sqrt{\cdots + \sqrt{x_k}}}}\ .$$

Examples of continued square roots have been around for quite a while; Pólya and Szegö in 1916 considered their convergence properties [3], and relevant problems date back to the late 19th century. A. Herschfeld's 1935 paper on continued square roots and similar expressions (he called them "infinite radicals" [1]) contained many interesting observations, among them two in particular. First, Herschfeld proved that a continued square root of real nonnegative terms x_n converges if, and only if, $(x_n)^{2^{-n}}$ is bounded. (This was independently discovered some 50 years later by Sizer, who also coined the term "continued square root" [4]). Second, Herschfeld noted in passing that the general form

$$\lim_{k\to\infty} x_0 + \left(x_1 + \left(x_2 + \left(\cdots + (x_k)^p \cdots\right)^p\right)^p\right)^p, \tag{1}$$

includes not only continued square roots ($p = 1/2$), but infinite series and continued fractions as well ($p = 1$ and $p = -1$, respectively). Herschfeld's investigation of expression (1) extended his continued square roots work to arbitrary roots; here is a somewhat restricted version of his generalization.

HERSCHFELD'S CONVERGENCE THEOREM. *For real nonnegative terms x_n and real $p, 0 < p < 1$, the expression* (1) *converges if, and only if, $(x_n)^{p^n}$ is bounded.*

Definitions, notation, and the main result The terms "continued fraction" and "continued root" are accepted, in spite of a certain slipperiness about the adjective "continued." Allowing for some syntactic guilt by association, therefore, we call expression (1) a *continued* (*pth*) *power*, especially when $p > 1$. In imitation of the continued root notation, we adopt the alternate notation ${}^p(x)$ for exponentiation and write continued powers as

$$\lim_{k\to\infty} x_0 +^p \left(x_1 +^p \left(x_2 +^p \left(\cdots +^p (x_k) \cdots\right)\right)\right),$$

or sometimes more informally as

$$x_0 +^p \left(x_1 +^p \left(x_2 +^p (\cdots)\right)\right).$$

We abbreviate a continued power with the notation $C_{k=0}^{\infty}(p, x_k)$, and use C_0^{∞} as a nickname when p and the sequence $\{x_n\}$ of terms are understood. The finite expression

$$x_0 +^p \left(x_1 +^p \left(x_2 +^p \left(\cdots +^p (x_n) \cdots\right)\right)\right)$$

we call the *nth approximant* of the continued power, shortened likewise to $C_{k=0}^{n}(p, x_k)$ and C_0^n. Just as an infinite series is rigorously defined as the limit of its partial sums, it is important to think of a continued power as the limit of its sequence of approximants, especially in view of certain difficulties encountered in generating an approximant from its predecessor.

Can continued powers converge for positive values of p not covered by Herschfeld's Theorem? For $p = 1$ the question has been, shall we say, extensively answered in the affirmative. What can we say for $p = 2$, 100, or even $10^{10^{10}}$? Some results are known for the case $p > 1$, including a ratio test strikingly reminiscent of d'Alembert's test for the convergence of series **[2]**. Here we prove a condition, analogous to Herschfeld's Theorem, that is sufficient for the convergence of a continued power when $p > 1$.

THEOREM. *For real $p > 1$, the continued pth power with real nonnegative terms x_n converges if $(x_n/R)^{p^n}$ is bounded, where $R = (p-1)/p^{p/(p-1)}$.*

Examples and amplifications To get the feel of the theorem and its elements, a few examples are helpful. Some properties of continued powers and their approximants, assumed or referred to casually in this section, receive more careful scrutiny in the proof of the theorem.

I. Perhaps the simplest kind of example is the continued square ($p = 2$) having nonnegative, constant terms:

$$C_{n=0}^{\infty}(2, c) = c +^2 \left(c +^2 \left(c +^2 (\cdots)\right)\right), c \geq 0.$$

For $p = 2$ we have $R = 1/4$. The theorem guarantees convergence if $(x_n/R)^{2^n} = (4c)^{2^n}$ is bounded, so any c between 0 and $1/4$, inclusive, will work. Unlike most continued powers, the limit of this continued square is easily found. If it converges (*i.e.*, if $0 \leq c \leq 1/4$), we may write

$$C_{n=0}^{\infty}(2, c) = c +^2 \left(C_{n=0}^{\infty}(2, c)\right),$$

and solve the equation as a quadratic to obtain $C_{n=0}^{\infty}(2, c) = \frac{1}{2} \pm \frac{1}{2}\sqrt{1-4c}$. (Note the corroborative evidence under the radical for $0 \leq c \leq 1/4$.) The choice of "plus" or "minus" is resolved as follows: When c is at its maximum of $1/4$, the continued square has a limit of $1/2$. We expect the limit for smaller values of c to be less than $1/2$, hence

$$C_{n=0}^{\infty}(2, c) = \tfrac{1}{2} \pm \tfrac{1}{2}\sqrt{1-4c}.$$

An immediate corollary to the theorem can be gleaned from this example: *For $p > 1$, a continued pth power of nonnegative constant terms c converges if $0 \leq c \leq R$.* In fact it can be shown that the continued power *diverges* if $c > R$. More about necessary conditions for convergence can be found in **[2]**. Now, it happens that as p grows, R approaches 1. Hence the "interval of convergence" for a continued power of constants grows larger as p increases.

II. Moving up the scale of complexity, the continued square

$$C_{n=0}^{\infty}\left(2, 2^{2^{-n}-2}\right) = 1/2 +^2 \left(\frac{1}{\sqrt{2^3}} +^2 \left(\frac{1}{\sqrt[4]{2^7}} +^2 \left(\frac{1}{\sqrt[8]{2^{15}}} +^2 (\cdots)\right)\right)\right)$$

converges, since

$$(x_n/R)^{2^n} = \left(2^{2^{-n}}\right)^{2^n} = 2$$

for all n. Or *did* we move up the scale of complexity? Because this example converges, we may multiply by $2/2$ and distribute the denominator to the right:

$$C_{n=0}^{\infty}(2, 2^{2^{-n}-2}) = 2/2\left[1/2 +^2\left(\frac{1}{\sqrt{2^3}} +^2\left(\frac{1}{\sqrt[4]{2^7}} +^2\left(\frac{1}{\sqrt[8]{2^{15}}} +^2(\cdots)\right)\right)\right)\right]$$
$$= 2\left[1/4 +^2\left(1/4 +^2\left(1/4 +^2\left(1/4 +^2(\cdots)\right)\right)\right)\right].$$

It's just a constant multiple of Example I for $c = 1/4$. The value of this continued square is $2(1/2) = 1$.

III. For an example that really *is* different, and shows the limitations of a convergence condition that is sufficient but not necessary, consider

$$C_{n=0}^{\infty}(2, (4^n + 1)/4^{n+1}) = 1/2 +^2\left(5/16 +^2\left(17/64 +^2(\cdots)\right)\right).$$

This fails the test of the theorem: $(x_n/R)^{2^n} = (1 + 1/4^n)^{2^n}$ is unbounded as $n \to \infty$. Yet this continued square does converge. Each term after the zeroth is less than the corresponding term in Example II. (Consider it an exercise to show why.) We can compare corresponding approximants in Examples II and III to see that III's approximants are increasing and yet smaller than II's, and thereby conclude that because II converges, so does III.

Proof of the theorem The gist of the proof is that the approximants of a continued power are nondecreasing, and are bounded by the limit of a convergent iterated map. The only trick needed is a careful justification for "reversing" the associativity of an approximant when it suits our purpose. (For proving the theorem's sufficiency we would be technically correct in assuming $R = (p-1)/p^{p/(p-1)}$. An expression so exotic, however, is probably better appreciated when unearthed, rather than unveiled. Therefore, abandoning technicality for pedagogy (and a little drama), we assume only that $R > 0$, and derive its value in terms of p as a consequence of the mechanics of the proof.)

Let us begin the proof by reading the condition of the theorem more precisely. Suppose there is a real $M > 0$ and an integer $N \geq 0$ such that $(x_n/R)^{p^n} < M$ for $n \geq N$. Using the equivalent expression $x_n < RM^{p^{-n}}$, we begin construction of a continued power's nth approximant, proceeding right-to-left as the associativity of the form suggests, and freely using the convention ${}^p(x) = x^p$:

$$\begin{aligned}
{}^p(x_n) &< R^p M^{p^{-n+1}} \\
x_{n-1} +^p(x_n) &< RM^{p^{-n+1}} + R^p M^{p^{-n+1}} \\
&= RM^{p^{-n+1}}(1 + R^{p-1}) \\
{}^p(x_{n-1} +^p(x_n)) &< R^p M^{p^{-n+2}}\{{}^p(1 + R^{p-1})\} \\
x_{n-2} +^p(x_{n-1} +^p(x_n)) &< RM^{p^{-n+2}} + R^p M^{p^{-n+2}}\{{}^p(1 + R^{p-1})\} \\
&= RM^{p^{-n+2}}(1 + R^{p-1} \cdot {}^p(1 + R^{p-1})).
\end{aligned}$$

With n fixed, and making the substitution $r = R^{p-1}$, an induction proof on the index i $(i \leq n - N)$ shows that

$$\begin{aligned}
&x_{n-i} +^p(x_{n-i+1} +^p(\cdots +^p(x_n)\cdots)) \\
&\quad < RM^{p^{-n+i}}(1 + r \cdot {}^p(1 + r \cdot {}^p(\cdots + r \cdot {}^p(1 + r)\cdots))),
\end{aligned}$$

where r appears i times on the right. When $i = n - N$, the left side becomes an expression for the nth approximant, truncated at x_N:

$$C^n_{k=N}(p, x_k) < RM^{p^{-N}}(1 + r \cdot^p (1 + r \cdot^p (\cdots + r \cdot^p (1 + r) \cdots))), \tag{2}$$

where there are now $(n - N)$ constants r on the right.

The proof focuses now on the behavior of inequality (2) as n grows without bound. We consider each side in turn, beginning with C^n_N. Although we regard a continued power as the limit of the sequence of approximants C^n_0, it suffices to consider just the sequence of truncated approximants C^n_N, because the identity

$$C^n_0 = x_0 +^p (x_1 +^p (\cdots +^p (x_{N-1} +^p (C^n_N)) \cdots))$$

assures us that C^n_0 converges if C^n_N converges as $n \to \infty$, the former being at most finitely larger than the latter. Now, bearing in mind that $p > 1$ and $x_n \geq 0$, the inequality $x_{n-1} \leq x_{n-1} +^p (x_n)$ immediately becomes

$$C^{n-1}_{n-1} \leq C^n_{n-1}, \qquad n \geq 1.$$

From this auspicious beginning we can build truncated approximants on each side, term-by-term, layer-by-layer, from right to left, to obtain inequalities of the form

$$C^{n-1}_i \leq C^n_i, \qquad 0 \leq i \leq n - 1.$$

These inequalities show that a sequence of truncated approximants having the same initial term is nondecreasing. In particular, as regards (2), the truncated approximants C^n_N are nondecreasing as n increases.

