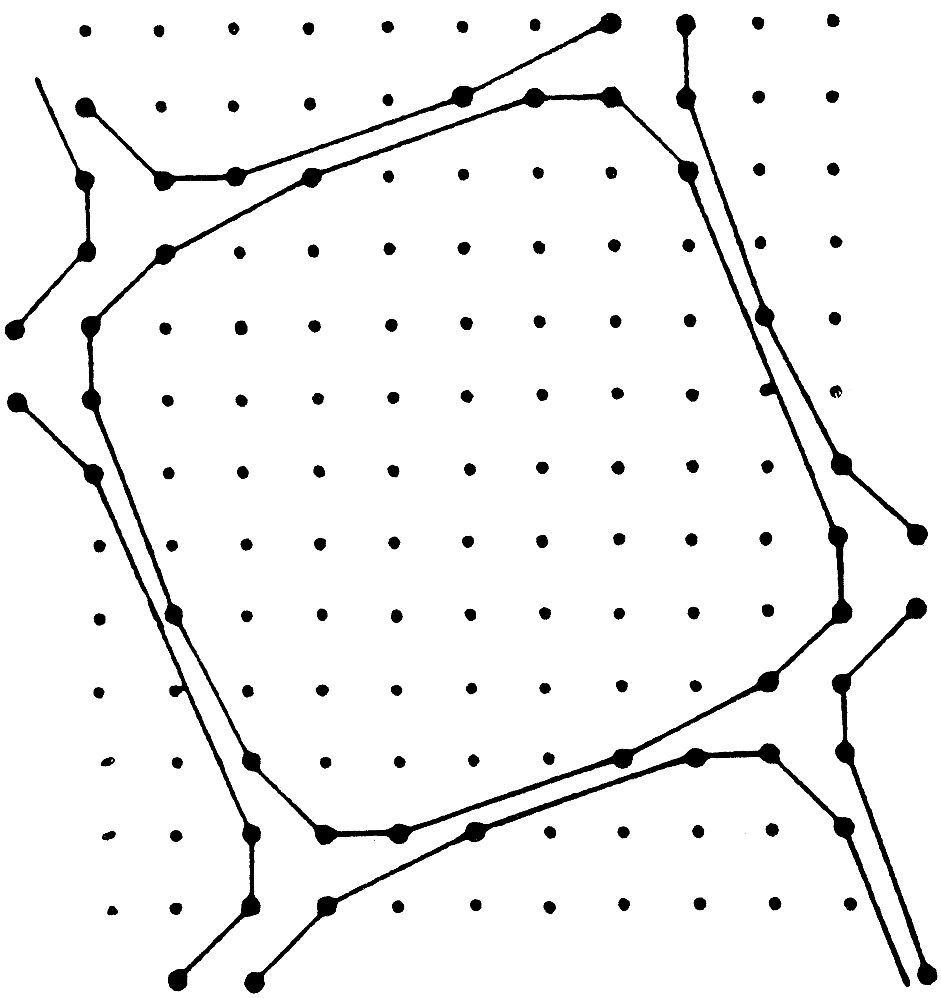
Vol. 64, No. 1 February 1991

MATHEMATICS MAGAZINE



- Translational Prototiles on a Lattice
- Assigning Driver's License Numbers
- Running Clubs—A Combinatorial Investigation

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.

A 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: Martha Siegel, Editor, Mathematics Magazine, Towson State University, Towson, MD 21204. 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.

AUTHORS

Ludwig W. Danzer received his Ph.D. in 1960 in Munich, spent a year at the University of Washington in Seattle, and earned his Habilitation in 1963 in Göttingen. A professor of mathematics at Dortmund since 1969, his main interest is discrete geometry, especially in packings, coverings, and tilings and their symmetries in Euclidean d-space. He was happy to find an application of the theory of continued fractions, once taught to him by Oskar Perron in 1949.

Grattan Murphy has earned degrees from Rockhurst College (B.S.) and St. Louis University (M.S. and Ph.D.). A faculty member of the University of Maine since 1965, he currently is professor and chair in mathematics. He has been active in the Northeast section of the MAA in a number of offices, including chair, and has worked with the section's Short course. A former elementary problems editor for the MONTHLY, his interests are in geometry and convexity.

John Reay studied music at Pacific Lutheran University and mathematics at the University of Washington (where he earned his Ph.D. in 1963). He now teaches at Western Washington University. His search for translational lattice prototiles with n vertices began as one of those fun problems to occupy a long airplane flight: $n \leqslant 6$ were found before a Seattle take-off; $n \leqslant 12$ by Denver; and $n \leqslant 20$ and the conjecture that became Theorem 2 existed before landing in New York.

Joseph A. Gallian received a B.A. degree from Slippery Rock and a Ph.D. from Notre Dame in 1971. His research areas are discrete mathematics and finite groups. His interest in the mathematics of identification numbers began one morning while looking at the UPC code on a box of corn flakes. Dr. Gallian directs undergraduate research, always looking out for applications for his abstract algebra class. Much of this material now appears in his book *Contemporary Abstract Algebra* (Heath, 1990).

Vol. 64, No. 1 February 1991

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, unemployed and emeritus members receive a 50% discount; 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

Translational Prototiles on a Lattice

LUDWIG DANZER
Universität Dortmund
D-4600 Germany

GRATTAN MURPHY
University of Maine
Orono, ME 04469

JOHN REAY
Western Washington University
Bellingham, WA 98225

If a set P is either a parallelogram, or else a centrally symmetric hexagon (so that opposite sides are parallel and have equal length), then the whole plane may be tiled with translates of P. Such a set P is called a *translational prototile* for the plane. FIGURE 1 shows two such tilings. Here we require that the union of the set of translates will cover all of the plane, and every two translates must have disjoint interiors [3].

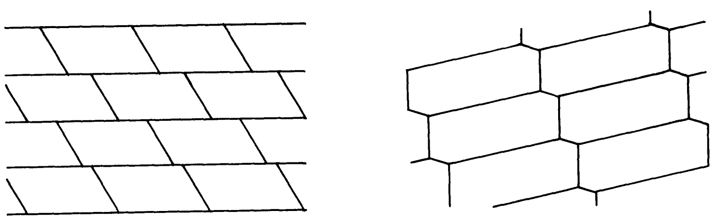


FIGURE 1

There are many other sets that are translational prototiles for the plane. For example, FIGURE 2 shows one that has many sides and that is not centrally symmetric. But suppose we restrict our attention to translational prototiles that are convex. Then it is a fairly easy exercise to show that the centrally symmetric hexagons and parallelograms, like those shown in FIGURE 1, are the only convex translational prototiles for the plane.

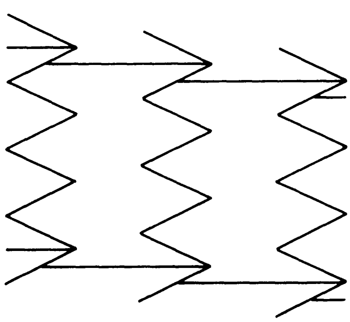


FIGURE 2

There is an equivalent way to state this fact. Suppose P is a convex polygon and we pack the plane as densely as we can with translates of P with disjoint interiors. If this densest packing actually covers the whole plane (we say that the packing has density 1), then P must be a centrally symmetric hexagon or parallelogram.

In this note we will consider the lattice of points in the plane having both coordinates integers, and use only lattice polygons whose vertices lie in this lattice. We will investigate the rich variety of convex lattice polygons whose disjoint translates can be used to cover the lattice. In contrast to the two classes of convex translational prototiles for tiling the whole plane, we will show there exist lattice-covering convex translational prototiles having any desired number of vertices. Although the characterization of all such translational prototiles is an open question, we will characterize all such triangles.

Notation, Examples, and Fundamental Results

A planar polygon P is called a *lattice polygon* if its vertices are in the *lattice* L of points where both coordinates are integers. We will call P a *translational lattice prototile* (TLP) if P is a convex lattice polygon, and there is a pair-wise disjoint family of translates of P that are also lattice polygons, whose union contains every lattice point of L.

Note that this definition is more restrictive than in the previously discussed case, where a tiling of the plane required only that the interiors of the tiles must be disjoint. Here the polygonal tiles that cover L must themselves be disjoint. For example, the hexagon shown in FIGURE 3 is clearly not a TLP, although there is a family of translates with pair-wise disjoint interiors (but not disjoint boundaries) that covers all lattice points.

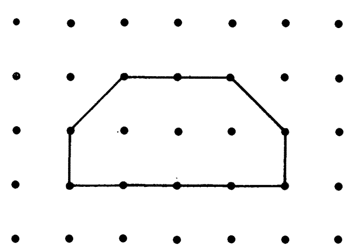


FIGURE 3

The TLP come in assorted sizes and shapes, not just hexagons and parallelograms. Examples like the one given in FIGURE 4 show that there exist triangular TLP with an arbitrarily large area. FIGURE 5 shows a triangular TLP with a lattice point as an interior point. FIGURE 6 shows a quadrilateral TLP whose opposite sides are not parallel. FIGURE 7 shows TLP with five and six vertices. (We conjecture that the hexagon shown in FIGURE 7 is the TLP with minimal perimeter, minimal number of interior lattice points, and minimal diameter among all TLP with six vertices.)

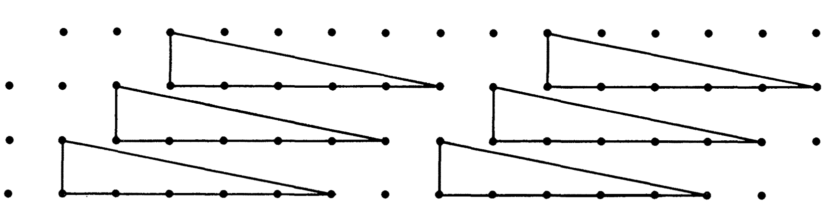


FIGURE 4
Case $n = 5$ of Theorem 1.

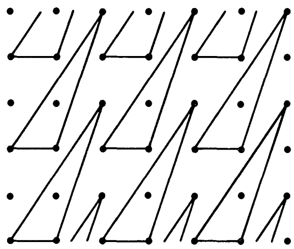


FIGURE 5

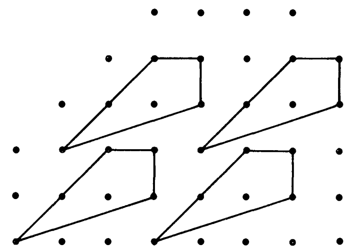


FIGURE 6

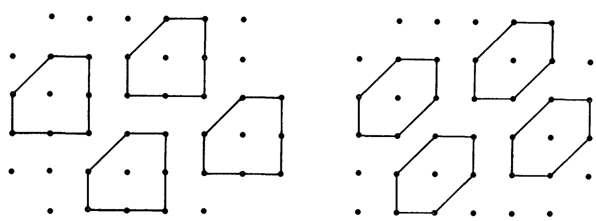


FIGURE 7
TLP with 5 and with 6 vertices.

There are several useful facts about lattice polygons and TLP that are either well known, or easy to show:

a) The area $A(P)$ of any lattice polygon P is given by Pick's Theorem: $A(P) = i + b/2 - 1$ where i is the number of lattice points interior to P and b is the number of lattice points on the boundary of P. It follows that the area of each TLP is an integer multiple of one half. Thus polygon P is a *primitive* lattice triangle (i.e., $L \cap P$ is just the set of three vertices of P) if and only if $A(P) = 1/2$. (See Coxeter [1] for a proof of Pick's Theorem, and Ding and Reay [2] for more recent references on Pick's Theorem.)

b) If $f(x) = Ax + b$ is an affine transformation of the plane that maps the lattice L onto itself, then the 2×2 matrix A has relatively prime integers in each row, and its determinant is $|A| = \pm 1$. Vector b is the lattice point $f(0)$. The image under f of any TLP is clearly another TLP with the same area, the same number of vertices, and the same number of interior lattice points. Thus the group $\mathscr{A}$ of lattice-preserving affine transformations induces an equivalence relation on the set of all TLP. (See Martin [4], Chapter 15, for a review of affine transformations.)

c) The line segment $[A, B]$ between lattice points A and B in the plane is called a *lattice segment* if A and B are the only lattice points on $[A, B]$. Coxeter ([1], p. 208) and others would say that the lattice point B is "visible" from the lattice point A. For each lattice segment $[A, B]$ there is a lattice-preserving affine transformation that maps segment $[A, B]$ onto the unit segment with end points $(0, 0)$ and $(0, 1)$.

Triangular Translational Lattice Prototiles

In this section we will characterize all triangular TLP. We begin with five easy observations:

d) It is clear that every lattice segment in the plane occurs as a side of some triangular TLP. Just choose any primitive lattice triangle with side $[A, B]$ and map its three vertices by an affine transformation $f \in \mathscr{A}$ onto the points $(0,0)$, $(1,0)$, and $(0,1)$ that form a TLP.

e) At least two edges of each triangular TLP must be lattice segments. To show this, assume that P is a triangular TLP with two edges that are not lattice segments. By b) we may assume, with no loss of generality, that one edge $[(0,0),(n,0)]$ of P is horizontal, as in FIGURE 8, and another edge $[(0,0),(jp,jq)]$ contains j lattice segments, with $n \geqslant j \geqslant 2$. Let P' and P'' be any translates of P that contain $A=(jp+1,jq)$ and $B=(n+p,q)$, respectively. Then A must be the translate of $(0,0)$ in P' (since P and P' are disjoint). Similarly B must be the translate of $(m,0)$ in P'', where $0 \leqslant m < n$. But this implies an intersection of P' and P'' at the point $C=(n-m+jp,jq)$, and thus P is not a TLP.

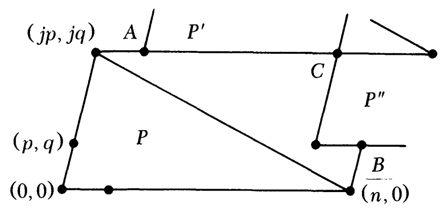


FIGURE 8

f) Since we are interested in characterizing a triangular TLP up to the equivalence classes formed by $\mathscr{A}$, we may assume without loss of generality that any triangular TLP has a horizontal edge from $(0,0)$ to $(n,0)$, and its other two edges are lattice segments meeting at a vertex (k,h) with $0 \leqslant k < h$. Thus it has exactly $b=n+2$ lattice points on its boundary, and by Pick's Theorem, exactly $i=n(h-1)/2$ lattice points in its interior.

g) If P is a triangular TLP with exactly one interior point, then all three sides of P must be lattice segments, and P is (affinely equivalent to) the triangle shown in FIGURE 5.

Proof. From f) it follows that $h=1+2/n$. Since h is an integer, n must be 1 or 2. If $n=2$, then $(k,h)=(1,2)$ as shown in FIGURE 9. If a disjoint translate P' of P contains $A=(2,1)$, then A is a translate of the origin. But then no disjoint translate of P contains $B=(2,2)$, so P cannot be a TLP. Hence, $n=1$, $h=3$, and k must be 2, i.e., P must be the prototile shown in FIGURE 5.

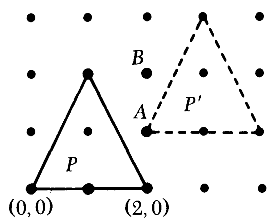


FIGURE 9

h) If $i>1$ then there is no triangular TLP with i interior points. In particular, the Diophantine equations $2i=n(h-1)$ and $0<k<h$ from f) have a small number of integer solutions when i is small, and reasoning similar to that in paragraph g)

excludes each case. (Recall that k and h must be relatively prime, as must $|k-n|$ and h, since two sides are lattice segments.) But the work required quickly increases when i gets large, and a different method of proof is given as part of the next theorem.

THEOREM 1. *The only triangular* TLP *(up to a lattice-preserving affine transformation) are the one particular example shown in* FIGURE 5, *with one interior point, and the countable family, parameterized by* $n \geqslant 1$, *with no interior points, as shown in* FIGURE 4.

Proof. The theorem follows from paragraph f) if $i=0$, and when $i=1$ it was proved in paragraph g). So let P be any lattice triangle with $i>1$ interior lattice points. Using the assumptions and notation of f) above, we show that P is not a TLP.

First consider the case when $n=1$, and P is a primitive lattice triangle with vertices at $(0,0)$ and $(1,0)$ and (k,h). Then P has $i=(h-1)/2$ interior points, so $h \geqslant 5$ must be odd. Here $k \geqslant 2$ since each edge is a lattice segment. We need a separate argument for the special subcase when $k=2$ as shown in FIGURE 10. Assume P is a TLP. Point $A=(1,i+1)$ is the lowest lattice point on the line $x=1$ that lies above triangle P. Clearly the only translate of P that contains A and is disjoint from P is $P'=\langle 0,i+1\rangle + P$, i.e., translate P vertically by moving $(1,0)$ to the point A. For $1 \leqslant m \leqslant i$ the point $(0,m)$ could only be covered by a translation of P that mapped the upper vertex $(2,h)$ onto $(0,m)$. Since $i \geqslant 2$, there are at least two such translates. But these translates of P are not disjoint. Hence, P cannot be a TLP.

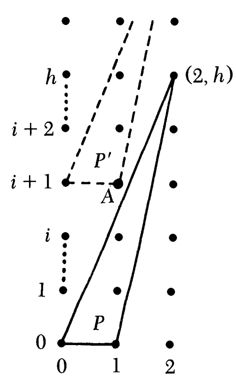


FIGURE 10

To complete the case when $n=1$, we may assume $3 \leqslant k < h$. (FIGURE 11 shows the cases when $(k,h)=(3,5)$ or $(5,7)$.) Let Q be the primitive lattice triangle with vertices $(0,0)$, (k,h) and $(k-1,h)$, and let S be the set of lattice points in the interior of Q. Since triangles Q and P are congruent, $|S|=i$. For each point C in S, let $P(C)$ denote the translated image of P that carries $(1,0)$ onto the point C. For all points C, D in set S, note that $P(C) \cap P(D) \neq 0$. For each point C in S, the only lattice translates of P that could possibly contain C and be disjoint from P are the translates $P(D)$ for some D in S. Hence, for P to be a TLP, it would be necessary for some $P(C)$ to contain all points of S. We will show that this does not happen. There are two points A and B of S that are closest to the line $y=hx/k$. Since $k \geqslant 3$, they will have different x-coordinates. Since h and k are relatively prime, the closer one,

A, must be at a distance $1/k$ directly above the line $y = hx/k$ and B will be at a distance $2/k$ directly above the line. If C is a point of S other than A, then clearly A is not in the triangle $P(C)$. Thus it is sufficient to show that B is not in $P(A)$. The least favorable case (e.g., $(k, h) = (3, 5)$) is when the x-coordinate of A is one less than the x-coordinate of B. In that case, the line through A with slope $h/(k-1)$ lies above B. Hence no triangle $P(C)$ contains all of S when $k \geqslant 3$. (Note that $S \subset P(A)$ if $k = 2$ as FIGURE 10 shows, and hence the special subcase above was necessary. Also note that $S \subset P(A)$ for cases such as $(k, h) = (9, 7)$ so the assumption $k < h$ was necessary.) This completes the proof if $n = 1$.

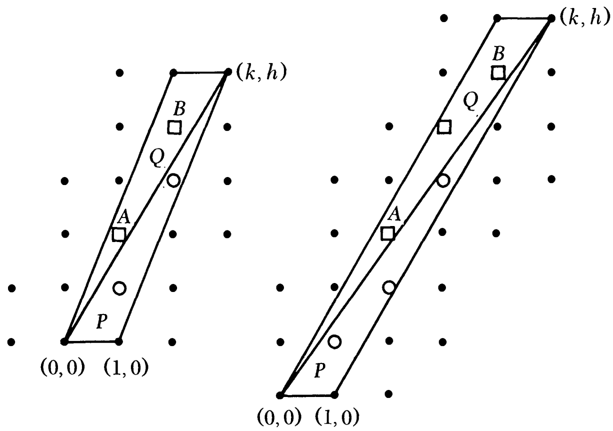


FIGURE 11

The cases $n \geqslant 2$ are similar and easier. Thus no triangular TLP has more than one interior point and the proof is complete.

COROLLARY 1. *Although the area of a triangular* TLP *may be arbitrarily large, its width is bounded by* $3(5)^{-1/2}$ *as shown by the dotted segment in* FIGURE 12. *Here* $i = 1$ *and the maximum edge length is minimized.*

(The *width* of P is the minimal distance between two parallel tangent lines on opposite sides of P. Width is not invariant under transformations $f \in \mathscr{A}$.)

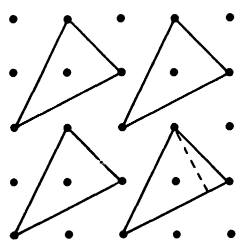


FIGURE 12

COROLLARY 2. *Once any copy of a triangular* TLP *is placed on the lattice L, there is a unique set of translations to disjoint lattice polygons that cover the lattice. That is, the translational cover of the lattice is completely determined by any one of its prototiles.*

Suppose we pack the plane, as densely as possible, with disjoint lattice polygons that are translates of a given lattice polygon P. If P is a TLP, then in a densest packing, every lattice point in the plane will be contained in a translate of P. In the general situation, not all lattice points will be covered by translated images of P. The *translation density* $T(P)$ of P may be defined, using limits in the usual way, as the fraction of the lattice points that lie in some translate. It is clear that $T(P) = T(f(P))$ for any $f \in \mathscr{A}$. Polygons such as the one in FIGURE 13 show that $T(P)$ may be made less than any positive number. We conjecture that $T(P) \geqslant 2/3$ if P is restricted to be a convex lattice polygon. Clearly, $T(P)$ is very near $2/3$ for triangles with large area that are, for example, nearly equilateral. The above theorem characterized those triangular lattice polygons P for which $T(P) = 1$. It would be interesting to have an algorithm for determining $T(P)$ for any lattice polygon, or to characterize the lattice triangles P with other values of $T(P)$.

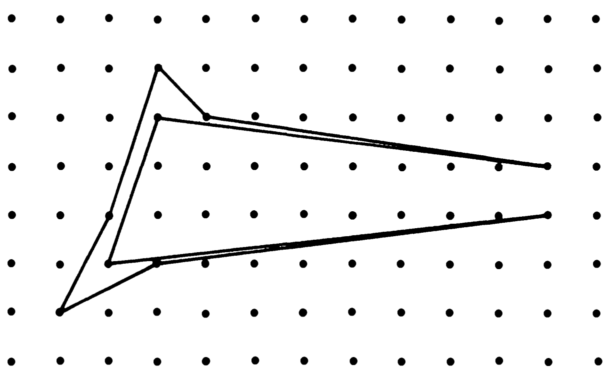


FIGURE 13

Convex Translational Lattice Prototiles with Many Vertices

Theorem 1 in the previous section characterized all triangular TLP. It is an open problem to find a similar characterization of TLP with four or more vertices. The case of convex lattice quadrilaterials (or, more specifically, parallelograms or squares) is particularly tempting. We are grateful to Doris Schattschneider for several interesting partial results and conjectures for this case of TLP with four vertices, which she suggested in her referee's report of this paper. (These will appear in a separate paper.) One example, which gives the flavor of these results, is the fact that there exist lattice squares that are not TLP. It is easy to show that the square in FIGURE 14 is the smallest one.

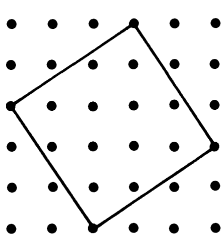


FIGURE 14

Another open problem that could be considered is to determine $i(v)$, the greatest lower bound on the number of lattice points that must be in the interior of any TLP with exactly v vertices. Clearly, $i(3) = i(4) = 0$, and the examples in FIGURE 7 show that $i(5) \leqslant 1$ and $i(6) \leqslant 1$. We conjecture that $i(v)$ increases without bound as v gets large. This conjecture makes sense, of course, only if we know there exist TLP with a large number of vertices. In the remainder of this section, therefore, we will consider TLP with many vertices. We will use well-known results about continued fractions (as found, for example, in chapter 7 of the number theory book by Niven and Zuckerman [5]) to show that there exist TLP with any number of vertices. But first let's look at another example.

As an example, several TLP with 15 vertices are shown in FIGURE 15. We conjecture that among all TLP with 15 vertices, the one shown in FIGURE 15(a) (with diameter $d = 10.77\ldots$ and perimeter $p = 31.237\ldots$) has the least diameter, and the one shown in FIGURE 15(b) (with $d = 11.18\ldots$ and $p = 29.38$) has the least perimeter.

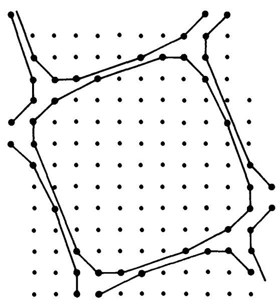


FIGURE 15(a)

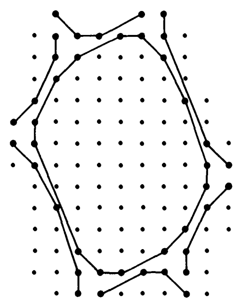


FIGURE 15(b)

We next give a constructive algorithm that produces a convex lattice polygon with any specified number $b \geqslant 3$ of vertices. In general it will have many lattice points in its interior, but it will always have width less than one. We will show that this polygon contains all lattice points that are in the interior and on the left boundary of a certain lattice parallelogram, and hence it will provide a constructive proof of the existence of TLP with arbitrarily many vertices.

Let $a_1, a_2, \ldots, a_k$ be any sequence of positive integers with $a_1 \geqslant 2$ and $k = b - 2 \geqslant 1$. Use the a_i to define a continued fraction m/n with g.c.d. $(m,n) = 1$:

$$\frac{m}{n} = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \ddots + \cfrac{1}{a_k + 1}}}}.$$

Define p_i, q_i recursively, as

$$\begin{aligned} p_0 &= 0, & p_1 &= 1, & p_i &= a_i p_{i-1} + p_{i-2} \\ q_0 &= 1, & q_1 &= a_1, & q_i &= a_i q_{i-1} + q_{i-2}. & i &\leqslant k \quad (*) \end{aligned}$$

Then the a_i are the numbers defining the continued fraction decomposition of m/n and p_i/q_i are the associated convergents. Since all of the a_i are positive, $p_1/q_1 > m/n$. Then it follows that $p_i/q_i > m/n > p_{i+1}/q_{i+1}$ if i is odd $(1 \leqslant i \leqslant k-1)$ [5, chapter 7].