We now turn our attention to the right-hand side of (2), and of course what interests us here is not the constant quantity $RM^{p^{-N}}$, but the long, continued-power-like expression in which r is repeated $(n - N)$ times. One approach to unraveling such a sequence of nested operations is through the theory of *functional iteration*, the application of a function f to an appropriate input and the successive application of f to the resulting output. The *kth iterate* of f at a (meaning $f \circ f \circ \ldots f \circ f(a)$ with f repeated k times) is written $f^{(k)}(a)$, with $f^{(0)}(a) = a$ and $f^{(1)}(a) = f(a)$ by definition; the iterates obey the forward recursion formula $f^{(k)} = f(f^{(k-1)})$. The form of (2) suggests that a function $f(x) = 1 + rx^p$ might receive an initial input of, say $x_0 = 1$, then by iteration of f a total of $(n - N)$ times the entire expression might be generated:

$$f^{(1)}(1) = 1 + r,$$
$$f^{(2)}(1) = 1 + r \cdot^p (1 + r),$$

and so on. The behavior of this iterated function can be analyzed pretty easily; if it converges, we'll use its limit as an upper bound for inequality (2).

This is the point in the proof where associativity must be considered very carefully. Both sides of expression (2) were constructed from right to left. The left-hand side, the truncated approximant C^n_N, is possessed of an unambiguous order of operations, which can be posed in the form of a backward recursion formula:

$$C^n_j = x_j +^p (C^n_{j+1}), \qquad N \leq j \leq n - 1.$$

The right-hand side of (2) is possessed of an *ambiguous* order of operations, meaning that forward and backward recursion formulas both apply. To see this, define

$$S^n_j = 1 + r \cdot^p (\cdots + r \cdot^p (1 + r) \cdots))$$

so that $(n-j)$ counts the number of times r appears, whereby for instance inequality (2) becomes

$$C_N^n \leq RM^{p^{-N}} S_N^n .$$

Then by construction, a backward recursion holds:

$$S_j^n = 1 + r \cdot^p \left(S_{j+1}^n \right), \qquad N \leq j \leq n-1.$$

But also, by luck, providence, and/or the absence of indices on the rs, we have the forward recursion

$$S_j^n = 1 + r \cdot^p \left(S_j^{n-1} \right), \qquad N \leq j \leq n-1. \tag{3}$$

Formula (3) is evidently our link to functional iteration. With $f(x) = 1 + rx^p$, interpret (3) as $S_j^n = f(S_j^{n-1})$ for $N \leq j \leq n-1$. Employing $S_N^{N+1} = 1 + r = f(1)$ as a basis for induction, we have in short order

$$S_N^n = f^{(n-N)}(1),$$

which validates our treatment of $1 + r\cdot^p(\cdots + r\cdot^p(1+r)\cdots))$ as an iterated function.

(If it seems that the preceding discourse makes a mountain of technicality out of a transparently obvious molehill, be cautioned: Already there are published accounts of continued square roots that are confused and even wrong about the associativity of infinitely nested expressions. The attempt at conciseness here follows the example set by Herschfeld, who carefully distinguished between "left infinite radicals" and "right infinite radicals"—having forward and backward recursion formulas, respectively—and found their convergence properties completely dissimilar.)

To summarize the proof thus far: The inequality (2) having been constructed for each $n \geq N$ in right-to-left fashion, we accept that its right-hand side can be interpreted as $(n-N)$ iterations of the function $f(x) = 1 + rx^p$ from the initial input $x_0 = 1$, all multiplied by the constant $RM^{p^{-N}}$. We continue now by examining the function f and its iterates more closely. Inasmuch as our entire discourse is concerned only with nonnegative real numbers; and whereas the exponent p is strictly greater than 1; and because $r = R^{p-1}$ is strictly positive, it follows by elementary calculus that the graph of $f(x) = 1 + rx^p$ must resemble one of the curves shown in FIGURE 1.

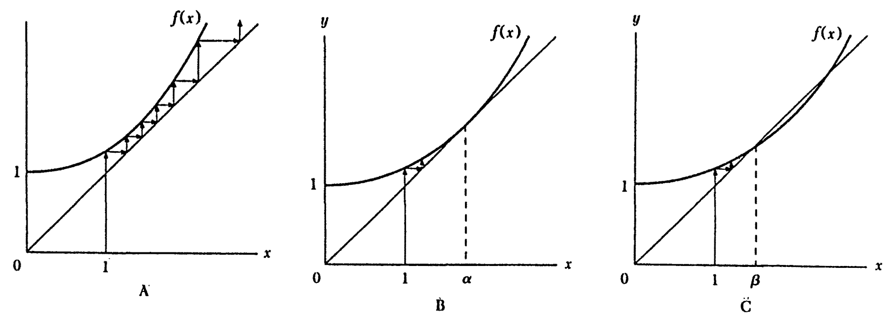


FIGURE 1

Beyond illustrating the possible graphs of f, FIGURE 1 shows in each case the iteration of f from $x_0 = 1$ in what has come to be called a "graphical analysis." From the point $(1, f(1))$, the horizontal skip to the line $y = x$ and the vertical bounce back up to $y = f(x)$ represent, respectively, the transfer of $f(1)$ from the range of f to its domain, followed by the application of f to the domain element $f(1)$ to produce $f^{(2)}(1)$. The process then continues from the point $(f(1), f^{(2)}(1))$.

With its foot anchored firmly at $(0,1)$, f stretches farther to the right as the scaling factor r decreases, and from FIGURE 1A it is clear that the iterates of f will race off to $+\infty$ until r reaches some critical value forcing f to touch the line $y=x$. When r attains this value and f is tangent to $y=x$, as in FIGURE 1B, the point of tangency (α, α) becomes the limit of the sequence of iterates. And as r drops below its critical value, f intersects $y=x$ in two points (FIGURE 1C), of which (β, β)—the left-most —is the limit point of the sequence of iterates. Note that α and β lie to the right of 1 on the x-axis because f increases from 1 on the y-axis.

Which graph optimally represents the function $f(x) = 1 + rx^p$ as we require it for our proof? We'd prefer that the iterates of f converge—which automatically rules out FIGURE 1A—and we'd like their limit to be the largest possible, which among the remaining choices puts FIGURE 1C out of the running. FIGURE 1B therefore gets the nod as the graph of f that best meets our needs. Given that f is tangent to $y=x$ at (α, α), clearly α is the largest number that bounds the iterates of f. Therefore α is the largest number that bounds the right side of inequality (2). We can be a bit more specific: At (α, α), we have $f'(\alpha) = 1$, which yields

$$\begin{aligned} \alpha &= (rp)^{1/(1-p)} \\ &= p^{1/(1-p)}/R. \end{aligned} \tag{4}$$

Equation (2) thereby becomes well-disciplined:

$$\begin{aligned} C_{k=N}^{n}(p, x_k) &< RM^{p^{-N}}\left(1 + r^{\cdot p}\left(1 + r^{\cdot p}\left(\cdots + r^{\cdot p}(1+r)\cdots\right)\right)\right) \\ &< RM^{p^{-N}}\left(p^{1/(1-p)}/R\right) \\ &= M^{p^{-N}} p^{1/(1-p)}. \end{aligned}$$

The last quantity above is composed entirely of known constants, and the truncated approximant C_N^n is therefore bounded for arbitrary $n \geq N$. Being increasing and bounded, the sequence of truncated approximants is convergent, and thus the continued power is convergent.

The only loose end to be taken care of is verifying the value of R given in the theorem. At (α, α) it is true that $f(\alpha) = \alpha$, so by substitution for α from equation (4) we get

$$1 + r(rp)^{p/(1-p)} = (rp)^{1/(1-p)},$$

which can be solved for r as follows:

$$\begin{aligned} 1 + r^{1/(1-p)}p^{p/(1-p)} &= r^{1/(1-p)}p^{1/(1-p)} \\ r^{1/(1-p)}p^{1/(1-p)}(1 - 1/p) &= 1 \\ r &= (1/p)[(p-1)/p]^{p-1} \\ &= (p-1)^{(p-1)}/p^p. \end{aligned}$$

Since $r = R^{p-1}$, we conclude that $R = (p-1)/p^{p/(p-1)}$, and the proof is complete.

REFERENCES

1. Aaron Herschfeld, On infinite radicals, *Amer. Math. Monthly* 42 (1935), 419–429.
2. Dixon J. Jones, Continued powers and roots, *Fibonnaci Quarterly* 29 (1991), 37–46.
3. G. Pólya and G. Szegö, *Problems and Theorems in Analysis*, Vol. 1, Springer-Verlag, New York, 1972.
4. Walter S. Sizer, Continued roots, this MAGAZINE 59 (1986), 23–27.

Computing the Fundamental Matrix for a Reducible Markov Chain

THEODORE J. SHESKIN
Industrial Engineering Department
Cleveland State University
Cleveland, OH 44115

1. Career opportunity at a consulting firm The primary purpose of this paper is to show new techniques for computing the fundamental matrix of a reducible Markov chain. We shall apply the techniques to a model of a career path at a consulting firm. Suppose a firm employs three types of consultants: probationary, associate, and partner. Probationary consultants are classified as grade 1, 2, 3, and 4. Entry-level is grade 1 probationary. Only probationary consultants can be promoted or discharged. Discharged employees are never rehired and consultants are never demoted. Associates are classified as associates in engineering, in computing, or in mathematics. All changes in job classification occur at the end of a year.

1. What is the expected length of time that a newly hired grade 1 consultant spends with the firm as a grade 2 consultant?
2. What is the expected length of time that a newly hired grade 1 consultant spends as a probationary consultant?
3. What is the probability that a grade 1 consultant will be promoted to partner?
4. In the long run what proportion of associate consultants will be associates in engineering?

To answer these questions we use the standard model: the career path of a consultant as a Markov chain.

2. Markov chain model A Markov chain **[2]**, **[5]** is a collection of random variables $\{X_t\}$ where the index t runs through the nonnegative integers. A state X_t represents the job classification of an employee at the end of year t. The conditional probabilities $P(X_{t+1} = j | X_t = i)$ are called transition probabilities. They do not change over time, that is, they are independent of t and are denoted by p_{ij}.

Our Markov chain model has 9 states indexed as follows.

9: discharged, 8: partner
7, 6, 5: associate in mathematics, computing, or engineering
4, 3, 2, 1: probationary consultant, grade 4, 3, 2, or 1,

which give rise to the transition matrix $P = [p_{ij}]$, $1 \le i, j \le 9$. As the p_{ij} are conditional probabilities, they must satisfy the properties

$$p_{ij} \ge 0, \quad \text{for all } i \text{ and } j.$$

$$\sum_{j=1}^{9} p_{ij} = 1, \quad \text{for all } i.$$

The directed graph in FIGURE 1 shows allowable transitions for the career of a consultant.

The states in our model are classified to describe the probabilistic behavior of employees. A state j is *accessible* from a state i if it is possible to enter j after one or more steps (time units), starting from i. In our model, states 5, 6, 7, 8, and 9 are accessible from states 1, 2, 3, and 4. If state j is accessible from state i and i is accessible from j, then states i and j are said to *communicate*. States 5 and 6 communicate. A set of states forms a *closed set* if no state outside the set is accessible from any state in the set. Once a process enters a closed set, it can never leave the set. If, in addition, each pair of states in a closed set communicate, the set is termed a *closed communicating class*. States 5, 6, and 7 form a closed communicating class. If a closed set contains only one state, that state is called an *absorbing state* and when the process enters an absorbing state, it can never leave. If i is an absorbing state, $p_{ii} = 1$. States 8 and 9 are absorbing states.

A Markov chain is *irreducible* if every state is accessible from every state. For an irreducible chain, a *steady state probability*, denoted by π_j, can be interpreted as the long run proportion of time that the process spends in state j. We can always find the values of π_j for an irreducible Markov chain ([**5**], pp. 151–152).

A nonirreducible Markov chain is termed *reducible*. Our consulting firm is modeled as a reducible Markov chain. In a reducible chain, any state that does not belong to a closed communicating class is called a *transient state*. If a Markov chain starts in a transient state, the chain is certain eventually to enter some closed communicating class ([**2**], p. 43). If a state is not transient, it is called *recurrent*. Starting from a recurrent state, eventual return to this state is certain. Our model has three classes of states: $\{8, 9\}$ absorbing states, $\{5, 6, 7\}$ recurrent states, and $\{1, 2, 3, 4\}$ transient states.

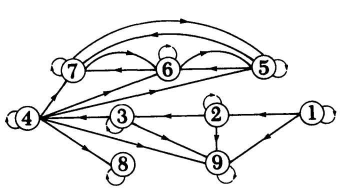


FIGURE 1

3. Canonical form of the transition matrix To answer our questions we arrange the transition matrix P in *canonical form*. Generally, states belonging to a communicating class are numbered consecutively. In a reducible chain with a closed communicating set of recurrent states and a set of transient states, we write P as the partitioned matrix

$$P = \begin{bmatrix} S & O \\ E & Q \end{bmatrix}, \tag{1}$$

where the square submatrix S is the transition matrix corresponding to the closed communicating class and the submatrix E specifies transitions from transient states to recurrent states. The submatrix Q governs transitions among the transient states.

4. State reduction The matrix construction algorithm is based on state reduction, an iterative procedure in which each iteration produces a reduced matrix one state smaller than its predecessor, resulting in a final reduced matrix from which the solution to the original problem can be obtained. Other state reduction algorithms are described in references [**1**], [**3**], and [**6**].

5. Probabilistic motivation The probabilistic motivation for state reduction is the following result in Kemeny and Snell ([**2**], pp. 114–115). Suppose that we have a Markov chain with an $n \times n$ transition matrix F partitioned as

$$F = \begin{array}{c c} & \begin{array}{cc} C & Y \end{array} \\ \begin{array}{c} C \\ Y \end{array} & \begin{bmatrix} Z & W \\ V & X \end{bmatrix} \end{array}. \tag{2}$$

Assume that we observe the process only when it is in a subset C of the states having c elements. A new Markov chain with c states, which we call a reduced process, is obtained. A single step in the reduced process corresponds in the original process to the transition, not necessarily in one step, from a state in C to another state in C. We compute the transition matrix D for the reduced process. Matrix Z describes transitions within C and has dimensions $c \times c$. Matrix W, with dimensions $c \times (n-c)$, governs transitions from C to Y, the subset of states outside of C. Matrix V describes transitions from Y to C and has dimensions $(n-c) \times c$. Matrix X, which governs transitions within Y, is $(n-c) \times (n-c)$. Let $Z = [z_{ij}]$, $W = [w_{ij}]$, $V = [v_{ij}]$, and $X = [x_{ij}]$. Let k and l be two states of C. We define

$$\begin{aligned} d_{kl} &= z_{kl} + \sum_{g,\, h \in Y} w_{kg}\left[x_{gh}^0 v_{hl} + x_{gh}^1 v_{hl} + x_{gh}^2 v_{hl} + \cdots\right] \\ &= z_{kl} + \sum_{g,\, h \in Y} w_{kg}\left[1 + x_{gh} + x_{gh}^2 + \cdots\right] v_{hl} \\ &= z_{kl} + \sum_{g,\, h \in Y} w_{kg}\left[1 - x_{gh}\right]^{-1} v_{hl} \quad \text{for } 1 \leqslant k, l \leqslant c. \end{aligned}$$

In matrix form we have

$$D = Z + W[I - X]^{-1} V \tag{3}$$

where I is the $(n-c) \times (n-c)$ identity matrix. Note D is $c \times c$.

6. Steady state probabilities for recurrent states Since our model of the consulting firm is a reducible Markov chain, we arrange the transition probability matrix P in canonical form.

$$P = \begin{array}{c c} & \begin{array}{ccccccccc} 9 & 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \end{array} \\ \begin{array}{c} 9 \\ 8 \\ 7 \\ 6 \\ 5 \\ 4 \\ 3 \\ 2 \\ 1 \end{array} & \left[\begin{array}{c|c|ccc|cccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0.50 & 0.30 & 0.20 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.30 & 0.45 & 0.25 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.10 & 0.35 & 0.55 & 0 & 0 & 0 & 0 \\ \hline 0.20 & 0.16 & 0.04 & 0.03 & 0.02 & 0.55 & 0 & 0 & 0 \\ 0.15 & 0 & 0 & 0 & 0 & 0.20 & 0.65 & 0 & 0 \\ 0.10 & 0 & 0 & 0 & 0 & 0 & 0.15 & 0.75 & 0 \\ 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0.10 & 0.85 \end{array}\right] \end{array}$$

The submatrix Q of transient states appears in the lower right-hand corner. The submatrix S of recurrent states represents an irreducible Markov chain.

$$Q = \begin{array}{c c} & \begin{array}{cccc} 4 & 3 & 2 & 1 \end{array} \\ \begin{array}{c} 4 \\ 3 \\ 2 \\ 1 \end{array} & \begin{bmatrix} 0.55 & 0 & 0 & 0 \\ 0.20 & 0.65 & 0 & 0 \\ 0 & 0.15 & 0.75 & 0 \\ 0 & 0 & 0.10 & 0.85 \end{bmatrix} \end{array},$$

$$S = \begin{array}{c} \\ 7 \\ 6 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccc} 7 & 6 & 5 \end{array} \\ \left|\begin{array}{ccc} 0.50 & 0.30 & 0.20 \\ 0.30 & 0.45 & 0.25 \\ 0.10 & 0.35 & 0.55 \end{array}\right| \end{array}.$$

To answer question 4 we solve the system of linear equations

$$\pi = \pi S,$$
$$1 = \pi_5 + \pi_6 + \pi_7,$$

to compute the steady state probability vector [5] for S,

$$\pi = \begin{array}{c} \begin{array}{ccc} 7 & 6 & 5 \end{array} \\ [0.291, 0.373, 0.336] \end{array}.$$

In the long run, 0.336 of the associate consultants will be in state 5, engineering.

7. Matrix construction algorithm The matrix construction algorithm we present is new. It contains two steps: *matrix augmentation* and *matrix reduction*. In the *matrix augmentation* step the transition matrix is truncated to obtain a submatrix Q governing transitions among the transient states. A null matrix and two identity matrices are adjoined to Q to form an augmented matrix B. In the *matrix reduction* routine, B is reduced to produce a matrix B_n that is the same size as Q.

In the matrix augmentation step, we assume that the transition matrix P for a reducible Markov chain is partitioned as in (1). For any reducible Markov chain the inverse of the matrix $(I - Q)$ exists and is called the *fundamental matrix*, N (see [2]). To construct an augmented matrix B of order $2n$ we assume that Q is $n \times n$. We adjoin a null matrix O, and two identity matrices I, each of order n, to Q. We arrange the augmented $2n \times 2n$ matrix in the form

$$B = \begin{bmatrix} O & I \\ I & Q \end{bmatrix}. \tag{4}$$

We let $B = [b_{ij}]$ and $Q = [q_{ij}]$.

The detailed steps of matrix reduction applied to the augmented matrix B are presented below.

1. Initialize $k = 2n$.
2. Let $B_k = B$.
3. Partition B_k as

$$B_k = \begin{array}{c} \begin{array}{cc} k-1 & 1 \end{array} \\ \left[\begin{array}{c|c} T_k & U_k \\ \hline R_k & Q_k \end{array}\right] \end{array} \begin{array}{c} \\ k-1 \\ 1 \end{array}.$$

4. $B_{k-1} = T_k + U_k[I - Q_k]^{-1}R_k$ where $[I - Q_k]^{-1} = (1 - b_{kk})^{-1}$.
5. Decrement k by 1. If $k = n$, stop. Otherwise, repeat steps 3 and 4.

Matrix reduction ends when $k = n$, indicating that the final reduced matrix, B_n, is of order n. The calculation of the fundamental matrix is based on the following application of equation (3): When a matrix B is partitioned as in (4), then

$$B_n = O + I[I - Q]^{-1}I = [I - Q]^{-1} = N. \tag{5}$$

Therefore, B_n is the fundamental matrix.

8. Operation count Each step of matrix reduction performs $(n-1)^2$ additions, 1 subtraction, (n^2-1) multiplications, and 1 division. Multiplying these numbers by n steps, we have (n^3-2n^2+2n) additions and subtractions and n^3 multiplications and divisions. If we count only multiplications and divisions, then the number of operations is approximately n^3. A standard method for inverting a matrix such as the Crout LU decomposition also performs approximately n^3 operations [4].

9. Analysis of consulting opportunity By applying results derived by Kemeny and Snell for the fundamental matrix of a reducible Markov chain, we answer the first three questions.

$$B = \begin{array}{c} \\ 8 \\ 7 \\ 6 \\ 5 \\ 4 \\ 3 \\ 2 \\ 1 \end{array} \begin{array}{c} \begin{array}{cccccccc} 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \end{array} \\ \left[\begin{array}{cccc|cccc} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ \hline 1 & 0 & 0 & 0 & 0.55 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0.20 & 0.65 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0.15 & 0.75 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0.10 & 0.85 \end{array}\right] \end{array} = \begin{bmatrix} O & I \\ I & Q \end{bmatrix}.$$

$$B_8 = \begin{array}{c} \\ 8 \\ 7 \\ 6 \\ 5 \\ 4 \\ 3 \\ 2 \\ 1 \end{array} \begin{array}{c} \begin{array}{cccccccc} 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \end{array} \\ \left[\begin{array}{ccccccc|c} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0.55 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0.20 & 0.65 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0.15 & 0.75 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 & 0 & 0.10 & 0.85 \end{array}\right] \end{array} = \begin{bmatrix} T_8 & U_8 \\ R_8 & Q_8 \end{bmatrix}.$$

$$B_7 = T_8 + U_8[I - Q_8]^{-1} R_8.$$

$$\vdots$$

$$B_4 = T_5 + U_5[I - Q_5]^{-1} R_5 = N = \begin{array}{c} \\ 4 \\ 3 \\ 2 \\ 1 \end{array} \begin{array}{c} \begin{array}{cccc} 4 & 3 & 2 & 1 \end{array} \\ \begin{bmatrix} 2.222 & 0 & 0 & 0 \\ 1.270 & 2.857 & 0 & 0 \\ 0.762 & 1.714 & 4 & 0 \\ 0.508 & 1.143 & 2.667 & 6.667 \end{bmatrix} \end{array}.$$

To answer question 1, if a process is presently in a transient state i, the expected number of periods that will be spent in a transient state j before entry into an absorbing or a recurrent state is the (i,j)th element of the fundamental matrix. Therefore, the expected time that a grade 1 consultant spends with the firm as a grade 2 consultant is 2.667 years, element $(1,2)$ of N.