The line segment S from $(0,0)$ to (m,n) is a lattice segment and has slope $n/m \geqslant 2$. The segments from $(0,0)$ to (p_i, q_i) are above S if i is even and below S if i is odd. The vertices of a polygon P may now be obtained in the following way. Take $(0,0)$, $(1,1) = (p_0 + 1, q_0)$, $(p_2 + 1, q_2)$, $(p_4 + 1, q_4), \ldots,$ and $(p_r + 1, q_r)$, where r is the index of the last even indexed pair. (Thus $r = k-1$ or $r = k$.) Also take (m,n), $(m - p_1 + 1, n - q_1), (m - p_3 + 1, n - q_3), \ldots,$ and $(m - p_t + 1, n - q_t) = (p_r + 1, q_r)$, where t is the last odd index. (Thus $t = k-1$ or $t = k$, and $t \neq r$.) Note that $m = p_k + p_{k-1} = p_r + p_t$ and $n = q_k + q_{k-1} = q_r + q_t$ follow from $(*)$ by taking the last denominator in the continued fraction to be $a_k + 1$. It also follows from the recursion formulas $(*)$ that each of these $b = k + 2$ points is not in the convex hull of the remaining $k+1$ points, and hence they do form the set of $k+2$ vertices of a convex lattice polygon P. See FIGURE 16. Since all the vertices of P except for $(0,0)$ and (m,n) are (by 5, Theorem 7.13) contained in the interior of the parallelogram R with vertices $(0,0), (1,0), (m,n), (m+1,n)$, it follows that the width of P is less than 1.

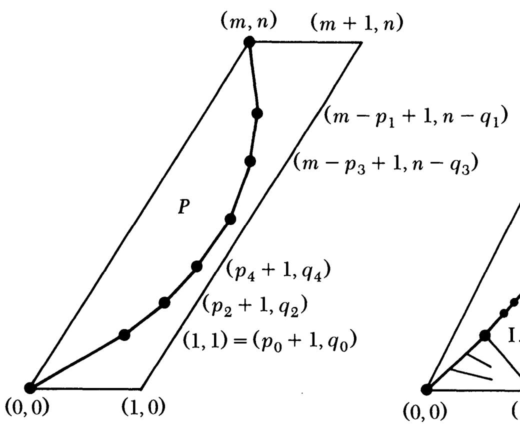


FIGURE 16

Thin Polygon *P*.

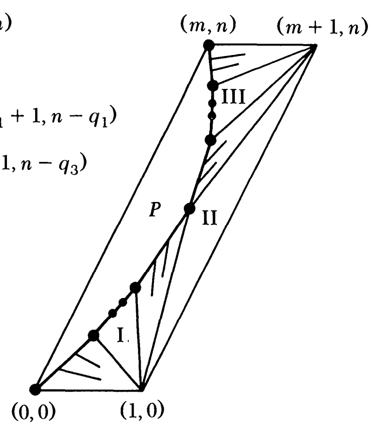


FIGURE 17

Triangulation of *R* \ *P*.

THEOREM 2. *For each positive integer $b \geqslant 3$ there exists a translational lattice prototile with b vertices and width less than one.*

Proof. It is sufficient to prove that any convex lattice polygon P with b vertices, which is formed by the algorithm given above, is a TLP. That is, it shows that disjoint translates of P to other lattice polygons may be used to cover the whole lattice L. To show this, it is sufficient to prove that P contains all the interior points of the parallelogram R [with vertices $(0,0)$, $(1,0)$, (m,n), and $(m+1,n)$], since then the translations that are integer combinations of the vectors $\langle 1,0 \rangle$ and $\langle 0, n+1 \rangle$ will work. We establish this by triangulating the region in R to the right of the polygon P, and showing that each triangle contains no lattice points except for $(1,0)$, $(m+1,n)$ or vertices and lattice points on the polygon P itself. There are three types of triangles to consider, with illustrations of each labeled as I, II, III in FIGURE 17. The areas of these triangles are

$$\frac{1}{2}\begin{vmatrix} 1 & 0 & 1 \\ p_{2j}+1 & q_{2j} & 1 \\ p_{2j+2}+1 & q_{2j+2} & 1 \end{vmatrix} = \frac{1}{2}\begin{vmatrix} 0 & 0 & 1 \\ p_{2j} & q_{2j} & 1 \\ p_{2j+2} & q_{2j+2} & 1 \end{vmatrix} = |p_{2j}q_{2j+2} - p_{2j+2}q_{2j}|/2$$

for case I, and

$$\frac{1}{2}\begin{vmatrix} 1 & 0 & 1 \\ m+1 & n & 1 \\ p_r+1 & q_r & 1 \end{vmatrix} = \frac{1}{2}\begin{vmatrix} 0 & 0 & 1 \\ m & n & 1 \\ p_r & q_r & 1 \end{vmatrix} = |mq_r - np_r|/2 = |p_k q_{k-1} - p_{k-1}q_k|/2 = \frac{1}{2}$$

for case II, and a similar result for case III. The areas in case I and case III are both of the form $|p_i q_{i-2} - p_{i-2} q_i|$, and since $p_i q_{i-2} - p_{i-2} q_i = (-1)^i a_i$ [5, Theorem 7.5],

$$\text{Area (I)} = (a_{2j+2})/2, \qquad \text{Area (III)} = (a_{2j+3})/2.$$

The corollary of Pick's theorem in paragraph a) shows that there are no lattice points on the triangle in case II except for the vertices. In cases I and III the segment on the polygon P may have lattice points other than the vertices. In case I, for example, the points $(p_{2j}+1, q_{2j}), (p_{2j+1}+p_{2j}+1, q_{2j+1}+q_{2j}), \dots, (a_{2j+2}p_{2j+1} + p_{2j}+1, a_{2j+2}q_{2j+1}+q_{2j}) = (p_{2j+2}+1, q_{2j+2})$ are all on that segment. Since these $a_{2j+2}+1$ points and $(1,0)$ are all boundary points, Pick's theorem again tells us there can be no interior points in the triangle of case I. Case III is entirely analogous. This establishes the theorem.

The same construction also yields a centrally symmetric convex TLP with any number $2b-2$ of vertices. To see this, just consider $Q = P \cup (\langle m,n\rangle + (-P))$ and the lattice of translations generated by $\langle 2,0\rangle$ and $\langle m+1, n\rangle$.

Note that if the TLP has $b \geqslant 4$ vertices, then the whole lattice covering is not uniquely determined, in general, by the position of the first translate, as it was in Corollary 2. An interesting example of this may be shown with the covering of L by hexagons, as shown in FIGURE 7. Choose one of these hexagons to be the prototile, and let its interior lattice point be the origin. Then reflection in the line $y = x$ maps the lattice L onto itself and maps the prototile onto itself, but gives a different covering by translates.

It is an open problem to consider what prototiles would be possible if, besides translation, one is allowed to turn them through 180°. The triangle in FIGURE 18 shows that this class is larger than the set of TLP.

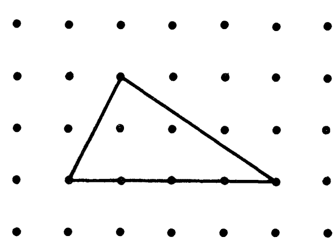


FIGURE 18

REFERENCES

1. H. S. M. Coxeter, *Introduction to Geometry*, John Wiley & Sons, New York, 1963. (Chapter 13.)
2. R. Ding and J. Reay, The boundary characteristic and Pick's theorem in the Archimedean planar tilings, *J. Combin. Theory Ser. A* 44 (1987), 110–119.
3. B. Grünbaum and G. Shephard, *Tilings and Patterns*, W. H. Freeman & Co., New York, 1987.
4. G. Martin, *Transformation Geometry*, Springer Verlag, New York, 1982.
5. I. Niven and H. Zuckerman, *An Introduction to the Theory of Numbers*, John Wiley & Sons, New York, 1960.

Assigning Driver's License Numbers

JOSEPH A. GALLIAN
University of Minnesota
Duluth, MN 55812

"You know my name
look up the number"

John Lennon and Paul McCartney, *You know my name*, single, B-side of *Let it be*, March 1970.

Introduction

Among the individual states, a wide variety of methods are used to assign driver's license numbers. The three most common methods, a sequential number, the social security number, and a computer-generated number, are uninteresting mathematically. On the other hand, many states encode data such as month and date of birth, year of birth, and sex in ways that involve elementary mathematics. Seven states go so far as to employ a check digit for possible detection of forgery or errors. Several states assign driver's license numbers by applying complicated hashing functions to the first, middle, and last names and formulas or tables for the month and date of birth. Surprisingly, the assignment of numbers is not always injective. In Michigan, for instance, there are 56 numbers whose inverse image has two or more members. New Jersey incorporates eye color into the number. Some states keep their method confidential. In a few instances, administrators of the license bureaus do not know the method used to assign numbers in their state! In this paper we discuss some of the methods we have uncovered.

Check Digit Schemes in General Use

Schemes for the assignment of identification numbers are extremely varied in methodology and in the information encoded. Most interesting to mathematicians are those that incorporate an extra digit for the detection of errors or fraud. Although the purpose of this paper is to analyze the methods used for driver's license numbers, it is worthwhile to begin with a brief survey of the methods employed to assign check digits to the most ubiquitous numbers in use and to provide a theoretical result that delineates their limitations.

The simplest and least effective method for assigning a check digit is to use the remainder or inverse of the remainder of the identification number modulo some number. For airline tickets, UPS packages, and Federal Express mail the check digit is the identification number modulo 7. At the bottom of FIGURE 1 we see the airline ticket number 17000459570 (the airline code 012 is not used in the calculation) is assigned a check digit 3 since $17000459570 \equiv 3 \bmod 7$.

U.S. postal money orders use the remainder modulo 9 while VISA traveler's checks use the inverse of the number modulo 9. Thus, the check digit for the VISA number 1002044679091 is 2 since $1002044679091 \equiv 7 \bmod 9$.

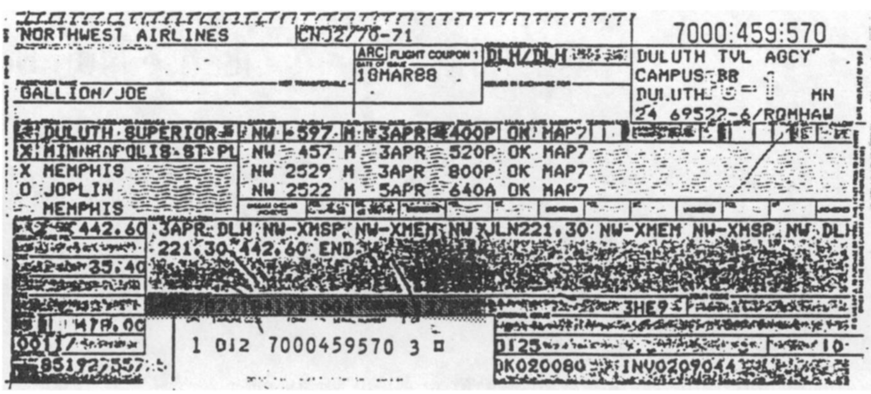


FIGURE 1
Airline ticket with number 17000459570 and check digit 3.

The modulo 7 schemes detect all errors involving a single digit except those where b is substituted for a and $|a-b|=7$. Likewise, an error of the sort $\cdots a_i \cdots a_j \cdots \to \cdots a_j \cdots a_i \cdots$ will go undetected if $|a_i - a_j| = 7$ or if 6 divides $j-i$.

The modulo 9 schemes are slightly better at detecting single-digit errors: Only a substitution of a 9 for a 0 or vice versa goes undetected. On the other hand, the only errors of the form $\cdots a_i \cdots a_j \cdots \to \cdots a_j \cdots a_i \cdots$ that are undetected are those involving the check digit itself. (A quick proof of this is to observe that the residue of a number modulo 9 is the residue of the sum of its digits modulo 9.)

Nearly all methods for assigning a check digit to a string of digits involve a scalar product of two vectors and modular arithmetic. For a string $a_1a_2 \cdots a_{k-1}$ and a modulus n, many schemes assign a check digit a_k so that

$$(a_1, a_2, \ldots, a_k) \cdot (w_1, w_2, \ldots, w_k) \equiv 0 \bmod n.$$

We call such schemes *linear* and we call the vector $(w_1, w_2, \ldots, w_k)$ the *weighting vector*. The Universal Product Code (UPC) used on grocery items employs the weighting vector $(3,1,3,1,3,1,3,1,3,1,3,1)$ with $n=10$; the International Standard Book Number (ISBN) utilizes $(10,9,8,7,6,5,4,3,2,1)$ and $n=11$; banks in the U.S. use $(7,3,9,7,3,9,7,3,9)$ with $n=10$; many Western countries use $(7,3,1,7,3,1,\ldots)$ with $n=10$ to assign check digits to numbers on passports. Notice that the division schemes mentioned at the outset of this section are also linear with weighting vectors of the form $(10^{k-2}, 10^{k-3}, \ldots, 10^0, \pm 1)$.

The error-detecting capability of linear schemes is given by the following theorem.

THEOREM. *Suppose a number $a_1a_2 \cdots a_k$ satisfies the condition $(a_1, a_2, \ldots, a_k) \cdot (w_1, w_2, \ldots, w_k) \equiv 0 \bmod n$. Then the single error occasioned by substituting a_i' for a_i is undetectable if and only if $(a_i' - a_i)w_i$ is divisible by n and a sole error of the form $\cdots a_i \cdots a_j \cdots \to \cdots a_j \cdots a_i \cdots$ is undetectable if and only if $(a_i - a_j)(w_i - w_j)$ is divisible by n.*

Proof. If a_i' is substituted for a_i, then the dot product of the correct number and the incorrect number differ by $(a_i' - a_i)w_i$. Thus, the error is undetected if and only if $(a_i' - a_i)w_i \equiv 0 \bmod n$.

Now consider an error of the form $\cdots a_i \cdots a_j \cdots \to \cdots a_j \cdots a_i \cdots$. Here the dot product of the correct number and the incorrect number differ by

$$(a_i w_i + a_j w_j) - (a_j w_i + a_i w_j) = (a_i - a_j)(w_i - w_j)$$

The conclusion now follows as before.