To answer question 2, if a process is presently in a transient state i, the expected number of periods that will be spent in all transient states before entry into an absorbing or a recurrent state is the sum of the elements in the ith row of the fundamental matrix. The expected time that a grade 1 consultant spends as a probationary consultant is equal to the sum of the expected times that she spends in grades 1 through 4, or $6.667 + 2.667 + 1.143 + 0.508 = 10.985$ years.

To answer question 3, if a process is presently in a transient state i, the probability of eventual absorption in an absorbing state j is the (i, j)th element of the matrix NE_1, where E_1 is the submatrix governing transitions from transient states to absorbing states. We have

$$NE_1 = \begin{array}{cc} & \begin{array}{cccc} 4 & 3 & 2 & 1 \end{array} \\ \begin{array}{c} 4 \\ 3 \\ 2 \\ 1 \end{array} & \begin{bmatrix} 2.222 & 0 & 0 & 0 \\ 1.270 & 2.857 & 0 & 0 \\ 0.762 & 1.714 & 4 & 0 \\ 0.508 & 1.143 & 2.667 & 6.667 \end{bmatrix} \end{array} \begin{array}{c} \begin{array}{cc} 9 & 8 \end{array} \\ \begin{bmatrix} 0.20 & 0.16 \\ 0.15 & 0 \\ 0.10 & 0 \\ 0.05 & 0 \end{bmatrix} \end{array} = \begin{array}{cc} & \begin{array}{cc} 9 & 8 \end{array} \\ \begin{array}{c} 4 \\ 3 \\ 2 \\ 1 \end{array} & \begin{bmatrix} 0.444 & 0.356 \\ 0.683 & 0.203 \\ 0.810 & 0.122 \\ 0.873 & 0.081 \end{bmatrix} \end{array}.$$

The probability that a grade 1 consultant will be promoted to partner is 0.081, element $(1, 8)$ of NE_1.

The matrix construction algorithm is an interesting new procedure for calculating the fundamental matrix for a reducible Markov chain.

REFERENCES

1. W. K. Grassmann, M. I. Taksar and D. P. Heyman, Regenerative analysis and steady state distributions for Markov chains, *Opns. Res.*, 33 (1985), 1107–1116.
2. J. G. Kemeny and J. L. Snell, *Finite Markov Chains*, Springer-Verlag, New York, 1976.
3. R. Lal and U. N. Bhat, Reduced system algorithms for Markov chains, *Mgmt Science*, 34 (1988), 1202–1220.
4. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes*, Cambridge University Press, Cambridge, U.K., 1986.
5. S. Ross, *Introduction to Probability Models*, 4th edition, Academic Press, New York, 1989.
6. T. J. Sheskin, A Markov chain partitioning algorithm for computing steady state probabilities, *Opns. Res.*, 33 (1985), 228–235.
7. T. J. Sheskin, Computing the fundamental matrix for a reducible Markov chain, poster paper, 1st International Workshop on the Numerical Solution of Markov Chains, North Carolina State University, Raleigh, NC (1990).

Baltimore, January 1992. Members of the Mathematics Magazine Editorial Board: Front row (L to R), Dianne McCann, Susanna Epp, Dan Kalman, Ken Ross, Martha Siegel. Back Row, Loren Larson, George Gilbert, David James, Bruce Reznick, Paul Campbell.

PROBLEMS

LOREN LARSON, *editor*
St. Olaf College

GEORGE T. GILBERT, *assistant editor*
Texas Christian University

Proposals

To be considered for publication, solutions should be received by May 1996.

1484. *Proposed by Lenny Jones, Shippensburg University, Shippensburg, Pennsylvania.*

Let $\sigma(n)$ be the sum of the positive divisors of the positive integer n and let $\phi(n)$ be Euler's totient function. For an arbitrary positive integer k, find all positive integers n that satisfy

$$n^k\sigma(n) \equiv 2 \pmod{\phi(n)}.$$

1485. *Proposed by Yasutoshi Nomura, Hyogo University of Teacher Education, Hyogo, Japan.*

Let $n > 1$ be a natural number and consider the statement Q_n:

There exist positive integers $x_1, x_2, \ldots, x_n$ for which the arithmetic-geometric-mean quotient $\dfrac{x_1^n + \cdots + x_n^n}{n x_1 \cdots x_n}$ is an integer greater than 1.

(a) Show that Q_2 is false.
(b) Show that Q_n is true for even $n > 2$ or for prime n congruent to 5 modulo 6.
(c)* Find another n for which Q_n is false or an infinite family for which it is true.

1486. *Proposed by Paul Bracken, University of Waterloo, Waterloo, Ontario, Canada.*

For $-1 < x$, $x \neq 0$, define the sequence $\theta_n(x)$ by

$$\log(1+x) = x - \frac{x^2}{2} + \cdots + (-1)^{n-1}\frac{\theta_n(x)x^n}{n}.$$

Show that the sequence (θ_n) is monotonic in n and find its limit.

ASSISTANT EDITORS: CLIFTON CORZAT, BRUCE HANSON, RICHARD KLEBER, KAY SMITH, *and* THEODORE VESSEY, *St. Olaf College and* MARK KRUSEMEYER, *Carleton College. We invite readers to submit problems believed to be new and appealing to students and teachers of advanced undergraduate mathematics. Proposals should be accompanied by solutions, if at all possible, and by any other information that will assist the editors and referees. A Problem submitted as a Quickie should have an unexpected succint solution. An asterisk (*) next to a problem number indicates that neither the proposer nor the editors supplied a solution.*

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

Solutions and new proposals should be sent to George T. Gilbert, Problems Editor-Elect, Department of Mathematics, Box 32903, Texas Christian University, Fort Worth, TX 76129 or mailed electronically to g.gilbert@tcu.edu. Electronic submission of TeX input files is acceptable. Readers who use e-mail should also provide an e-mail address.

1487. *Proposed by Edward Kitchen, Santa Monica, California.*

Given circles $\mathscr{C}$ and $\mathscr{C}'$ with centers O and O', and circles $\mathscr{C}_1$ and $\mathscr{C}_2$ externally tangent to $\mathscr{C}$ at points M_1 and M_2 and internally tangent to $\mathscr{C}'$ at points N_1 and N_2, prove that the lines M_1N_1, M_2N_2, and OO' are concurrent.

1488. *Proposed by Heinz-Jürgen Seiffert, Berlin, Germany.*

Let n be a positive integer. Show that if $0 < x_1 \le x_2 \le \cdots \le x_n$, then

$$\left(\prod_{i=1}^{n}(1+x_i)\right)\left(\sum_{j=0}^{n}\prod_{k=1}^{j}\frac{1}{x_k}\right) \ge 2^n(n+1),$$

with equality if, and only if, $x_1 = x_2 = \cdots = x_n = 1$. (Empty products are understood to be unity.)

Quickies

Answers to the Quickies are on page 406.

Q841. *Proposed by Murray S. Klamkin, University of Alberta, Edmonton, Alberta, Canada, and Stanley Rabinowitz, MathPro Press, Westford, Massachusetts.*

Prove that the sequence $u_n = 1/n$, $n = 1, 2, \ldots$, cannot be the solution of a nonhomogeneous linear finite-order difference equation with constant coefficients.

Q842. *Proposed by Ruby Rose, Pacific Lutheran University, Tacoma, Washington.*

Let A and B be two 2×2 matrices. Prove that if AB is a linear combination of I, A, and B, then so is BA.

Q843. *Proposed by Murray S. Klamkin, University of Alberta, Edmonton, Alberta, Canada.*

Evaluate the following two $n \times n$ determinants:

(i) $D_1 = \det(a_{rs})_{r,s=1,\ldots,n}$, where $a_{rr} = a_r$, $r = 2, \ldots, n$, and all the remaining elements are 1;

(ii) $D_2 = \det(b_{rs})_{r,s=1,\ldots,n}$, where $b_{rr} = b_r$, $r = 1, 2, \ldots, n$, and all remaining elements are 1.

Solutions

Newman-Conway sequence **December 1994**

1459. *Proposed by D. M. Bloom, Brooklyn College of CUNY, Brooklyn, New York.*

The now notorious Newman-Conway sequence $1, 1, 2, 2, 3, 4, 4, 4, 5, 6, 7, 7, \ldots$ is defined by the recurrence $P(1) = P(2) = 1$, $P(n) = P(P(n-1)) + P(n - P(n-1))$, $(n \ge 3)$. Richard Guy, in this MAGAZINE, February 1990, p. 17, wrote: "I have an earlier manuscript of Conway in which he has written '$P(2^k) = 2^{k-1}$ (easy), $P(2n) \le 2P(n)$ (hard), ... '." Prove Conway's "hard" inequality: $P(2n) \le 2P(n)$.

Solution by the proposer.

First, it is easy to show by induction that $P(n+1)-P(n)=0$ or 1 for all n (so that $P(n)\le n$ when $n\ge 1$). Next, suppose $P(2n)>2P(n)$ for some n, and let n be the smallest such. Then $P(2(n-1))\le 2P(n-1)$, and since P increases only in steps of 0 or 1, it follows that $P(n-1)=P(n)$ (say $=r$).

Clearly $n\ge 3$, so that by the recurrence condition for P,

$$P(n)=P(P(n-1))+P(n-P(n-1))=P(r)+P(n-r).$$

If $P(2n-1)=2r$, then by the recurrence,

$$P(2n)=P(2r)+P(2n-2r)\le 2P(r)+2P(n-r)=2P(n),$$

contrary to hypothesis. If instead, $P(2n-1)=2r+1$, then $P(2n-2)\le 2P(n-1)=2r$ so that

$$\begin{aligned}2r+1=P(2n-1)&=P(P(2n-2))+P(2n-1-P(2n-2))\\&=P(2r)+P(2n-1-2r)\le P(2r)+P(2n-2r)\\&\le 2P(r)+2P(n-r)=2P(n)=2r,\end{aligned}$$

again a contradiction.

Hence $P(2n)>2P(n)$ is impossible and the proof is complete.

Also solved by the Con Amore Problem Group (Denmark), J. S. Frame, and Achilleas Sinefakopoulos (student, Greece).

An orthic triangle inequality **December 1994**

1460. *Proposed by Doru Popescu Anastasiu, Slatina, Romania.*

Let ABC be an acute triangle with altitudes AA', BB', CC'. Let A_1, B_1, C_1 be the second intersection points of lines AA', BB', CC' with the circumcircle of triangle ABC. Show that

$$AA_1^2\sin 2A+BB_1^2\sin 2B+CC_1^2\sin 2C>24S_0,$$

where S_0 denotes the area of triangle $A'B'C'$.

Solution by Lajos Csete, Markotabödöge, Hungary.

We will show more generally, that for any integer $n>1$,

$$AA_1^n\sin 2A+BB_1^n\sin 2B+CC_1^n\sin 2C\ge 2^{2n+1}(R\rho_0)^{n/2-1}S_0,$$

where ρ_0 denotes the radius of the incircle of $A'B'C'$. In the case $n=2$,

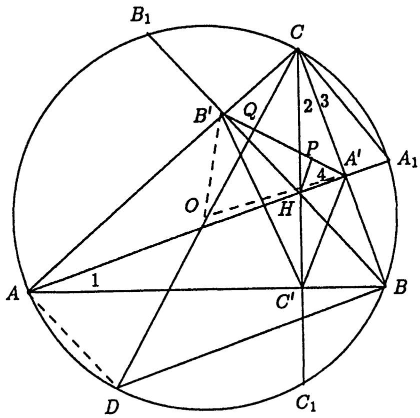
$$AA_1^2\sin 2A+BB_1^2\sin 2B+CC_1^2\sin 2C\ge 32S_0.$$

We begin by considering the following figure: H is the orthocenter, O is the circumcenter, CD is a diameter of the circumcircle of ABC of radius R, and P is on $A'B'$ so that HP is perpendicular to $A'B'$; S denotes the area of triangle ABC, and p_0 is the perimeter of $A'B'C'$.

We will need the following facts ((i)–(iv) have obvious analogues to vertices B and C):

(i) $AH=2R\cos A,\ BC=2R\sin A,$
(ii) $\rho_0=HA'\cos A,$
(iii) $AH\cdot HA'=2R\rho_0,$
(iv) $HA'=A'A_1,$

(v) $S = 2R^2 \sin A \sin B \sin C$,
(vi) $p_0 = 4R \sin A \sin B \sin C$,
(vii) $\sin 2A + \sin 2B + \sin 2C = 4 \sin A \sin B \sin C$.



Proofs: AH is parallel to DB because AA_1 and DB are both perpendicular to BC; DA is parallel to BH because DA and BB_1 are both perpendicular to AC. Therefore $AHBD$ is a parallelogram and $AH = DB = 2R\cos D = 2R\cos A$. Also, $BC = 2R\sin D = 2R\sin A$. This proves (i).

$CA'HB'$ is a cyclic quadrilateral, and therefore $\angle ACH = \angle 4$ $(= \angle HA'B')$. Since $\angle ACH$ and $\angle A$ are complementary, so are $\angle 4$ and $\angle A$. It follows that $\rho_0 = HP = HA' \sin\angle 4 = HA' \cos A$. This proves (ii).

(iii) follows directly from (i) and (ii).

$\angle 1 = \angle 2$ since their corresponding sides are perpendicular to one another. $\angle 1 = \angle 3$ since both cut off the same arc in the circumcircle. Therefore $\angle 2 = \angle 3$ and, since AA_1 is perpendicular to BC, triangle CHA' is congruent to CA_1A'. From this it follows that $HA' = A'A_1$. This proves (iv).

$S = \frac{1}{2}BC \cdot AC \sin C = \frac{1}{2}(2R\sin A)(2R\sin B)\sin C$ and this leads to (v).

$\angle CA'B = \angle A$ since we have seen that $\angle 4$ is complementary to $\angle A$. Also, $\angle DCB$ is complementary to $\angle A$ since $\angle D = \angle A$. Therefore $\angle CA'Q$ and $\angle DCA'$ are complementary. It follows that CD is perpendicular to $A'B'$. Thus the area of $CB'OA'$ is $\frac{1}{2}RA'B'$. Similarly, the area of $BA'OC'$ is $\frac{1}{2}RC'A'$, and the area of $AC'OB'$ is $\frac{1}{2}RB'C'$. Adding these together yields $S = \frac{1}{2}Rp_0$. This, together with (v) yields (vi).

The area of $ABHC$ is the same as $ABC - HBC = \frac{1}{2}BC(AA' - HA') = \frac{1}{2}BC \cdot AH = \frac{1}{2}(2R\sin A)(2R\cos A) = R^2 \sin 2A$. In a similar way, the areas of $BCHA$ and $CAHB$ are $R^2 \sin 2B$ and $R^2 \sin 2C$. Adding these we have

$$\begin{aligned} &R^2(\sin 2A + \sin 2B + \sin 2C) \\ &\quad = [ABHC] + [BCHA] + [CAHB] \\ &\quad = ([ABC] - [HBC]) + ([ABC] - [HCA]) + ([ABC] - [HAB]) \\ &\quad = 2[ABC]. \end{aligned}$$

This, together with (v), yields (vii).

Using (i), (ii), and (iv), together with the arithmetic-geometric mean inequality, we have

$$AA_1^n = \left(2R\cos A + \frac{2\rho_0}{\cos A}\right)^n$$

$$\ge 2^n\left(2R\cos A\frac{2\rho_0}{\cos A}\right)^{n/2}$$
$$= 4^n(R\rho_0)^{n/2},$$

with equality if, and only if, $AH = HA_1$. Hence

$$\begin{aligned}AA_1^n\sin 2A + BB_1^n\sin 2B + CC_1^n\sin 2C &\ge 4^n(R\rho_0)^{n/2}(\sin 2A + \sin 2B + \sin 2C)\\ &= 4^n(R\rho_0)^{n/2}(4\sin A\sin B\sin C)\\ &= 4^n(R\rho_0)^{n/2-1}(\rho_0)(4R\sin A\sin B\sin C)\\ &= 4^n(R\rho_0)^{n/2-1}\rho_0 p_0\\ &= 4^n(R\rho_0)^{n/2-1}2S_0\\ &= 2^{2n+1}(R\rho_0)^{n/2-1}S_0,\end{aligned}$$

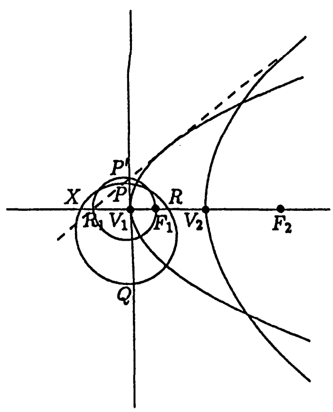
with equality if, and only if, ABC is equilateral.

Also solved by the Con Amore Problem Group (Denmark), Miguel Amengual Covas (Spain), Jiro Fukuta (Japan), Mostafa Ghandehari, Nick Lord (England), O. P. Lossers (The Netherlands), Francisco Bellot Rosado (Spain), Hoe Teck Wee (student, Singapore), Sammy and Jimmy Yu (students), David Zhu, and the proposer.

Tangent to parabolas with ruler and compass **December 1994**

1461. *Proposed by Václav Konečný, Ferris State University, Big Rapids, Michigan.*

Suppose we are given two parabolas with the same axes, having vertices at V_1 and V_2 and foci at F_1 and F_2, respectively. Construct a common tangent, if one exists,



using only a compass and straightedge. Assume that the unit of length is given.

Solution by Nick Lord, Tonbridge School, Kent, England.

With x-axis as common axis, we may take the parabolas to be $y^2 = 4ax$ and $y^2 = 4A(x-b)$. The tangent to $y^2 = 4ax$ at $(at^2, 2at)$ is $ty = x + at^2$ and meets $y^2 = 4A(x-b)$ at the point whose x-coordinate satisfies

$$(x + at^2)^2 = t^2 4A(x-b),$$

or equivalently,

$$x^2 + 2t^2(a-2A)x + t^2(a^2t^2 + 4Ab) = 0.$$

For a common tangent, this has repeated roots so

$$4(a-2A)^2t^4 = 4t^2(a^2t^2+4Ab)$$

and

$$t^2 = \frac{b}{A-a}. \tag{1}$$

For a solution, $b>0$, $A>a$, or, $b<0$, $A<a$. We give a construction in the former case; the latter is similar.

Consider the tangent with slope $\sqrt{\frac{A-a}{b}}$ that goes through $\left(-\frac{ab}{A-a},0\right)$ on the x-axis and $\left(0, a\sqrt{\frac{b}{A-a}}\right)$ on the y-axis.

Steps of construction:

1. Construct the "y-axis," perpendicular to the x-axis through V_1.
2. Locate X on the negative x-axis with $F_1X = F_2V_2$.
3. Locate P on the positive y-axis, and Q on the negative y-axis, with $V_1P = V_1F_1$ and $V_1Q = V_1V_2$.
4. Let R be the second point of intersection of the circumcircle of PXQ with the x-axis, and let R_1 be on the x-axis with $RV_1 = V_1R_1$.
5. From the circle with diameter F_1R_1, locate P' on the positive y-axis.
6. Then R_1P' gives the required tangent.

In coordinates, $F_1(a,0)$, $F_2(A+b,0)$, $V_1(0,0)$, $V_2(b,0)$. Then $XV_1 = A-a$ and $(V_1R)(XV_1) = (V_1P)(V_1Q)$ imply that $V_1R = ab/(A-a)$. Also, $(V_1P')^2 = (V_1R_1)(V_1F_1) = a^2b/(A-a)$ implies $V_1P' = a\sqrt{b/(A-a)}$. Thus R_1P' goes through $(-ab/(A-a),0)$ with slope $\sqrt{(A-a)/b}$ as required.

Also solved by H. Guggenheimer, Adam and Po Kee Wong, and the proposer.

Roots of a fourth-degree polynomial **December 1994**

1462. *Proposed by Arthur L. Holshouser and Benjamin G. Klein, Davidson College, Davidson, North Carolina.*

Let λ be a given positive number. Let $p(x) = ax^2+bx+c$, where a, b, and c are nonzero real numbers. Assume that no roots of $p(x)$ lie on the line $\Lambda = \{t+i\lambda t : t\in\mathbf{R}\}$. Find a homogeneous fourth-degree polynomial $H_\lambda(a,b,c)$ such that the number of roots of $p(x)$ that lie below Λ is given by

$$1 - \frac{\operatorname{sgn}(ab)+\operatorname{sgn}(H_\lambda(a,b,c))}{2}.$$

Solution by Michael H. Andreoli, Miami Dade Community College (North), Miami, Florida.

One such polynomial is $H_\lambda = b^3c - 4abc^2 + \lambda^2b^3c = (b^2-4ac+\lambda^2b^2)bc$.