Since the most common moduli are 10 and 11, the following corollary is worth mention.

COROLLARY. *Suppose an identification number* $a_1 a_2 \cdots a_k$ *satisfies*

$$(a_1, a_2, \ldots, a_k) \cdot (w_1, w_2, \ldots, w_k) \equiv 0 \bmod n$$

where $0 \leqslant a_i < n$ *for each* i. *Then all single-digit errors occurring in the ith position are detectable if and only if* w_i *is relatively prime to* n *and all errors of the form* $\cdots a_i \cdots a_j \cdots \to \cdots a_j \cdots a_i \cdots$ *are detectable if and only if* $w_i - w_j$ *is relatively prime to* n.

The above theorem verifies our claims about the error-detection capability of the schemes used on money orders and airline tickets. It also explains why the bank and passport schemes will detect some errors of the form $\cdots abc \cdots \to \cdots cba \cdots$ while the UPC code will detect no such errors. Observe that because 11 is prime the ISBN code detects 100% of all single-digit errors and 100% of all errors involving the interchange of two digits. But there is a price to pay for using the modulus 11: The number a_k needed to satisfy the condition may be 10, which is two digits. In this case, an alphabetic character such as X or A is used or such numbers are not issued. As we will see below there are schemes that use the modulus 11 that do not resort to an alphabetic character, but there is a price to pay for this too: Not all transposition errors are detectable. More information about check digits schemes in use can be found in **[1]**, **[2]**, **[3]**, **[4]**, **[6]**, **[7]**, **[8]**.

Check Digits on Driver's License Numbers

The state of Utah assigns an eight-digit driver's license number in sequential order, say $a_1 a_2 \cdots a_8$, then appends a check digit a_9 using a linear scheme with weighting vector $(9, 8, 7, 6, 5, 4, 3, 2, 1)$ and modulus 10. This method is identical to that used by the American Chemical Society for its chemical registry numbers. Assuming that all errors are equally likely,[1] this method detects 73/81 or 90.1% of all single-digit errors and 100% of all transposition errors (i.e., errors of the form $\cdots ab \cdots \to \cdots ba \cdots$).[2]

To verify the single-error detection rate, observe from our theorem that in positions 2, 4, 6, and 8 substitution of b for a will go undetected when $|a - b| = 5$; in position 5, a substitution of b for a will go undetected when a and b have the same parity. Thus in each of positions 2, 4, 6, and 8 there are 10 undetected errors among 90 possible errors while in the fifth position, 40 of the 90 possible errors are undetected.

[1] In practice all errors are not equally likely. One study [7, p. 15] revealed that a substitution of a "5" for a "3" was 17 times as likely as a substitution of a "9" for a "1." However, available data are insufficient to assign reliable probabilities to the various error possibilities.

[2] A highly publicized error of this kind recently occurred when Lt. Col. Oliver North gave U.S. Assistant Secretary of State Elliot Abrams an incorrect Swiss bank account number for depositing $10 million for the contras. The correct number begins with "386"; the number North gave to Abrams begins with "368."

So, in all, 80 of 810 errors are undetected.

Someone working for the Canadian province of Quebec, probably having seen a scheme like the one used by Utah somewhere, came up with the laughable weighting vector $(12, 11, 10, 9, \ldots, 2, 1)$ with modulus 10 to assign a check digit. Of course, any error in the third position is undetectable and weights of 12 and 11 have the same effect as the weights 2 and 1.

Newfoundland uses the weighting vector $(1, 2, 3, 4, 5, 6, 7, 8, 1)$ with modulus 10. This is nearly identical to the Utah scheme except that it will not detect the event that the first and last digit are interchanged.

Three states use a modified linear scheme with modulus 11. New Mexico and Tennessee append a check digit a_8 to $a_1a_2 \cdots a_7$ as follows: First calculate

$$x = -(a_1, a_2, \ldots, a_7) \cdot (2, 7, 6, 5, 4, 3, 2) \bmod 11.$$

If $x = 0$, a_8 is 1; if $x = 10$, $a_8 = 0$; otherwise $a_8 = x$. This method catches 100% of all single-digit errors. Furthermore, the only errors of the form $\cdots a_i \cdots a_j \cdots \to \cdots a_j \cdots a_i \cdots$ that go undetected are those where $i = 1$ and $j = 7$ (an unlikely error indeed) and some involving the check digits 0 and 1. Assuming that all transposition errors are equally likely,[3] this method detects 98.2% of such errors. The Vermont scheme is the same as the one used by New Mexico except that when $x = 0$, the letter "A" is the check. This method, like the ISBN method, yields a 100% detection rate for both single-digit and transposition errors, but utilizes two formats for numbers. Notice that there would be nothing lost if the weighting vector began with 8 instead of 2 and there would be a slight gain since errors of the form $a_1a_2 \cdots a_7a_8 \to a_7a_2 \cdots a_1a_8$ would be detectable.

The state of Washington and the province of Manitoba use a check digit scheme devised by IBM in 1964 to assign a check digit. The license number is a blend of 12 alphabetic and numeric characters. To compute the Washington check digit, alphabetic characters are assigned numeric values as follows: $* \to 0, A \to 1, B \to 2, \ldots, I \to 9, J \to 1, K \to 2, \ldots, R \to 9, S \to 2, T \to 3, \ldots, Z \to 9$. (Notice the aberration at S.) The 12-character license number, after an alphabetic to numeric conversion, then corresponds to a string of digits $a_1a_2 \cdots a_{12}$ with a_{10} as the check digit calculated as $|a_1 - a_2 + a_3 - a_4 + \cdots + a_9 - a_{11} + a_{12}| \pmod{10}$. Interestingly, the use of the absolute value actually makes the method nonlinear and reduces the error detection capability of the scheme. It would have been better to use the linear scheme with weighting vector $(1, 9, 1, 9, \ldots, 1) \bmod 10$.

South Dakota and Saskatchewan employ another nonlinear scheme developed by IBM to assign its check digit. In South Dakota, a six-digit computer-generated string is assigned a check digit as follows. Each of the second, fourth, and sixth digits is multiplied by 2 and the digits of the resulting products are summed (e.g., a 7 yields $1 + 4 = 5$ while a 3 yields 6). This resulting total is then added to the digits in the first, third, and fifth positions. The check digit is the inverse modulo 10 of this tally. (Alternatively, the check digit is $(10 - ((\Sigma_{i\text{ even}}\ (2a_i + \lfloor 2a_i/10 \rfloor) + \Sigma_{i\text{ odd}}\ a_i) \pmod{10})) \bmod 10$.) Thus, the check digit for 263743 is $-(1 + 2 + 1 + 4 + 6 + 2 + 3 + 4) \bmod 10 = 7$. This method is used by credit card companies, many libraries, and drug stores in the U.S. and by banks in West Germany, although in some instances it is the digits in the odd positions that are multiplied by 2. It detects 100% of all

---

[3] In reality, the likelihood of a transposition error depends on the pair of digits as well as the positions. But as before, reliable data for these occurrences are unavailable.

single-digit errors and 97.8% of transposition errors. To see that all single-digit errors are detected, observe that distinct digits contribute distinct values to the sum. To compute the detection rate for errors of the form $\cdots ab \cdots \to \cdots ba \cdots$, suppose such an error is undetected. We consider four cases. For simplicity, assume a in the correct number is in position 2, 4 or 6. The alternative case gives the same result.

Case 1. $a, b < 5$
Then $2a + b \equiv 2b + a \bmod 10$.
Thus $a - b = 0$ and $a = b$.

Case 2. $a < 5$, $b \geqslant 5$
Then $2a + b \equiv 2b - 9 + a \bmod 10$.
It follows that $b - a = 9$ so that $b = 9$ and $a = 0$.

Case 3. $a \geqslant 5$, $b < 5$
Then $2a - 9 + b \equiv 2b + a \bmod 10$.
So, $a - b = 9$ and $a = 9$ and $b = 0$.

Case 4. $a \geqslant 5$, $b \geqslant 5$
Then $2a - 9 + b \equiv 2b - 9 + a \bmod 10$.
Thus $a - b = 0$ and $a = b$.

So all transposition errors except $09 \leftrightarrow 90$ are detected. Since there are 90 possible transposition errors, the error detection rate is 88/90 or 97.8%.

It is worth noting that Gumm [4] has shown that it is not possible to improve upon these rates with any system that uses addition modulo 10 to compute the check digit without utilizing an extra character, as was the case for the New Mexico scheme.

Wisconsin appends a check digit to a 13-digit string. Unfortunately, I have not been able to figure out how this scheme works. I know it isn't linear; for if so, the weighting vector $(w_1, w_2, \ldots, w_{13}, w_{14})$ could be determined by gathering up a large number of valid license numbers to produce a system of linear equations with the w_i's as the unknowns. I have done this for modulo 10 and 11 to no avail. To circumvent any peculiarity that might arise involving a check digit of 10 in a modulo 11 scheme (e.g., New Mexico), I avoided numbers with a check digit of 0 or 1.

Encoding Personal Data

Here is the driver's license number of a Wisconsin resident: E 425-7276-9176-07. What information about the holder can you deduce from this number: year of birth, day and month of birth, sex, name? None of these is obvious. Let's go the other direction. I am a resident of Minnesota. I was born on January 5, 1942, and my middle name is Anthony. From this can you deduce my driver's license number?

Eleven states assign their driver's license numbers with hashing functions applied solely to personal data. A good hash function should be fast and minimize collisions (see [5, pp. 506–544] for a detailed discussion of this topic). Of course, there will be occasions when two or more individuals have enough personal data in common that collisions will occur. Most states have a tie-breaking mechanism to handle this situation. Coding license numbers only from personal data enables automobile insurers, government entities, and law enforcement agencies to determine the numbers when necessary.

Washington uses a complicated blend of name, check digit, and codes for the month and date of birth to assign its numbers. This 12-digit identifier consists of the first five letters of the surname; the first and middle initials (* is used when a name has less than five characters, or there is no middle initial); the year of birth subtracted from 100 (we suspect this is done to disguise the year of birth); a check digit; a code for the month of birth; and a code for the day of birth. For instance, Fielding Mellish (no middle name) born on 10/29/42 receives the identifier MELLI F* 587P9. When checked against a file of 1.6 million items, this scheme yielded duplicates at the rate of 0.03% and only one number appeared as many as four times. (Most of the duplications represented twins.) To ensure that the correspondence between individuals and numbers is injective, 17 alternate codes for month and year of birth are available. For example, an S can be used instead of a B for January or a Z instead of a 9 for the year of birth. Interestingly, the check digit is invariant under all alternate coding. The primary code and one alternate for months is given in TABLE 1 and the code for the days is given in TABLE 2. Notice the absence of completely predictable patterns.

TABLE 1. Washington code for months.

Months	Codes	Alternate Codes
January	B	S
February	C	T
March	D	U
April	J	1
May	K	2
June	L	3
July	M	4
August	N	5
September	O	6
October	P	7
November	Q	8
December	R	9

TABLE 2. Washington code for days.

1 - A	7 - G	13 - L	19 - R	25 - 5	31 - U
2 - B	8 - H	14 - M	20 - 0	26 - 6	
3 - C	9 - Z	15 - N	21 - 1	27 - 7	
4 - D	10 - S	16 - W	22 - 2	28 - 8	
5 - E	11 - J	17 - P	23 - 3	29 - 9	
6 - F	12 - K	18 - Q	24 - 4	30 - T	

Illinois, Florida, and Wisconsin encode the surname, first name, middle initial, date of birth, and sex by a quite sophisticated scheme. The first character of the license number is the first character of the name. The next three characters are obtained by applying the "Soundex Coding System" to the surname as follows:

1. Delete all occurrences of h and w.
2. Assign numbers to the remaining letters as follows:
   $b, f, p, v \to 1$ $\qquad l \to 4$
   $c, g, j, k, q, s, x, z \to 2$ $\qquad m, n \to 5$
   $d, t \to 3$ $\qquad r \to 6$
   (No values are assigned to a, e, i, o, u, and y.)
3. If two or more letters with the same numeric value are adjacent, omit all but the first. (Here a, e, i, o, u, and y act as separators.) For example, Schworer becomes Sorer and Hughgill becomes Ugil.

4. Delete the first character of the original name if still present.
5. Delete all occurrences of a, e, i, o, u, and y.
6. Use the first three digits corresponding to the remaining letters; append trailing zeros if less than three letters remain.

Here are some examples: Schworer → S-660; Hughgill → H-240; Skow → S-000; Sachs → S-200; Lennon → L-550; McCartney → M-263.

We parenthetically remark that the Soundex System was designed so that likely misspellings of a name would nevertheless result in the correct coding of the name. For example, frequent misspellings of my name are: Gallion, Gillian, Galian, Galion, Gilliam, Gallahan, and Galliam. Observe that all of these yield the same coding as Gallian. We also mention that the above method differs somewhat from the system called Soundex by Knuth in [5, p. 392].

The next three digits are determined by summing numbers that correspond to the first name and middle initial. The scheme for doing this begins with the block 000 for the letter A and makes jumps of 20 for especially common names and each subsequent letter of the alphabet. A small portion of this scheme is given in TABLE 3. The values assigned to the middle initial are given in TABLE 4.

So Aaron G. Schlecker would be coded as S426-007 (S426 from Schlecker; 000 for Aaron + 7 for "G"), while Anne P. Schlecker would be coded as S426-055.