Define the quantities:

$$M = \text{the number of roots of } P(x) \text{ that lie below } \Lambda$$

$$Q = b^2 - 4ac + \lambda^2 b^2, \text{ so that } H_\lambda = Q \cdot bc$$
$$N = 1 - \tfrac{1}{2}(\operatorname{sgn}(ab) + \operatorname{sgn}(H_\lambda)).$$

The quadratic formula gives six cases that determine M:

(i) $4ac < 0$, in which case $M = 1$.
(ii) $0 < 4ac \le b^2$ and $\operatorname{sgn}(ab) = -1$, in which case $M = 2$.
(iii) $0 < 4ac \le b^2$ and $\operatorname{sgn}(ab) = 1$, in which case $M = 0$.
(iv) $4ac > b^2$ and $4ac - b^2 > \lambda^2 b^2$, in which case $M = 1$.
(v) $4ac > b^2$ and $4ac - b^2 < \lambda^2 b^2$ and $\operatorname{sgn}(ab) = 1$, in which case $M = 0$.
(vi) $4ac > b^2$ and $4ac - b^2 < \lambda^2 b^2$ and $\operatorname{sgn}(ab) = -1$, in which case $M = 2$.

Note also that

(a) $N = 0$ if, and only if, $\operatorname{sgn}(ab) = \operatorname{sgn}(H_\lambda) = 1$.
(b) $N = 1$ if, and only if, $\operatorname{sgn}(ab) = -\operatorname{sgn}(H_\lambda)$.
(c) $N = 2$ if, and only if, $\operatorname{sgn}(ab) = \operatorname{sgn}(H_\lambda) = -1$.

In case (i), $\operatorname{sgn}(Q) = 1$ and $\operatorname{sgn}(ac) = -1$. Therefore, $\operatorname{sgn}(bc) = -\operatorname{sgn}(ab)$, and so $\operatorname{sgn}(H_\lambda) = \operatorname{sgn}(Q) \cdot \operatorname{sgn}(ab)$. It follows from (b) that $N = 1$.

In the remaining cases, $\operatorname{sgn}(ac) = 1$, so that $\operatorname{sgn}(bc) = \operatorname{sgn}(ab)$.

Therefore, $\operatorname{sgn}(H_\lambda) = \operatorname{sgn}(Q) \cdot \operatorname{sgn}(bc) = \operatorname{sgn}(Q) \cdot \operatorname{sgn}(ab)$. It is straightforward to verify that in each case $M = N$.

Also solved by the proposers.

Partition identity **December 1994**

1463. *Proposed by Christos Athanasiadis, Massachusetts Institute of Technology, Cambridge, Massachusetts.*

Let n be a positive integer. For any partition λ of n and any $1 \le i \le n$, let $m_i = m_i(\lambda)$ be the number of i's in λ. Also let z_λ be the quantity $1^{m_1} m_1!\, 2^{m_2} m_2! \cdots n^{m_n} m_n!$. Show that, for $1 \le k \le n$,

$$\sum_{\lambda \vdash n} \frac{m_k(\lambda)}{z_\lambda} = \frac{1}{k}$$

where, in the summation, λ runs over all partitions of n.

Solution by O. P. Lossers, Technical University Eindhoven, Eindhoven, The Netherlands.

For a partition λ of n the number of permutations π in $\operatorname{Sym}(n)$ with cycle structure λ equals $n!/z_\lambda$. (This is well known and not difficult to prove: Start with a linear ordering on the symbols $1, 2, \ldots, n$; this can be done in $n!$ ways. Then put a pair of brackets around the first m_1 singletons, then around the following m_2 pairs, m_3 triples, and so on. This defines a permutation with cycle structure λ, and the number of linear orders giving rise in this way to this particular permutation is exactly z_λ.)

Now we count in two ways the number of (ordered) pairs (π, γ), where $\pi \in \operatorname{Sym}(n)$, and γ is a k-cycle occurring in the cyclic representation of π.

The total number of possible k-cycles on n symbols is

$$\frac{n!}{k(n-k)!},$$

and each one of them occurs in $(n-k)!$ permutations. So counting the number of pairs in this way we get $n!/k$. On the other hand, a permutation π with cycle structure λ occurs in $m_k(\lambda)$ pairs, so adding up and combining we find

$$\sum_{\lambda \vdash n} m_k(\lambda)\frac{n!}{z_\lambda} = \frac{n!}{k}.$$

Also solved by J. C. Binz (Switzerland), David Callan, Richard Holzsager, Nick Lord (England), José Heber Nieto (Venezuela), Michael Reid, Dennis Walsh, Hoe Teck Wee (student, Singapore), and the proposer.

Answers

Solutions to the Quickies on page 400.

A841. Assume to the contrary that it is possible. Then there exist constants a_i, not all zero, such that

$$a_0 + \frac{a_1}{n} + \frac{a_2}{n+1} + \cdots + \frac{a_r}{n+r-1} = 0 \tag{1}$$

for $n = 1, 2, \ldots$. It then follows that the left-hand side of (1), which is a rational function of n, must identically vanish for all n. Letting $n \to 0$, it follows that $a_1 = 0$. Then letting $n \to -1$, it follows that $a_2 = 0$, and similarly, all the a_i are zero, and this is a contradiction.

In a similar way, it follows that no strictly rational function can be the solution of a linear finite-order difference equation with constant coefficients.

A842. Let $P_D(x)$ be the characteristic polynomial of the matrix D. If D is a 2×2 matrix then $P_D(x)$ is a quadratic, and since $P_D(D) = 0$ by the Cayley-Hamilton Theorem, $D^2 = d_1 D + d_2 I$, for some constants d_1 and d_2. In particular, $A^2 = a_1 A + a_2 I$, $B^2 = b_1 B + b_2 I$, and $(A+B)^2 = c_1(A+B) + c_2 I$. Assume that $AB = aA + bB + cI$. Then,

$$\begin{aligned} c_1(A+B) + c_2 I &= (A+B)^2 = A^2 + B^2 + AB + BA \\ &= (a_1 A + a_2 I) + (b_1 B + b_2 I) + aA + bB + cI + BA, \end{aligned}$$

and the result follows.

A843. (i) On setting any $a_r = 1$, D_1 vanishes. Hence $(a_2 - 1)(a_3 - 1)\cdots(a_n - 1)$ is a factor of D_1. The remaining factor can only be the constant 1 since D is a polynomial in the a_i's with leading term $a_2 a_3 \ldots a_n$.

(ii) If $b_r = 1$ for some r, the determinant reduces to the case considered in (i), so we will assume that none of the b_r is equal to 1. Replace b_r by $x/x_r + 1$, so that

$$x_1 x_2 \cdots x_n D_2 = D_2' = \det(c_{rs}) \quad \text{where } c_{rr} = x + x_r \text{ and } c_{rs} = x_r \text{ for } r \neq s.$$

By setting $x = 0$ in D_2', we get n rows that are proportional. Hence x^{n-1} is a factor of D_2'. The other factor must be linear in x having the form $x + \lambda$ since the coefficient of x^n in D_2' must be 1. It is clear that $\lambda = \sum x_i$, since the coefficient of x^{n-1} can only come in from the main diagonal. Finally,

$$D_2' = \left(x + \sum_{i=1}^{n} x_i\right)x^{n-1} \quad \text{and} \quad D_2 = \left(1 + \sum_{i=1}^{n} \frac{1}{b_i - 1}\right)\prod_{i=1}^{n}(b_i - 1).$$

Alternatively, one can split the first row as $(1, 1, \ldots, 1) + (b_1 - 1, 0, \ldots, 0)$, and then use the linearity of the determinant in a row, part (a), and induction.

Comment Problem 1434 was also solved by the Anchorage Math Solutions Group, and Problem 1445 by Brian Wolk (student).

REVIEWS

PAUL J. CAMPBELL, *editor*
Beloit College

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

Saari, Donald G., A chaotic exploration of aggregation paradoxes, *SIAM Review* 37 (1) (March 1995) 37–52.

In this exciting article, author Saari provides a common explanation for voting paradoxes, nontransitive dice, and Simpson's paradox in statistics. He outlines an approach based on chaotic symbolic dynamics and symmetry groups, creating a mathematical theory of all paradoxes that can occur and their properties. He concludes that because the Borda count election method "allows shockingly fewer paradoxes" than other election methods (including the plurality voting used by SIAM, the single-transferable-vote method used by the AMS, and the approval voting used by the MAA), it "is, by far, the superior choice." The nonparametric Kruskal-Wallis rank test also "admits far more paradoxes and difficulties than previously imagined"; still, "of all possible nonparametric ranking methods, the KW test is by far the best choice."

Keller, Joseph B., How many shuffles to mix a deck?, *SIAM Review* 37 (1) (March 1995) 88–89. Mann, Brad, How many times should you shuffle a deck of cards?, *The UMAP Journal* 15 (4) (1994) 303–332; also in *Topics in Applied Probability and Its Applications*, ed. Laurie Snell, 1995; CRC Press, Boca Raton, FL. S. Brent Morris, Practitioner's commentary: Card shuffling, *The UMAP Journal* 15 (4) (1994) 333–338. Bayer, Dave, and Persi Diaconis, Trailing the dovetail shuffle to its lair, *Annals of Applied Probability* 2 (2) (1992) 294–13.

The "common wisdom"—at least since the researches of Persi Diaconis 10 years ago, publicized only more recently—is that seven riffle shuffles suffice to randomize a 52-card deck (in a riffle shuffle, the deck is cut into two possibly unequal packets, then the two packets are interleaved in any possible way). Mann's article is an exposition of the analysis by Diaconis and Dave Bayer. They fasten on seven as a convenient middle point between too few shuffles and diminishing returns from doing a great many more; but they also show that roughly $\frac{3}{2}\log_2 k$ shuffles of a k-card deck are necessary and sufficient to meet one of their criteria. For $k = 52$, this means 11–12 shuffles. Keller gives a much simpler argument to show that roughly $\log_2 k$ shuffles are needed just to "fix" the probability for the bottom card (make the probability close to $1/k$ that the original bottom card is on the bottom after the shuffling). While these authors regard the randomness of the shuffled deck as lying along a continuum, Morris focuses on discrete aspects of the perfect shuffle (riffle shuffles in which the deck is cut exactly in half and the cards perfectly interlaced); he describes applications to magic, dishonest gaming, efficient networks for parallel processors, and the design of dynamic computer memories, and he gives an extensive bibliography.

Solow, Daniel, *The Keys to Advanced Mathematics: Recurrent Themes in Abstract Reasoning*, BookMasters Distribution Center (1444 U.S. Route 42, RD # 11, Mansfield, OH 44903; (800) 247–6553); xx + 476 pp, $39.95 (P) + $4.95 shipping and handling. ISBN 0–9644519–0–5.

Solow's previous book, *How to Read and Do Proofs* (2nd ed., Wiley, 1990), is (in my opinion) the best of the books that try to teach students the ins and outs of proving; after working through the book, students never again are stuck about how to begin a proof. That book confined its examples to precalculus mathematics. His new book takes on the concepts in discrete mathematics, linear algebra, abstract algebra, and real analysis—not just doing proofs, but mastering combining special cases into a single framework; generalizing a problem; abstraction; identifying similarities and differences; abstracting from the specific to the general; translating between visual, verbal, and symbolic form; understanding and creating definitions; and working with axiomatic systems. This book "is designed for those who believe in separating the teaching of conceptual ideas from specific subject matter." It would be excellent for a one-semester "transition" course to advanced mathematics, as a supplement to a course in any of the four subjects featured, or as a self-study guide. Forty pages are devoted to complete solutions to the odd-numbered problems; a complete solutions manual is available.

Peterson, Ivars, Progressing to a set of consecutive primes, *Science News* 148 (9 September 1995) 167.

$1,089,533,431,247,059,310,875,780,378,922,957,732,908,036,492,993,138,195,385,213,$ $105,561,742,150,447,308,967,213,141,717,486,151 + 210k$ are primes for $k = 0,\dots,6$. This sequence of seven primes in arithmetic progression beats the old record of six. It was discovered by Harvey Dubner (semiretired electrical engineer, Westwood, NJ) and Harry L. Nelson (Lawrence Livermore National Laboratory, retired), using seven specially modified computers running for two weeks. Maybe they can go to eight, in a couple of computer-years. How high could they go? No one knows, as it is conjectured—but unproved—that there are arbitrarily long sequences of primes in arithmetic progression.

Asimov, Daniel, There's no space like home: Why look over the $\mathbb{R}^n$-bow? Three dimensions offer topological thrills galore, *The Sciences* (September-October 1995) 20–25.

"You might expect that as the number of dimensions goes up, space gets stranger and more interesting. Not so." Asimov contends that three-space "is the first Euclidean space in which it is possible to get hopelessly lost" (because the probability of return of a random walk is less than 1), and also "is the highest numbered space that has any hope of being sensibly sorted out" (in terms of classification of its manifolds). At the same time, three-space is the lone holdout to the Poincaré conjecture (about when two manifolds are topologically equivalent). Asimov writes with enthusiasm from a topologist's point of view and hence finds no need to mention the glories of the Leech lattice in 24-space.

Pedoe, Dan, *Circles: A Mathematical View*, MAA, 1995; xxxvii + 102 pp, $18.95 (P). ISBN 0–88385–518–6.

This new edition of the 1957 classic on circles includes a major new addition, Chapter 0 (27 pp), dedicated to "the geometrical terms which may not be familiar to a new generation of students." An appendix (also in the original edition) reprints a biography of Karl Wilhelm Feuerbach, by Laura Guggenbuhl.

Stenseth, Nils Christian, Snowshoe hare populations: Squeezed from below and above, *Science* 269 (25 August 1995) 1061–1062; Krebs, C.J., *et al.*, Impact of food and predation on the snowshoe hare cycle, 1112–1115.

The ten-year cyclical variation in the populations of the Canadian snowshoe hare and its predator the Arctic lynx is well-known to students of differential equations as a "real-world" application of linear differential equations. To biologists, however, the cause of the cycling is more "obscure" than the simple mathematical model makes out; and the cycling has been attributed by biologists variously to food, predators, disease, sunspots, or a sequential interaction of food shortage followed by predation. The new light that these recent articles shed on the situation, via a controlled experiment, is that the effects on hare density of their supply of food and the density of their predators are superadditive. The results suggest that the cycling is produced by an interaction between food supply and predation.

Stewart, Ian, The never-ending chess game, *Scientific American* (October 1995) 182–183.

The rules of chess provide that the game is drawn if a player demonstrates that 50 moves have been made by each player without a capture or a pawn move. (In recent years, however, positions have been discovered from which a player can force a win but the mate takes more than 50 moves without a capture or pawn move.) Also, the rules provide that if the same position occurs three times with the same player to move, that player may claim a draw. Stewart explores a potential new termination rule: that the game should be drawn "if the same sequence of moves, in which the pieces are in exactly the same positions, repeats three times in a row." He goes on to exhibit pointless games that do not violate this rule (but would terminate under the 50-move rule), thus showing that the proposed rule is no protection "against players colluding to play stupidly," and suggesting to readers other mathematical questions about nonrepeating sequences.

Matthews, Robert, It's a lottery, *New Scientist* (22 July 1995) 38–42.

This article skips lightly through the "enthusiasm" and "ambivalence" people feel about randomness. It touches briefly on humans as randomizers, clustering of events, randomized algorithms, the law of large numbers, the Poisson distribution (V2 bomb hits on London), determining π (not by needle-tossing but from choosing pairs of stars at random), pseudo-random number generators, and the use of random numbers in cryptography.

Gardner, Martin, *New Mathematical Diversions*, rev. ed., MAA, 1995; 268 pp, \$19.95 (P). ISBN 0–88385–517–8.

This delightful volume, the third of Martin Gardner's collections of columns from *Scientific American*, originally appeared in 1966. Here it is reprinted with a Postscript updating the 20 columns, plus a new bibliography. Order a copy as a holiday gift for a curious junior-high or high-school student!

Matthews, Robert, The man who played God, *New Scientist* (2 September 1995) 36–40.

Older readers (and younger ones who have explored the library) will remember Donald Knuth's little romance novelette, *Surreal Numbers: How Two Ex-Students Turned on to Pure Mathematics and Found Total Happiness* (1974). Surreal numbers, invented by John Horton Conway (now of Princeton University), are an extension of the reals that contains infinitely large and infinitely small numbers. Now Martin Kruskal (Rutgers University) is "the leading authority on surreals and their chief evangelist"; he hopes to apply them to handle the divergent series produced in attempts at a unified theory of forces.

McIver, Annabelle, Why be formal, *New Scientist* (26 August 1995) 34–38.

"[T]he Ministry of Defence in Britain and the Canadian Energy Control Board [think nuclear power] have made formal methods mandatory for their software projects." This article gives simple examples of formal specifications, mentions invariants, and discusses formal validation (proving mathematically that the program implements the specification). The article mentions that Intel's testing did in fact find the infamous bug in its Pentium chip but "too late in the design to correct it without significant delay to the launch" (a less tragic echo of management's rush-to-launch that caused the Challenger space shuttle disaster). Why aren't formal methods always used in software projects—or at least, more often? Cost is a consideration, as is the "sheer boredom" of working through the great many small steps in long proofs; fortunately, automated proof-checkers are in the wings.

Marshall, Sandra P., *Schemas in Problem Solving*, Cambridge Univ. Pr., 1995; xi + 424 pp, $49.95. ISBN 0-521-43072-0.

At the college level, we encounter students who, if their lives depended on it, could not successfully set up a simple word problem; this, after they have spend years dealing with such problems (or perhaps successfully avoiding them?). Author Marshall explores the nature of problem-solving schemas in connection with the arithmetic "story" problems. She identifies five simple situations that alone or in combination account for all the relations in such problems: Change (one thing changes value over time), Group (small groups are combined into a large group), Compare (two things are contrasted to determine the smaller or the larger), Restate (a relationship is described between two different things), and Vary (a relationship is preserved; the problem is frequently couched in the wording "if... then"). Each situation can be approached to develop a problem-solving schema based on identification knowledge, elaboration knowledge, planning knowledge, and execution knowledge. She discusses theoretical and practical issues in using schemas to improve learning, gives examples of schema-based instruction (including computer-based instruction), and relates results of experimentation.

Dedekind, Richard, *What Are Numbers and What Should They Be?*, revised, translated, and edited by H. Pogorzelski, W. Ryan, and W. Snyder, Research Institute for Mathematics (383 College Ave., Orono, ME 04473), 1995; viii + 91 pp, $49.99 (post free). ISBN 0-964-3023-1-4.

This new translation of Dedekind's *Was sind und was sollen die Zahlen?* (1888) is accompanied by valuable footnotes by the translators, which help interpret the meaning of Dedekind's text. This is the work in which Dedekind tried to start from set theory and establish a foundation for the natural numbers (Dedekind cuts made their appearance in his 1872 *Stetigkeit und irrationale Zahlen*). The translators have modernized Dedekind's notation and terminology but omitted his prefaces and footnotes in favor of their own. The translators contend "that the true founder of the set-theoretic foundations of mathematics is Dedekind and not Cantor as many imagine" (the two met in 1872 and corresponded thereafter). Surprisingly, there is no mention of or comparison with the only previous translation into English, by W.W. Beman, which is included in the volume *Essays on the Theory of Numbers* (Open Court, 1901; Dover, 1963), still in print. The publisher plans new translations of Gauss's *Disquisitiones arithmeticae* and Dedekind's *Stetigkeit und irrationale Zahlen*, as well as translations of a previously untranslated paper of Cantor's on foundations of set theory and of lectures on number theory by Dirichlet and Dedekind.

NEWS AND LETTERS

ACKNOWLEDGEMENTS

In addition to our associate editors, the following have assisted the MAGAZINE *by refereeing papers during the past year. We appreciate the time and care they have given.*

Ayoub, Ayoub B., *The Pennsylvania State University, Abington, PA*
Barbeau, Edward J., *University of Toronto, Ontario, Canada*
Bivens, Irl C., *Davidson College, Davidson, NC*
Bloch, Norman J., *SUNY College at Brockport, Brockport, NY*
Boules, Raouf, *Towson State University, Towson, MD*
Casey, Stephen D., *The American University, Washington, DC*
Castellanos, Dario, *Valencia, Venezuela*
Chollet, John, *Towson State University, Towson, MD*
Fisher, J. Chris, *University of Regina, Saskatchewan, Canada*
Ghahramani, Saeed, *Towson State University, Towson, MD*
Godbole, Anant P., *Michigan Technological University, Houghton, MI*
Goldstein, Larry J., *Pheonixville, PA*
Green, Judy, *Marymount University, Arlington, VA*
Grimaldi, Ralph P., *Rose-Hulman Institute of Technology, Terre Haute, IN*
Guy, Richard, *University of Calgary, Alberta, Canada*
Johnson, Jerry A., *Oklahoma State University, Stillwater, OK*
Katz, Victor J., *University of District of Columbia, Washington, DC*
Krieger, Henry, *Harvey Mudd College, Claremont, CA*
Kung, Sidney J., *Jacksonville University, Jacksonville, FL*
Lay, David C., *University of Maryland, College Park, MD*
Maurer, Stephen B., *Swarthmore College, Swarthmore, PA*
May, Coy, *Towson State University, Towson, MD*
Mayfield, Elizabeth, *Hood College, Frederick, MD*
Michaels, John G., *State University of New York, Brockport, NY*
Morrison, John, *Towson State University, Towson, MD*
Nelsen, Roger B., *Lewis and Clark College, Portland, OR*
Penrice, Stephen, *Emory University, Atlanta, GA*
Richman, Dan, *Baltimore, MD*
Schattschneider, Doris M., *Moravian College, Bethlehem, PA*
Senechal, Marjorie, *Smith College, Northampton, MA*
Stern, Leonid, *Towson State University, Towson, MD*
Tattersall, James J., *Providence College, Providence, RI*
Tucker, Alan, *State University of New York at Stony Brook, Stony Brook, NY*
Wagon, Stan, *Macalester College, St. Paul, MN*
Zorn, Paul, *St. Olaf College, Northfield, MN*

INDEX TO VOLUME 68

AUTHORS

Dear Editor:

This is to point a slight error in "The Phi Number System Revisited," by Cecil Rousseau (October 1995). The Lemma on page 283 is incorrect. If one lets α equal ϕ, then the hypothesis of the Lemma is satisfied in that $\alpha \in \mathbb{Z}[\phi]$ and $1 \le \alpha < \sqrt{5}$. However, the conclusion is not satisfied as

$$|N(\alpha)| = |N(\phi)| = 1 = |N(1/\phi - 1)| = |N(\alpha - 1)|.$$

This does not affect the main theorem as, in the proof of the theorem, we know $1 \le \alpha < \phi$ and the Lemma can be proved for this more restrictive case.