The last five digits of Illinois and Florida numbers capture the year and date of birth as well as the sex. In Illinois, each day of the year is assigned a three-digit number in sequence beginning with 001 for January 1. However, each month is assumed to have 31 days. Thus, March 1 is given 063. These numbers are then used to identify the month and day of birth of male drivers. For females, the scheme is identical except January 1 begins with 601. The last two digits of the year of birth, separated by a dash (probably for camouflage), are listed in the 5th and 4th positions from the end of the driver's license number. Thus, a male born on July 18, 1942, would have the last five digits 4-2204 while a female born on the same day would have 4-2804. When necessary, Illinois adds an extra character to avoid duplications.

TABLE 3. Illinois, Florida, Wisconsin given name or first initial code.

000	–	*A*
020	–	Albert, Alice
040	–	Ann, Anna, Anne, Annie, Arthur
060	–	*B*
080	–	Bernard, Bette, Bettie, Betty
100	–	*C*

TABLE 4. Illinois, Florida, Wisconsin middle initial code.

0 - none	10 - *J*
1 - *A*	11 - *K*
2 - *B*	12 - *L*
3 - *C*	13 - *M*
4 - *D*	14 - *N*, *O*
5 - *E*	15 - *P*, *Q*
6 - *F*	16 - *R*
7 - *G*	17 - *S*
8 - *H*	18 - *T*, *U*, *V*
9 - *I*	19 - *W*, *X*, *Y*, *Z*

Among the 9,397,518 licenses on file on January 1, 1987, this occurred in 14,856 instances. Of these, 55 numbers corresponded to three individuals (excluding the extra digit). No number corresponded to more than three people.

The scheme to identify birthdate and sex in Florida is the same as Illinois except each month is assumed to have 40 days and 500 is added for women. For example, the five digits 49583 belong to a woman born on March 3, 1949.

Wisconsin employs the same scheme as Florida to generate the first 12 of their 14 characters. The thirteenth character is an integer issued sequentially beginning with 0 to people who share the same first 12 characters. The fourteenth character is a check digit.

A Missouri driver's license number has 16 characters. The first 13 characters are obtained by applying a hashing function to the first five letters of the surname, the first three letters of the first name and the middle initial. (The method of encoding is similar to that used by Florida.) The final three characters are a function of the month and day of birth and sex. For a male born in month m and day d the three digits are $63m + 2d$. For a female, the corresponding formula is $63m + 2d + 1$. Thus the number of a woman born on March 4 has the final three characters 198. To avoid duplications, Missouri assigns a 17th character. Among the first 3,921,922 numbers issued, 31,719 have a 17th character.

Last, we discuss the scheme employed by Minnesota, Michigan, and Maryland. The number is a function of last name, first name, middle name, month and date of birth. The first four characters are determined by the Soundex System, as was the case for Illinois, Florida, and Wisconsin. The first and middle names account for the next six characters and the same algorithm is applied to both names. In the majority of cases the first two characters of the name determine the desired three digits for each name (see TABLE 5 for a sample); for common pairs of leading letters such as Al or Ja, the third letter is invoked (see TABLE 6); 11 three-digit numbers are uniquely assigned to the 11 most popular names (e.g., 189 ↔ Edward; 210 ↔ Elizabeth).

TABLE 5. Minnesota, Michigan, Maryland code for first and middle names beginning with *A* except *Al*.

		A	027		
Aa	028	*Aj*	037	*As*	072
Ab	029	*Ak*	038	*At*	073
Ac	030	*Al*	—	*Au*	074
Ad	031	*Am*	066	*Av*	075
Ae	032	*An*	067	*Aw*	076
Af	033	*Ao*	068	*Ax*	077
Ag	034	*Ap*	069	*Ay*	078
Ah	035	*Aq*	070	*Az*	079
Ai	036	*Ar*	071		

The final three digits are based on month and date of birth (but not year). Each day of the year is assigned a three-digit number in a monotonically increasing fashion. Although the usual pattern is to alternate increments of 3 and 2, there are numerous seemingly random increments at unpredictable dates. The month of March illustrates this behavior well. Notice from TABLE 7 that March 1 is assigned 159. Subsequent days are assigned values by increments of 3 and 2 in alternating fashion until March 8. Then there is an increment of 5. Notice the jump of 20 between March 19 and March 20.

These gaps serve a practical purpose. In the event that there are two or more individuals born on the same month and date and with names so similar that the

TABLE 6. Minnesota, Michigan, Maryland code for first and middle names beginning with *Al*.

		Al	039		
Ala	040	*Alj*	049	*Als*	058
Alb	041	*Alk*	050	*Alt*	059
Alc	042	*All*	051	*Alu*	060
Ald	043	*Alm*	052	*Alv*	061
Ale	044	*Aln*	053	*Alw*	062
Alf	045	*Alo*	054	*Alx*	063
Alg	046	*Alp*	055	*Aly*	064
Alh	047	*Alq*	056	*Alz*	065
Ali	048	*Alr*	057		

TABLE 7. Minnesota, Michigan, Maryland code for dates in March.

	March - 158	
1 - 159	11 - 187	21 - 229
2 - 162	12 - 189	22 - 232
3 - 164	13 - 192	23 - 234
4 - 167	14 - 194	24 - 237
5 - 169	15 - 197	25 - 239
6 - 172	16 - 199	26 - 242
7 - 174	17 - 202	27 - 244
8 - 177	18 - 204	28 - 247
9 - 182	19 - 207	29 - 249
10 - 184	20 - 227	30 - 252
		31 - 254

hashing function does not distinguish between them (e.g., Jill Paula Smith and Jimmy Paul Smythe), the first person who applies for a license is assigned the number given by the algorithm while the second person is assigned the next higher number thereby using one of the numbers in the gap for birthdays. For example, if Jill Paula Smith is born on March 2 and is the first to receive the combination S530-441-675-162 as determined by the algorithm, then the next person who yields the same number is assigned S530-441-675-163 instead. Once all of the higher numbers in a gap have been assigned, lower numbers are used. Thus the third applicant with a name yielding the combination S530-441-675 born on March 2 would be assigned the last three digits 161. As of 1984, this scheme had not yielded any duplications among 4,468,080 people in Maryland while of Michigan's 6,332,878 drivers by 1987 there are 56 that have a number not uniquely their own. In fact, Michigan has two numbers that are each shared by four individuals and three that are each shared by three individuals. A common cause of duplication is the custom of naming a son after the father. When both share the same birthday a duplication occurs.

Summary

TABLE 8 summarizes the information the author has discovered about the coding of driver's license numbers. Unfortunately our knowledge is incomplete. Several states (e.g., Florida, New York, Minnesota, Missouri, Wisconsin) keep their methods confidential. In some of these cases we were able to determine the coding scheme by examining data. A question mark after the letter X indicates the corresponding item is used in the coding, but we do not know the method involved. The expression (A) after an X indicates that the corresponding item is part of a scheme that is an alternative to the social security number.

TABLE 8. Summary of Schemes for Assigning Driver's License Numbers.

State	*Social Security Number*	*Computer or Sequential Number*	*Check Digit*	*Last Name Coded*	*First Name Coded*	*Middle Name Coded*	*Year of Birth Coded*	*Month of Birth Coded*	*Day of Birth Coded*	*Sex Coded*
Alabama		X								
Alaska		X								
Arizona	X									
Arkansas	X	X(A)								
California		X								
Colorado		X								
Connecticut		X						X		
Delaware		X					X			
Florida				X	X	X	X	X	X	X
Georgia	X	X(A)								
Hawaii	X									
Idaho	X	X(A)								
Illinois				X	X	X	X	X	X	X
Indiana	X	X(A)								
Iowa	X									
Kansas		X								
Kentucky	X									
Louisiana		X								
Maine		X		X	X		X	X	X	
Maryland				X	X	X		X	X	
Massachusetts	X									
Michigan				X	X	X		X	X	
Minnesota				X	X	X		X	X	
Mississippi	X									

TABLE 8.

State	*Social Security Number*	*Computer Sequential Number*	*Check Digit*	*Last Name Coded*	*First Name Coded*	*Middle Name Coded*	*Year of Birth Coded*	*Month of Birth Coded*	*Day of Birth Coded*	*Sex Coded*
Missouri				X(?)	X(?)	X(?)		X	X	X
Montana	X				X (A)		X (A)	X (A)	X (A)	X (A)
Nebraska		X								
Nevada	X	X(A)								
New Hampshire				X	X		X	X	X	
New Jersey				X	X	X	X	X		X
New Mexico		X	X mod 11							
New York				X(?)	X(?)	X(?)	X	X(?)	X(?)	?
North Carolina		X								
North Dakota	X	X (A)								
Ohio		X								
Oklahoma	X	X(A)								
Oregon		X								
Pennsylvania		X								
Rhode Island		X								
South Carolina		X								
South Dakota		X	X mod 10				X	X		
Tennessee		X	X mod 11							
Texas		X								
Utah		X	X mod 10							
Vermont		X	X mod 11							
Virginia	X	X(A)								
Washington			X mod 10	X	X	X	X	X	X	
West Virginia		X								
Wisconsin			X(?)	X	X	X		X	X	X
Wyoming		X	X(?)							

REFERENCES

1. Steve Connor, The invisible border guard, *New Scientist* Jan. 5 (1984), 9–14.
2. Joseph A. Gallian, The zip code bar code, *The UMAP Journal* 7 (1986), 191–194.
3. Joseph A. Gallian and Steven Winters, Modular arithmetic in the marketplace, *Amer. Math. Monthly* 95 (1988), 548–551.
4. H. Peter Gumm, A new class of check digit methods for arbitrary number systems, *IEEE Transactions on Information Theory* 31 (1985), 102–105.
5. Donald E. Knuth, *The Art of Computer Programming*, Vol. 3, Addison-Wesley, Reading, MA, 1973.
6. Philip M. Tuchinsky, International standard book numbers, *The UMAP Journal* 5 (1985), 41–54.
7. J. Verhoeff, *Error Detecting Decimal Codes*, Mathematical Centrum, Amsterdam, 1969.
8. E. F. Wood, Self-checking codes–an application of modular arithmetic, *Mathematics Teacher* 80 (1987), 312–316.

NOTES

Chains in Power Sets

ROGER B. NELSEN
HARVEY SCHMIDT, JR.
Lewis and Clark College
Portland, OR 97219

When studying ordered sets a natural question that arises is: *How many chains are there in a typical finite, partially ordered set?* Generic questions about partially ordered sets (posets) have a long history and this particular question about the number of chains in a partially ordered set has connections to questions dealing with ordered partitions and trees that may be traced back to work by Cayley [**3**] in 1859.

Our main goal in this article is to develop recursive and closed formulas for the total number of chains and the number of chains of a particular length in the power set of an n-element set, partially ordered by set inclusion. We also discuss a canonical relationship between these chains and the preferential arrangements of an n-element set discussed in [**7**] and [**9**].

Another objective is to use this question about chains to make two pedagogical points. First, frequently in mathematics there may be several approaches to solving a problem with each approach illuminating a different aspect of the problem. Chain counting provides such an occasion for illustrating a wide variety of techniques from classical combinatorics and, as a consequence, we sketch more than one line of attack for most of our results.

Second, most of the results in this article are subsumed in more advanced and modern treatments of combinatorics. (Explicit reference to the literature is contained in our concluding remarks.) However, the classical combinatorial techniques used here are elementary and accessible to those familiar with counting techniques, generating functions, distribution and occupancy problems, Stirling numbers, and elementary differential equations. In particular, the generating function techniques exploit the remarkable correspondence, discovered by Laplace, between set theoretic operations and operations on formal power series. A comprehensive discussion of these classical techniques may be found in most advanced undergraduate texts on combinatorics (e.g., [**11**, Chapters 2, 3, 5, and 6]).

Before we begin, a word of caution is in order. As in elementary number theory, questions about partially ordered sets may be easy to pose but quite difficult to resolve. Indeed, a simply stated open question [**12**, p. 103], closely related to the one we discuss here, is: *What is the number of antichains in the power set of an n-element set?* (An antichain is a subset of a poset in which no two elements are comparable.)

We start by fixing some notation. Identifying an arbitrary n-element set in the natural way with $X_n = \{1, 2, \dots, n\}$, we partially order the power set $\mathscr{P}(X_n)$ by set inclusion. If C: $N_0 \subset N_1 \subset \cdots \subset N_k$ is a chain in $\mathscr{P}(X_n)$, we say C has *length* k. In particular, chains consisting of a single element of $\mathscr{P}(X_n)$ have length 0. Furthermore, $a_{n,k}$ and a_n will denote, respectively, the number of chains of length k and the total number of chains in $\mathscr{P}(X_n)$. Finally, if C: $N_0 \subset N_1 \subset \cdots \subset N_k$ is a chain of length k in $\mathscr{P}(X_n)$ then necessarily $k \leqslant n$ since each containment in C is proper. Consequently, $a_{n,k} = 0$ if $k > n$ and thus $a_n = \sum_{k=0}^{n} a_{n,k} = \sum_{k=0}^{\infty} a_{n,k}$. (If $n = 0$ then $X_n = \varnothing$ and $a_{0,0} = 1$.)

Recurrence relations for $a_{n,k}$ In this section we establish two recurrence relations for $a_{n,k}$ relating (the lengths of) chains in $\mathscr{P}(X_n)$ to (the lengths of) chains in $\mathscr{P}(X_m)$ for $0 \leqslant m \leqslant n$ and an expression relating chains in $\mathscr{P}(X_n)$ to the classical Stirling numbers.

Our first goal is to establish the one-step recursion

$$a_{n+1,k} = k \cdot a_{n,k-1} + (k+2) a_{n,k}. \tag{1}$$

This proof is motivated by the observation that every chain in $\mathscr{P}(X_{n+1})$ induces a chain in $\mathscr{P}(X_n)$ simply by intersecting that chain, node by node, with the set X_n.