Richard E. Stone
100 Birnamwood Drive
Burnsville, MN 55337

Dear Editor:

I certainly liked the argument in "A One-Sentence Proof That $\sqrt{2}$ Is Irrational," by David M. Bloom (October 1995). The author and other readers might be interested in what happened today, when I, for no good reason, was flipping through an old *American Mathematical Monthly* that was lying on my incredibly messy desk. To my amazement, I found basically the same argument in "A Simple Irrationality Proof for the Quadratic Surds," Vol. 75, pp. 772-773. There, the author gets slightly different mileage out of the last equation $\sqrt{k}$ be m/n where m and n are positive and one assumes that either (a) n is the smallest possible value; or (b) m is the smallest; or (c) $m+n$ is the smallest. Each one of these three is contradicted by the equation cited above, hence the author claims he has produced three different proofs of irrationality.

Rick Kreminski
East Texas State University
Commerce, TX 75429

Dear Editor:

There was a calculus manuscript I saw 20 years ago that had a beautiful bunch of algebraic methods for getting the integrals of elementary functions by methods such as those in "Limitless Integrals and a New Definition of the Logarithm," by David Shelupsky (October 1995). I had forgotten the author's name, but thought Vic Klee might remember it as he was the series advisor I was working with at the time. I wrote him, in part:

"The October *Mathematics Magazine* has a nice note on these sorts of formulas by the physicist David Shelupsky. The only problem with it is his notion that all of this is new. It may be original to him, or he may have amplified ideas he not recall hearing, but this is hardly new.

This is a case of no outlook on the past and light reviewing. I must have learned this general method from Burrows Hunt, one of my teachers at Reed. I see it is in his *Calculus and Linear Algebra*, W.H. Freeman and Co., 1967, p. 287, and it was certainly an old observation when I learned it from Hunt in the 1950s.

Those who do not know the past or fail to recall it will have the pleasure of rediscovering it and claiming it for their own as if it were new.

As for me, I am plagued with a memory that I can't put my finger on right now. Last night, after tossing and turning, I had the name I seek clear, after quite a few rounds of approximation. This morning it is gone. Can you refresh it?"

Klee supplied the answer: "Warren Stenberg."

Peter Renz
Academic Press
1300 Boylston St.
Chestnut Hill, MA 02167

UNITED STATES POSTAL SERVICE™

Statement of Ownership, Management, and Circulation
(Required by 39 U.S.C. 3685)

1. Publication Title: MATHEMATICS MAGAZINE

2. Publication No.: 0 0 2 5 - 5 7 0 X

3. Filing Date: October 16, 1995

4. Issue Frequency: bi-monthly except July and August

5. No. of Issues Published Annually: Five

6. Annual Subscription Price: Library $85; Indiv. Member $22

7. Complete Mailing Address of Known Office of Publication *(Street, City, County, State, and ZIP+4) (Not Printer)*
MAA, 1529 Eighteenth St., NW, Washington, DC 20036-1385

8. Complete Mailing Address of Headquarters or General Business Office of Publisher *(Not Printer)*
MAA, 1529 Eighteenth St., NW, Washington, DC 20036-1385

9. Full Names and Complete Mailing Addresses of Publisher, Editor, and Managing Editor *(Do Not Leave Blank)*

Publisher *(Name and Complete Mailing Address)*
The Mathematical Association of America, 1529 Eighteenth St., NW
Washington, DC 20036-1385

Editor *(Name and Complete Mailing Address)*
Martha J. Siegel, Towson State Univ., Dept. of Mathematics, Baltimore, MD 21204

Managing Editor *(Name and Complete Mailing Address)*
Harry Waldman, MAA, 1529 Eighteenth St., NW, Washington, DC 20036-1385

10. Owner *(If owned by a corporation, its name and address must be stated and also immediately thereafter the names and addresses of stockholders owning or holding 1 percent or more of the total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address as well as that of each individual must be given. If the publication is published by a nonprofit organization, its name and address must be stated.) (Do Not Leave Blank.)*

Full Name	Complete Mailing Address
The Mathematical Association of America	1529 Eighteenth St., NW
	Washington, DC 20036-1385

11. Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages, or Other Securities. If none, check here. ☒ None

Full Name	Complete Mailing Address

12. For completion by nonprofit organizations authorized to mail at special rates. The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes: *(Check one)*
☒ Has Not Changed During Preceding 12 Months
☐ Has Changed During Preceding 12 Months
(If changed, publisher must submit explanation of change with this statement)

PS Form **3526**, October 1994 *(See Instructions on Reverse)*

13. Publication Name: MATHEMATICS MAGAZINE

14. Issue Date for Circulation Data Below: June 1995

15. Extent and Nature of Circulation	Average No. Copies Each Issue During Preceding 12 Months	Actual No. Copies of Single Issue Published Nearest to Filing Date
a. Total No. Copies *(Net Press Run)*	18,871	16,605
b. Paid and/or Requested Circulation (1) Sales Through Dealers and Carriers, Street Vendors, and Counter Sales *(Not Mailed)*	0	0
(2) Paid or Requested Mail Subscriptions *(Include Advertisers' Proof Copies/Exchange Copies)*	15,012	12,960
c. Total Paid and/or Requested Circulation *(Sum of 15b(1) and 15b(2))*	15,012	12,960
d. Free Distribution by Mail *(Samples, Complimentary, and Other Free)*	1,488	1,018
e. Free Distribution Outside the Mail *(Carriers or Other Means)*	0	0
f. Total Free Distribution *(Sum of 15d and 15e)*	1,488	1,018
g. Total Distribution *(Sum of 15c and 15f)*	16,500	13,978
h. Copies Not Distributed (1) Office Use, Leftovers, Spoiled	2,371	2,627
(2) Return from News Agents	0	0
i. Total *(Sum of 15g, 15h(1), and 15h(2))*	18,871	16,605
Percent Paid and/or Requested Circulation *(15c / 15g x 100)*	90.98%	92.7%

16. This Statement of Ownership will be printed in the Dec. 1995 issue of this publication. ☐ Check box if not required to publish.

17. Signature and Title of Editor, Publisher, Business Manager, or Owner: Membership Manager — Kay Lamont — Date: 10-16-95

I certify that all information furnished on this form is true and complete. I understand that anyone who furnishes false or misleading information on this form or who omits material or information requested on the form may be subject to criminal sanctions *(including fines and imprisonment)* and/or civil sanctions *(including multiple damages and civil penalties).*

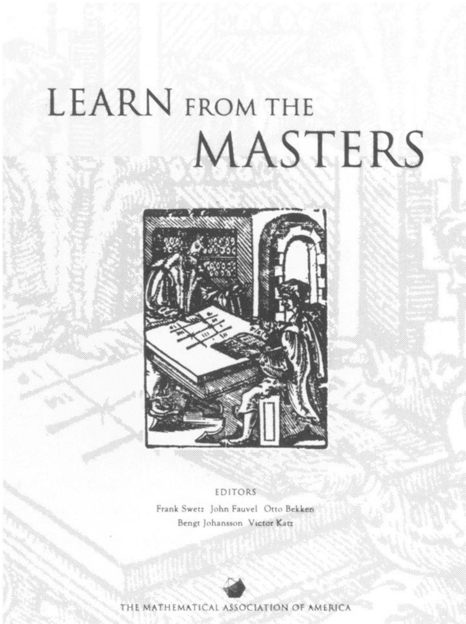
Instructions to Publishers

1. Complete and file one copy of this form with your postmaster on or before October 1, annually. Keep a copy of the completed form for your records.
2. Include in items 10 and 11, in cases where the stockholder or security holder is a trustee, the name of the person or corporation for whom the trustee is acting. Also include the names and addresses of individuals who are stockholders who own or hold 1 percent or more of the total amount of bonds, mortgages, or other securities of the publishing corporation. In item 11, if none, check box. Use blank sheets if more space is required.
3. Be sure to furnish all information called for in item 15, regarding circulation. Free circulation must be shown in items 15d, e, and f.
4. If the publication had second-class authorization as a general or requester publication, this Statement of Ownership, Management, and Circulation must be published; it must be printed in any issue in October or the first printed issue after October, if the publication is not published during October.
5. In item 16, indicate date of the issue in which this Statement of Ownership will be printed.
6. Item 17 must be signed.

Failure to file or publish a statement of ownership may lead to suspension of second-class authorization.

PS Form **3526**, October 1994 *(Reverse)*

Learn from the Masters



Frank Swetz, John Fauvel, Otto Bekken, Bengt Johansson, Victor Katz, Editors

Provides high school and college teachers with important historical ideas and insights which can be immediately applied in the classroom.

This book is for college and high school teachers who want to know how they can use the history of mathematics as a pedagogical tool to help their students construct their own knowledge of mathematics. Often, a historical development of a particular topic is the best way to present a mathematical topic, but teachers may not have the time to do the research needed to present the material. This book provides its readers with historical ideas and insights which can be immediately applied in the classroom.

The book is divided into two sections: the first on the use of history in high school mathematics, and the second on its use in university mathematics. So, high school teachers planning a discussion of logarithms, will find here the historical background of that idea along with suggestions for incorporating that history in the development of the idea in class. College teachers of abstract algebra will benefit by reading the three articles in the book dealing with aspects of that subject and considering their ideas for presenting groups, rings, and fields.

The articles are diverse, covering fields such as trigonometry, mathematical modeling, calculus, linear algebra, vector analysis, and celestial mechanics. Also included are articles of a somewhat philosophical nature, which give general ideas on why history should be used in teaching and how it can be used in various special kinds of courses. Each article contains a bibliography to guide the reader to further reading on the subject.

This book grew out of a conference in Norway which brought together mathematicians and mathematics educators from a dozen countries who were interested in the use of the history of mathematics as a pedagogical tool in the teaching of mathematics. Since the conference which provided the genesis of this book took place in Norway near the home where Niels Henrik Abel spent his final days, the book's title comes from a note scribbled in one of Abel's notebooks: "It appears to me that if one wants to make progress in mathematics one should study the masters." The authors hope that readers will benefit from Abel's advice and show their students how they too can Learn from the Masters.

312 pp., Paperbound, 1995
ISBN-0-88385-703-0
List: $26.50
MAA Member: $20.95
Catalog Code: LRM/JR

ORDER FROM:
THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, NW Washington, DC 20036
1-800-331-1622 (301) 617-7800 FAX (301) 206-9789

Membership Code: ____ ____ ____ ____ ____ ____

Name ______________________

Address ______________________

City ______________________

State ____________ Zip ____________

Qty.	Catalog Code	Price	Amount
______	LRM/JR	______	______
______	______	______	______
		TOTAL	______

Payment ☐ Check ☐ VISA ☐ MasterCard

Credit Card No. ______________ Expires ___/___

Signature ______________________

CONTENTS

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS

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