If C: $N_0 \subset N_1 \subset \cdots \subset N_k$ is a chain of length k in $\mathscr{P}(X_{n+1})$, where $n \geqslant 0$, then $\tilde{C} = C \cap X_n$: $N_0 \cap X_n \subseteq N_1 \cap X_n \subseteq \cdots \subseteq N_k \cap X_n$ is a chain in $\mathscr{P}(X_n)$. Note that there are two possibilities: Either the chain $\tilde{C}$ still has length k, or some of the links in the chain C collapse so that $\tilde{C}$ has length less than k. We will call the chain C *non-degenerate* if $\tilde{C}$ has length k; otherwise, call C *degenerate*. A little more can be said in the latter case, for if $N_i \cap X_n = N_{i+1} \cap X_n$ for some i, then it must be that $N_{i+1} = N_i \cup \{n+1\}$ since N_i is a proper subset of N_{i+1}. Hence, exactly one link in a degenerate chain C collapses and this occurs at the point in C where $N_i \subseteq X_n$ but $N_{i+1} \not\subseteq X_n$. Consequently, if C is degenerate, the length of $\tilde{C}$ is always $k-1$.

The next step is to count the number of chains of each type in $\mathscr{P}(X_{n+1})$. To obtain the number of non-degenerate chains in $\mathscr{P}(X_{n+1})$, let $\tilde{C}$: $N_0 \subset N_1 \subset \cdots \subset N_k$ be a fixed chain of length k in $\mathscr{P}(X_n)$. For $i = 0, 1, \ldots, k$, observe that $\tilde{C}$ can be induced by intersection from the chain

$$\tilde{C}_i: N_0 \subset N_1 \subset \cdots \subset N_{i-1} \subset N_i \cup \{n+1\} \subset \cdots \subset N_k \cup \{n+1\}$$

in $\mathscr{P}(X_{n+1})$. If we define $\tilde{C}_{k+1} = \tilde{C}$, then this construction produces $k+2$ distinct, non-degenerate chains of length k.

Next note that, if C: $N_0 \subset N_1 \subset \cdots \subset N_k$ is a non-degenerate chain of length k in $\mathscr{P}(X_{n+1})$ with $N_{i-1} \subseteq X_n$ but $N_i \not\subseteq X_n$ and $\tilde{C} = C \cap X_n$, then $C = \tilde{C}_i$, so every non-degenerate chain of length k in $\mathscr{P}(X_{n+1})$ can be generated in this fashion. (If $N_i \subseteq X_n$ for all i, then $\tilde{C} = C \cap X_n = \tilde{C}_{k+1}$.) Finally, if $\tilde{C}$ and $\overline{C}$ are two distinct chains of length k in $\mathscr{P}(X_n)$ with $\tilde{C}_i \cap X_n = \tilde{C}$ and $\overline{C}_j \cap X_n = \overline{C}$, then $\tilde{C}_i \neq \overline{C}_j$ for $i, j = 0, 1, \ldots, k+1$. Consequently there are exactly $(k+2) \cdot a_{n,k}$ non-degenerate chains of length k in $\mathscr{P}(X_{n+1})$.

The number of degenerate chains of length k in $\mathscr{P}(X_{n+1})$ can be obtained in a similar manner, for the chain $\tilde{C}$: $N_0 \subset N_1 \subset \cdots \subset N_{k-1}$ of length $k-1$ in $\mathscr{P}(X_n)$ may be induced by intersection from any one of the k chains

$$\tilde{C}_i: N_0 \subset N_1 \subset \cdots \subset N_i \subset N_i \cup \{n+1\} \subset \cdots \subset N_{k-1} \cup \{n+1\}$$

in $\mathscr{P}(X_{n+1})$ of length k. From this it is straightforward to see that there are exactly $k \cdot a_{n,k-1}$ degenerate chains of length k in $\mathscr{P}(X_{n+1})$. Combining the observations of the preceding paragraphs leads directly to equation (1).

The second recurrence relation we develop is

$$a_{n,k} = \sum_{j=0}^{n-1} (-1)^{n-j+1}\binom{n}{j}(a_{j,k} + a_{j,k-1}). \tag{2}$$

This recursion follows from a classical approach involving an application of the inclusion-exclusion principle and the observation that a given chain in $\mathscr{P}(X_n)$ either does, or does not, have X_n as one of its nodes. If we let $a'_{n,k}$ denote the number of chains of length k in $\mathscr{P}(X_n)$ that do not have X_n as a node and let $a''_{n,k}$ denote the number that do, then $a_{n,k} = a'_{n,k} + a''_{n,k}$.

First we deal with $a'_{n,k}$. Observe that if X_n is not a node of a chain C in $\mathscr{P}(X_n)$, then C must be a chain in $\mathscr{P}(X_n \setminus \{i\})$ for some $i \in X_n$. (Here $A \setminus B$ denotes the usual difference of the sets A and B.) If $A_{i,k}$ denotes the set of chains of length k in $\mathscr{P}(X_n \setminus \{i\})$ then the inclusion-exclusion principle [11, Chapter 3, (1)] asserts that

$$\begin{aligned} a'_{n,k} &= |A_{1,k} \cup A_{2,k} \cup \cdots \cup A_{n,k}| \\ &= \sum_i |A_{i,k}| - \sum_{i \neq j} |A_{i,k} \cap A_{j,k}| + \cdots + (-1)^{n+1}|A_{1,k} \cap \cdots \cap A_{n,k}|. \end{aligned}$$

Certainly $|A_{i,k}| = a_{n-1,k}$ for all i. Moreover, $A_{i,k} \cap A_{j,k}$ is just the set of chains of length k in $\mathscr{P}(X_n \setminus \{i\}) \cap \mathscr{P}(X_n \setminus \{j\}) = \mathscr{P}(X_n \setminus \{i,j\})$, so that $|A_{i,k} \cap A_{j,k}| = a_{n-2,k}$ whenever $i \neq j$. Continuing in this way through intersections of $n-k$ different $A_{i,k}$'s and recalling that $a_{j,k} = 0$ whenever $j < k$, we obtain

$$\begin{aligned} a'_{n,k} &= na_{n-1,k} - \binom{n}{2}a_{n-2,k} + \cdots + (-1)^{n-k+1}\binom{n}{n-k}a_{k,k} \\ &= \sum_{i=1}^{n-k} (-1)^{i+1}\binom{n}{i}a_{n-i,k} = \sum_{i=1}^{n} (-1)^{i+1}\binom{n}{i}a_{n-i,k}. \end{aligned}$$

To compute $a''_{n,k}$ notice that if X_n is a node of the chain C of length k in $\mathscr{P}(X_n)$, then X_n must be the largest node in C; consequently $C = C^* \cup \{X_n\}$, where C^* is a chain of length $k-1$ in $\mathscr{P}(X_n \setminus \{i\})$ for at least one $i \in X_n$. Using the notation of the previous paragraph and applying the results above, we have

$$a''_{n,k} = |A_{1,k-1} \cup A_{2,k-1} \cup \cdots \cup A_{n,k-1}| = \sum_{i=1}^{n} (-1)^{i+1}\binom{n}{i}a_{n-i,k-1}.$$

Combining the expressions for $a'_{n,k}$ and $a''_{n,k}$ yields

$$\begin{aligned} a_{n,k} &= \sum_{i=1}^{n} (-1)^{i+1}\binom{n}{i}(a_{n-i,k} + a_{n-i,k-1}) \\ &= \sum_{j=0}^{n-1} (-1)^{n-j+1}\binom{n}{j}(a_{j,k} + a_{j,k-1}), \end{aligned}$$

as asserted in (2).

Our third expression relates the numbers $a_{n,k}$ to the Stirling numbers of the second kind $S(n,k)$, specifically

$$a_{n,k} = (k+2)!S(n,k+2) + 2(k+1)!S(n,k+1) + k!S(n,k). \tag{3}$$

Our derivation of this expression exploits a canonical mapping between the chains being discussed here and the set of ordered partitions of an n-element set, the so-called rational preferential arrangements discussed in [7] and [9].

Recall that one of the interpretations of a Stirling number of the second kind $S(n,k)$, henceforth just called "Stirling number," is the number of unordered partitions of an n-element set into k classes [4, p. 204]. Since each ordering of these k classes induces a "preferential arrangement" of X_n, allowing for indifference, the number of ordered partitions of X_n into k classes is simply $k!S(n,k)$.

The mapping between chains and ordered partitions leading to (3) can now be described. For an ordered partition P of X_n into k (non-empty and disjoint) classes we will write $P\colon P_1|P_2|\cdots|P_k$. Then each chain $C\colon N_0 \subset N_1 \subset \cdots \subset N_k$ of length k in $\mathscr{P}(X_n)$ corresponds to an ordered partition P of X_n having either $k+2$, $k+1$ or k classes, respectively, as follows:

i) If $N_0 \neq \varnothing$ and $N_k \neq X_n$ then map C to the partition P with $k+2$ classes, where $P\colon N_0|N_1 \setminus N_0|\cdots|N_k \setminus N_{k-1}|X_n \setminus N_k$;
ii) if $N_0 = \varnothing$ and $N_k \neq X_n$ then map C to the partition P with $k+1$ classes, where $P\colon N_1|N_2 \setminus N_1|\cdots|N_k \setminus N_{k-1}|X_n \setminus N_k$; and finally,
iii) if $N_0 \neq \varnothing$ and $N_k = X_n$ then map C to the partition P with $k+1$ classes, where $P\colon N_0|N_1 \setminus N_0|\cdots|N_{k-1} \setminus N_{k-2}|X_n \setminus N_{k-1}$.
iv) If $N_0 = \varnothing$ and $N_k = X_n$ then map C to the partition P with k classes, where P: $N_1|N_2 \setminus N_1|\cdots|N_{k-1} \setminus N_{k-2}|X_n \setminus N_{k-1}$.

Moreover, it is clear that every ordered partition of X_n into either $k+2$ or k classes can be obtained uniquely from a chain of length k in $\mathscr{P}(X_n)$ in this manner and, in addition, each partition into $k+1$ classes can be derived from two distinct chains of length k. Expression (3) now follows directly from the observation that $k!S(n,k)$ counts the number of ordered partitions of X_n into k classes.

Relations (1) and (3) are closely linked; beginning with (3) and using the well-known recurrence relation $S(n+1,k) = S(n,k-1) + kS(n,k)$ for the Stirling numbers, it is straightforward to derive (1) directly without resorting to another counting argument.

A closed sum for $a_{n,k}$ This section is devoted to the development of the closed sum formula

$$a_{n,k} = \sum_{j=0}^{k} (-1)^j \binom{k}{j} (k+2-j)^n. \tag{4}$$

Our first approach is a classical one: We consider a formal power series, the exponential generating function of the form

$$A_k(x) = \sum_{n=0}^{\infty} a_{n,k} \frac{x^n}{n!} \qquad (k \geqslant 0), \tag{5}$$

where we use the yet-to-be-determined sequence $\{a_{n,k} \mid n = 0,1,2,\dots\}$ as the coefficients in the series. The strategy is to "operate" on this series in some manner and, using a recurrence relation, to obtain a functional equation that, when solved, explicitly determines $A_k(x)$ and hence the coefficients in (5). (See Chapter 1 of either [11] or [14].)

We begin our proof of (4) by observing that $a_{n,0} = 2^n$ since each subset of X_n is a chain of length 0, so that

$$A_0(x) = \sum_{n=0}^{\infty} 2^n \frac{x^n}{n!} = e^{2x}.$$

We will use recurrence relation (1) to derive the functions $A_k(x)$ for $k \geqslant 1$. It is also possible to develop this result beginning with (2); we leave this to the interested reader, as both arguments are similar. Differentiating (5), re-indexing, and introducing (1), we have

$$\begin{aligned} A'_k(x) &= \sum_{n=1}^{\infty} a_{n,k} \frac{x^{n-1}}{(n-1)!} = \sum_{n=0}^{\infty} a_{n+1,k} \frac{x^n}{n!} \\ &= \sum_{n=0}^{\infty} \{k \cdot a_{n,k-1} + (k+2) \cdot a_{n,k}\} \frac{x^n}{n!} \\ &= k \cdot A_{k-1}(x) + (k+2) \cdot A_k(x); \end{aligned} \tag{6}$$

that is, $A_k(x)$ satisfies a first-order, linear, differential equation.

From the initial conditions $A_k(0) = a_{0,k} = 0$ (for $k \geqslant 1$), the fact that $A_0(x) = e^{2x}$, and equation (6), we infer that $A_1(x) = e^{2x}(e^x - 1)$, whereupon induction yields

$$A_k(x) = e^{2x}(e^x - 1)^k. \tag{7}$$

Using the binomial theorem and then expanding $A_k(x)$ in a power series gives

$$\begin{aligned} A_k(x) &= e^{2x}(e^x - 1)^k = e^{2x} \sum_{j=0}^{k} (-1)^j \binom{k}{j} (e^x)^{k-j} \\ &= \sum_{j=0}^{k} (-1)^j \binom{k}{j} e^{(k-j+2)x} \\ &= \sum_{j=0}^{k} \left\{ (-1)^j \binom{k}{j} \sum_{n=0}^{\infty} (k-j+2)^n \frac{x^n}{n!} \right\} \\ &= \sum_{n=0}^{\infty} \left\{ \sum_{j=0}^{k} (-1)^j \binom{k}{j} (k-j+2)^n \right\} \frac{x^n}{n!}. \end{aligned} \tag{8}$$

Equating coefficients in (5) and (8) concludes our proof of (4).

We note in passing that another derivation of (4) can be obtained by using the well-known expression

$$S(n,k) = \frac{1}{k!} \sum_{j=0}^{k} (-1)^j \binom{k}{j} (k-j)^n \tag{9}$$

for the Stirling numbers in (3) and then simplifying the sums to obtain (4).

The numbers $a_{n,k}$ can be expressed more compactly, and perhaps more naturally, using the Δ operator, one of several discrete analogues of the derivative from the calculus of finite differences. If we define $\Delta f(x) = f(x+1) - f(x)$ and $\Delta^{k+1} f(x) = \Delta(\Delta^k f(x))$ for $k \geqslant 1$, then it is easy to verify that

$$\Delta^k x^n = \sum_{j=0}^{k} (-1)^j \binom{k}{j} (x + k - j)^n$$

(e.g., see [**14**, p. 36, (26)]). Thus, (4) may be rewritten as

$$a_{n,k} = \Delta^k 2^n \tag{10}$$

(see [**14**, 3.11.1(a)]). Incidentally, the similar expression

$$S(n,k) = \frac{1}{k!}\Delta^k 0^n$$

for the Stirling numbers follows directly from (9).

The total number of chains in $\mathscr{P}(\mathbf{X}_n)$ Our goal for this section is to obtain a series expression for the numbers a_n that imitates the classical formula

$$\omega(n) = \frac{1}{e}\sum_{j=0}^{\infty} j^n/j!, \qquad n \geqslant 1,$$

for the Bell numbers $\omega(n)$ discovered by Dobinski in 1877 (see [**4**, p. 210, (4d)]). Recall that the Bell numbers $\omega(n)$ are defined to be the total number of unordered partitions of an n-element set so that

$$\omega(n) = \sum_{k=1}^{n} S(n,k), \qquad n \geqslant 1.$$

The relationship (3) between $a_{n,k}$ and the Stirling numbers will enable us to establish a similar series expression for a_n, namely

$$a_n = 2\sum_{j=2}^{\infty} j^n 2^{-j}, \qquad n \geqslant 1. \tag{11}$$

To verify this, we first relate a_n to the total number $f_n = \sum_{k=0}^{n} k!S(n,k)$ of ordered partitions or preferential arrangements of X_n (see [**7**], [**9**], and [**13**, #1191]). Using (3) in the sum $a_n = \sum_{k=0}^{n} a_{n,k}$ and re-indexing, we may write

$$\begin{aligned} a_n &= 4\sum_{k=0}^{n} k!S(n,k) - 1 \\ &= 4f_n - 1, \qquad n \geqslant 1, \end{aligned} \tag{12}$$

a result that we will discuss more completely in the next section (see [**14**, 3.15.9 and 3.15.10]).

Returning to our derivation of a series expression for a_n, and using (9) and (12), we have

$$\begin{aligned} a_n &= 4\sum_{k=0}^{n} k!S(n,k) - 1 \\ &= 4\sum_{k=0}^{n}\left\{\sum_{j=0}^{k}(-1)^j\binom{k}{j}(k-j)^n\right\} - 1 \\ &= 4\sum_{k=0}^{n}\left\{\sum_{j=0}^{k}(-1)^{k-j}\binom{k}{j}j^n\right\} - 1. \end{aligned}$$

But, resorting to [10], we can write

$$\sum_{k=0}^{n}\sum_{j=0}^{k}(-1)^{k-j}\binom{k}{j}(x+j)^n = \frac{1}{2}\sum_{j=0}^{\infty}(x+j)^n 2^{-j},$$

so that, when $x = 0$, it follows that

$$a_n = 4\left\{\frac{1}{2}\sum_{j=0}^{\infty} j^n 2^{-j}\right\} - 1$$

$$= 2\sum_{j=2}^{\infty} j^n 2^{-j},$$

as asserted.

A second approach to (11) is afforded by another generating function argument, this time for the sequence $\{a_n \mid n = 0, 1, 2, \dots\}$. If we let $A(x) = \sum_{n=0}^{\infty} a_n(x^n/n!)$ and replace a_n by $\sum_{k=0}^{\infty} a_{n,k}$, then, upon reversing the order of summation, it is easy to see that

$$A(x) = \frac{e^{2x}}{2 - e^x}. \tag{13}$$

Now expanding $A(x)$ in a power series, we have

$$A(x) = \frac{1}{2}e^{2x}\frac{1}{1 - \frac{1}{2}e^x} = \frac{1}{2}e^{2x}\sum_{j=0}^{\infty}\left(\frac{e^x}{2}\right)^j$$

$$= \frac{1}{2}\sum_{j=0}^{\infty} e^{(j+2)x}\cdot 2^{-j} = 2\sum_{j=0}^{\infty} e^{(j+2)x}\cdot 2^{-(j+2)}$$

$$= 2\sum_{j=2}^{\infty}\left\{\sum_{n=0}^{\infty} j^n\frac{x^n}{n!}\right\}\cdot 2^{-j} = \sum_{n=0}^{\infty}\left\{2\sum_{j=2}^{\infty} j^n\cdot 2^{-j}\right\}\frac{x^n}{n!}.$$

Comparing the latter expression with the definition of $A(x)$ and equating coefficients yields (11).

Chains and preferential arrangements TABLE 1 contains the values of a_n and $a_{n,k}$ for $0 \leqslant k \leqslant n \leqslant 10$. Since neither (4) nor (11) lends itself well to calculation, as is often the case with alternating and infinite sums, recurrence relation (1) was used to generate the $a_{n,k}$'s with the boundary condition that $a_{n,0} = 2^n$ for all n; then a_n was obtained simply by summing the $a_{n,k}$'s on k.

TABLE 1. a_n and $a_{n,k}$ for $0 \leqslant k \leqslant n \leqslant 10$

		$a_{n,k}$; $k =$										
n	a_n	0	1	2	3	4	5	6	7	8	9	10
0	1	1										
1	3	2	1									
2	11	4	5	2								
3	51	8	19	18	6							
4	299	16	65	110	84	24						
5	2163	32	211	570	750	480	120					
6	18731	64	665	2702	5460	5880	3420	720				
7	189171	128	2059	12138	35406	57120	52080	25200	5040			
8	2183339	256	6305	52760	213444	484344	650160	514088	221760	40320		
9	28349043	512	19171	223290	1225230	3759840	6972840	8013600	5594400	2177280	362880	
10	408990251	1024	58025	931052	6796020	27459960	67609080	105945840	106444800	66528000	23587200	3628800

In (12) we were able to obtain, by means of the Stirling numbers, the explicit relationship $a_n = 4f_n - 1$ between chains and ordered partitions. In fact, it is possible to demonstrate this connection in other ways. We leave it as an exercise for the reader to develop (12) directly, without recourse to the Stirling numbers, by using a slight modification of the mapping between chains and ordered partitions that was used to establish (3).

Instead, we use an approach to (12) that employs the generating functions for the sequences $\{a_n\}$ and $\{f_n\}$ and that, at the same time, allows us to pose a general question about counting structures defined in terms of the set X_n. In [**7**, eq. (2) and **9**, eq. (4)] it is shown that $1/(2-e^x)$ is the generating function for the sequence $\{f_n\}$, so we are led to generalize the function $A(x)$ from (13) to

$$P(x,t)=\frac{e^{tx}}{2-e^x}=\sum_{n=0}^{\infty}p_n(t)\frac{x^n}{n!}.$$

Then $P(x,2)=A(x)$ and $P(x,0)=1/(2-e^x)$, so that $a_n=p_n(2)$ and $f_n=p_n(0)$.

Using $P(x,t)$, it is straightforward to obtain many identities among the polynomials $p_n(t)$ that parallel those known for the Bernoulli and Euler polynomials [**1**, p. 117 and p. 136]. These identities can be used to obtain many interesting recurrence formulas for the a_n's and f_n's. However, for our purposes we need only one, namely,

$$2p_n(t)-p_n(t+1)=t^n, \tag{14}$$

which can be seen by noticing that

$$\sum_{n=0}^{\infty}\left[2p_n(t)-p_n(t+1)\right]\frac{x^n}{n!}=2P(x,t)-P(x,t+1)$$

$$=\frac{2e^{xt}}{2-e^x}-\frac{e^{(t+1)x}}{2-e^x}=e^{tx}=\sum_{n=0}^{\infty}t^n\frac{x^n}{n!}.$$

It is now possible to demonstrate (12) formally. Applying (14) twice, once with $t=0$ and once with $t=1$, we get $2p_n(0)-p_n(1)=0$ and $2p_n(1)-p_n(2)=1$, from which it follows that

$$a_n=p_n(2)=2p_n(1)-1=2[2p_n(0)]-1=4p_n(0)-1=4f_n-1.$$

Concluding remarks We begin with a natural question that can be framed in terms of the function $P(x,t)$ introduced in the preceding section. We have shown that, for two selected values of t, the function $P(x,t)$ generates sequences of numbers associated with certain structures defined on X_n or $\mathcal{P}(X_n)$, namely preferential arrangements of X_n when $t=0$ and chains in $\mathcal{P}(X_n)$ when $t=2$. An obvious question concerns the sequences generated for other values of t. While the recursion given in (14) identifies inductively the numeric character of these sequences, we have not been able to discover any natural structures on X_n or $\mathcal{P}(X_n)$ that are enumerated by such sequences.

Finally, as mentioned in the introduction, during the last 20 years much of classical combinatorics has undergone a great deal of modern treatment and generalization, leading to a broad range of applications. One approach using the notion of an *incidence algebra* was initially discussed in [**5**] and excellent, and perhaps more accessible, accounts are also contained in Chapters 2 and 3 of [**14**] or in parts of Chapters II, IV, and V of [**1**]. Each of these sources contains an extensive bibliography and a "Notes" section at the end of each chapter providing historical commentary on the references. An introduction and survey to a homological algebra approach (Cohen-Macauley posets) to counting is given in [**2**] together with citations for the original papers as well as applications to group theory and ring theory (see also [**14**, Chapter 3]). Finally, an alternate approach to a general theory of generating functions, utilizing the notion of *tagged configurations*, may be found in Chapters 2 and 3 of **[6]**.

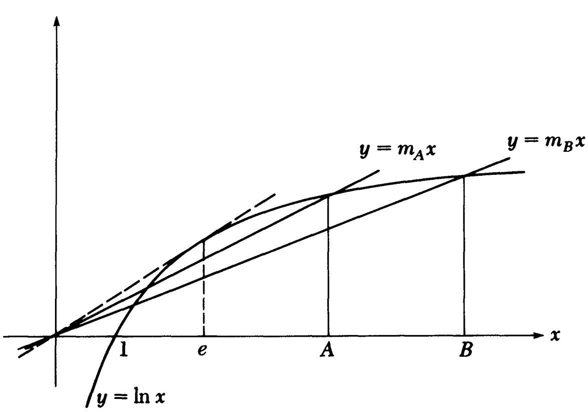
Acknowledgements. The authors wish to thank several referees for their generous help in developing the bibliography and their colleagues at Lewis and Clark College for many useful discussions regarding the organization of this article.

REFERENCES

1. M. Aigner, *Combinatorial Theory*, Springer-Verlag, New York, 1979.
2. A. Björner, A. M. Garsia, and R. P. Stanley, *An Introduction to Cohen-Macauley Partially Ordered Sets*, (I. Rival, ed.), D. Reidel Publishing Co., Boston, 1982, 583–615.
3. A Cayley, On the analytical forms called trees, *Phil. Mag.* 18 (1859), 374–378.
4. L. Comtet, *Advanced Combinatorics: the Art of Finite and Infinite Expansions*, Reidel, Holland, 1974.
5. P. Doubilet, G.-C. Rota, and R. P. Stanley, On the foundations of combinatorial theory (VI): The idea of a generating function, *Proc. 6th Berkeley Symposium on Math. Stat. and Prob.*, Vol. 2: *Probability Theory*, University of California, (1972), 267–318.
6. J. P. Golden and D. M. Jackson, *Combinatorial Enumeration*, Wiley-Interscience, New York, 1983.
7. O. A. Gross, Preferential arrangements, *Amer. Math. Monthly* 69 (1962), 4–8.
8. D. J. Kleitman and B. L. Rothschild, Asymptotic enumeration of partial orders on a finite set, *Trans. Amer. Math. Soc.* 205 (1975), 205–220.
9. Elliott Mendelson, Races with ties, this MAGAZINE 55 (1982), 170–175.
10. R. B. Nelsen, Elementary Problem 3062, *Amer. Math. Monthly* 91 (1984), 580 (for the solutions see [1987, 376]).
11. J. Riordan, *An Introduction to Combinatorial Analysis*, John Wiley & Sons, New York, 1978.
12. I. Rival (ed.), Unsolved problems, *Order* 1 (1984), 103–105.
13. N. J. A. Sloane, *A Handbook of Integer Sequences*, Academic Press, New York, 1973.
14. R. P. Stanley, *Enumerative Combinatorics*, Vol. I, Wadsworth and Brooks/Cole, Monterey, CA, 1986.

Proof without Words: Comparing B^A and A^B for $A < B$.

$$e \leqslant A < B \Rightarrow m_A > m_B$$
$$\Rightarrow \frac{\ln A}{A} > \frac{\ln B}{B}$$
$$\Rightarrow A^B > B^A$$



—CHARLES D. GALLANT
ST. FRANCIS XAVIER UNIVERSITY
ANTIGONISH, NOVA SCOTIA
CANADA B2G 1C0

Chapter 3]). Finally, an alternate approach to a general theory of generating functions, utilizing the notion of *tagged configurations*, may be found in Chapters 2 and 3 of **[6]**.

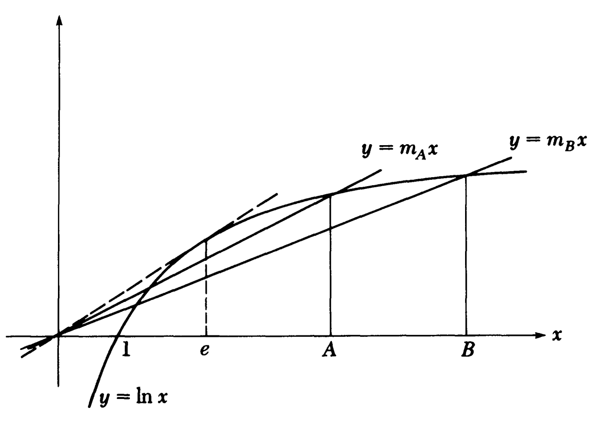
Acknowledgements. The authors wish to thank several referees for their generous help in developing the bibliography and their colleagues at Lewis and Clark College for many useful discussions regarding the organization of this article.

REFERENCES

1. M. Aigner, *Combinatorial Theory*, Springer-Verlag, New York, 1979.
2. A. Björner, A. M. Garsia, and R. P. Stanley, *An Introduction to Cohen-Macauley Partially Ordered Sets*, (I. Rival, ed.), D. Reidel Publishing Co., Boston, 1982, 583–615.
3. A Cayley, On the analytical forms called trees, *Phil. Mag.* 18 (1859), 374–378.
4. L. Comtet, *Advanced Combinatorics: the Art of Finite and Infinite Expansions*, Reidel, Holland, 1974.
5. P. Doubilet, G.-C. Rota, and R. P. Stanley, On the foundations of combinatorial theory (VI): The idea of a generating function, *Proc. 6th Berkeley Symposium on Math. Stat. and Prob.*, Vol. 2: *Probability Theory*, University of California, (1972), 267–318.
6. J. P. Golden and D. M. Jackson, *Combinatorial Enumeration*, Wiley-Interscience, New York, 1983.
7. O. A. Gross, Preferential arrangements, *Amer. Math. Monthly* 69 (1962), 4–8.
8. D. J. Kleitman and B. L. Rothschild, Asymptotic enumeration of partial orders on a finite set, *Trans. Amer. Math. Soc.* 205 (1975), 205–220.
9. Elliott Mendelson, Races with ties, this MAGAZINE 55 (1982), 170–175.
10. R. B. Nelsen, Elementary Problem 3062, *Amer. Math. Monthly* 91 (1984), 580 (for the solutions see [1987, 376]).
11. J. Riordan, *An Introduction to Combinatorial Analysis*, John Wiley & Sons, New York, 1978.
12. I. Rival (ed.), Unsolved problems, *Order* 1 (1984), 103–105.
13. N. J. A. Sloane, *A Handbook of Integer Sequences*, Academic Press, New York, 1973.
14. R. P. Stanley, *Enumerative Combinatorics*, Vol. I, Wadsworth and Brooks/Cole, Monterey, CA, 1986.

Proof without Words: Comparing B^A and A^B for $A < B$.

$$\begin{aligned} e \leqslant A < B &\Rightarrow m_A > m_B \\ &\Rightarrow \frac{\ln A}{A} > \frac{\ln B}{B} \\ &\Rightarrow A^B > B^A \end{aligned}$$



—CHARLES D. GALLANT
ST. FRANCIS XAVIER UNIVERSITY
ANTIGONISH, NOVA SCOTIA
CANADA B2G 1C0

Finding the Volume of an Ellipsoid Using Cross-Sectional Slices

JAMES E. MCKENNEY
Northern Kentucky University
Highland Heights, KY 41076

While I was consulting at a major medical complex, a medical researcher asked how he could find the volume of an object from cross-sectional slices through the object. (What actually happened was that foreign objects were imbedded in the lungs of a rat and, after some appropriate time period, a section of lung tissue was removed. Cross-sections were taken of the extracted section and some of the cross-sections contained slices through the mass formed by the body of the rat as it encased the implanted object. The researcher wanted to know the volume of the mass formed by the body when the section of tissue extracted contained only a portion of the mass.) However, the cross-sections of the mass appeared to be elliptical, and the researcher indicated that an ellipsoidal shape for the entire mass was expected. Thus I was led to consider (separate from this particular problem) whether the volume of an ellipsoid can be found exactly from some minimum number of parallel cross-sections. The purpose of this paper is to show how the volume can be calculated from three equally spaced slices.

Assume an ellipsoid is positioned in three-dimensional space with the orientation of the ellipsoid unknown. Axes may be chosen so that the ellipsoid is centered at the origin, slices through the ellipsoid are made parallel to the xy-plane, and the x and y axes lie along the principal axes of the equatorial ellipse cut by $z = 0$. Thus, without loss of generality, we may suppose that the ellipsoid has equation

$$X^tAX = 1 \tag{1}$$

where $X^t = (x, y, z)$ and $A = \begin{bmatrix} \lambda_1 & 0 & p \\ 0 & \lambda_2 & q \\ p & q & c \end{bmatrix}$, $\lambda_i > 0$. Rewriting (1) for $z = k$ leads to

$$\lambda_1\left(x + \frac{pk}{\lambda_1}\right)^2 + \lambda_2\left(y + \frac{qk}{\lambda_2}\right)^2 = 1 - k^2c + \frac{p^2k^2}{\lambda_1} + \frac{q^2k^2}{\lambda_2} \tag{2}$$

$$= 1 - k^2|A|/\lambda_1\lambda_2 \tag{3}$$

where $|A| = c\lambda_1\lambda_2 - q^2\lambda_1 - p^2\lambda_2$. Thus, whenever $k^2 < \lambda_1\lambda_2/|A|$, the cross-section of the ellipsoid at $z = k$ is the ellipse with easily determined semiaxes a and b and area πab centered at $k[-p/\lambda_1, -q/\lambda_2, 1]$. Now let

$$S = |A|/\lambda_1\lambda_2, \tag{4}$$

so that

$$a^2 = \frac{1 - k^2S}{\lambda_1}, \qquad b^2 = \frac{1 - k^2S}{\lambda_2} \quad \text{and} \quad (ab)^2 = \frac{(1 - k^2S)^2}{\lambda_1\lambda_2}. \tag{5}$$

As stated earlier, the purpose of this investigation is to determine the volume of the ellipsoid described in (1). It is well-known that an ellipsoid with semiaxes a, b, c has volume

$$V = 4\pi abc/3$$

and that $1/a^2, 1/b^2, 1/c^2$ are the eigenvalues for the symmetric matrix A. Hence

$$V = 4\pi|A|^{-1/2}/3. \tag{6}$$

A process to find the volume of the ellipsoid is now unfolding: The volume can be found from (6) if $|A|$ is known; $|A|$ can be found from (4) if S, λ_1 and λ_2 are known; S, λ_1 and λ_2 can be found from (5) for various values of a, b and k.

As three values must be found from (5), three cross-sectional ellipses need to be available so that (a_i, b_i), $i = 1,2,3$ can be used to find S, λ_1 and λ_2. Let $z = k_i$, $i = 1,2,3$ represent the values of z for the three slices. If the slices are equally spaced, then we may take $k_i = k_1 + (i-1)\Delta z$, $i = 1,2,3$, $\Delta z \neq 0$. Letting (a_i, b_i) be the semiaxes for the ellipse at $z = k_i$, $i = 1,2,3$ and using (5), we obtain

$$(a_i b_i)^2 \lambda_1 \lambda_2 = (1 - k_i^2 S)^2, \qquad i = 1,2,3.$$

Solving for $\lambda_1\lambda_2$ and equating, we further obtain

$$\frac{(1-k_1^2S)^2}{(a_1b_1)^2} = \frac{(1-k_2^2S)^2}{(a_2b_2)^2} = \frac{(1-k_3^2S)^2}{(a_3b_3)^2}. \tag{7}$$

From (4),

$$k^2 < \frac{\lambda_1\lambda_2}{|A|} = \frac{1}{S}.$$

Therefore, $0 < 1 - k^2S$, so that taking positive roots of (7) leads to

$$\frac{1-k_1^2S}{a_1b_1} = \frac{1-k_2^2S}{a_2b_2} = \frac{1-k_3^2S}{a_3b_3}.$$

Solving for S, we obtain

$$\frac{a_1b_1 - a_2b_2}{a_1b_1(k_1+\Delta z)^2 - a_2b_2k_1^2} = S = \frac{a_1b_1 - a_3b_3}{a_1b_1(k_1+2\Delta z)^2 - a_3b_3k_1^2} \tag{8}$$

Solving for k_1, we have

$$k_1 = \frac{-\Delta z}{2}\left[\frac{3a_1b_1 - 4a_2b_2 + a_3b_3}{a_1b_1 - 2a_2b_2 + a_3b_3}\right]. \tag{9}$$

Notice that k_1 is determined by information from the three cross-sectional ellipses and Δz, the distance between slices. We can now compute the volume of an ellipsoid from which three equally spaced cross-sectional slices have been taken:

Step 1. Compute a_1b_1, a_2b_2, and a_3b_3, the product of the semiaxes for each ellipse (each product is the area/π). Also note the value of Δz, the distance between slices.

Step 2. Compute k_1 from (9).

Step 3. Compute S from (8).
Step 4. Compute $\lambda_1\lambda_2$ from (7).
Step 5. Compute $|A| = \lambda_1\lambda_2 S$.
Step 6. Compute $V = \frac{4}{3}\pi|A|^{-1/2}$.

Note that it is possible to interchange (a_1, b_1) and (a_3, b_3), but the center slice must be correctly identified.

For example, let $\Delta z = 1$ cm, and suppose measurements of the cross-sections give (in cm)

first slice:	$a_1 = .620$,	$b_1 = 1.460$
second slice:	$a_2 = .920$,	$b_2 = 2.165$
third slice:	$a_3 = 1.000$,	$b_3 = 2.353$

The algorithm gives a volume $V \simeq 25.1$.

REFERENCES

1. Franklin A. Graybill, *Theory and Application of the Linear Model*, Wadsworth Publishing Co., Belmont, CA, 1965, 19–21.
2. Neil H. Timm, *Multivariate Analysis with Applications in Education and Psychology*, Brooks/Cole Publishing Co., Monterey, CA, 1975, 2–87.

Representing Primes by Binary Quadratic Forms

STEVEN GALOVICH
Carleton College
Northfield, MN 55057

JEREMY RESNICK
Schenley High School
Pittsburgh, PA

The problem of representing a prime via a binary quadratic form is one of the oldest and richest in number theory. The prototypical result in this subject is the assertion, first stated and perhaps proved by Fermat in 1640 and proved by Euler in the 1740s, that every positive prime congruent to 1 modulo 4 is a sum of two integer squares. Most proofs of this theorem have the following format: First, it is shown that if $p \equiv 1 \pmod 4$, then there exists $x \in \mathbf{Z}$ (the ring of integers) such that $x^2 \equiv -1 \pmod p$; i.e. -1 is a quadratic residue mod p. Then from the existence of the integer x, one deduces that there exist $a, b \in \mathbf{Z}$ such that $p = a^2 + b^2$.

In [2] Larson gives a proof of Fermat's statement that is interesting for at least two reasons. First, Larson's argument, which is based on ideas of Kraitchik [1] and Pólya [3], does not use the fact that -1 is a quadratic residue mod p. Second, his proof consists of a pleasant blend of algebraic and geometric arguments. The purpose of this paper is to apply Larson's method to prove the following theorem.

Step 3. Compute S from (8).
Step 4. Compute $\lambda_1\lambda_2$ from (7).
Step 5. Compute $|A| = \lambda_1\lambda_2 S$.
Step 6. Compute $V = \frac{4}{3}\pi|A|^{-1/2}$.

Note that it is possible to interchange (a_1, b_1) and (a_3, b_3), but the center slice must be correctly identified.

For example, let $\Delta z = 1$ cm, and suppose measurements of the cross-sections give (in cm)

first slice:	$a_1 = .620,$	$b_1 = 1.460$
second slice:	$a_2 = .920,$	$b_2 = 2.165$
third slice:	$a_3 = 1.000,$	$b_3 = 2.353$

The algorithm gives a volume $V \simeq 25.1$.

REFERENCES

1. Franklin A. Graybill, *Theory and Application of the Linear Model*, Wadsworth Publishing Co., Belmont, CA, 1965, 19–21.
2. Neil H. Timm, *Multivariate Analysis with Applications in Education and Psychology*, Brooks/Cole Publishing Co., Monterey, CA, 1975, 2–87.

Representing Primes by Binary Quadratic Forms

STEVEN GALOVICH
Carleton College
Northfield, MN 55057

JEREMY RESNICK
Schenley High School
Pittsburgh, PA

The problem of representing a prime via a binary quadratic form is one of the oldest and richest in number theory. The prototypical result in this subject is the assertion, first stated and perhaps proved by Fermat in 1640 and proved by Euler in the 1740s, that every positive prime congruent to 1 modulo 4 is a sum of two integer squares. Most proofs of this theorem have the following format: First, it is shown that if $p \equiv 1 \pmod 4$, then there exists $x \in \mathbf{Z}$ (the ring of integers) such that $x^2 \equiv -1 \pmod p$; i.e. -1 is a quadratic residue mod p. Then from the existence of the integer x, one deduces that there exist $a, b \in \mathbf{Z}$ such that $p = a^2 + b^2$.

In [2] Larson gives a proof of Fermat's statement that is interesting for at least two reasons. First, Larson's argument, which is based on ideas of Kraitchik [1] and Pólya [3], does not use the fact that -1 is a quadratic residue mod p. Second, his proof consists of a pleasant blend of algebraic and geometric arguments. The purpose of this paper is to apply Larson's method to prove the following theorem.

THEOREM. *Let* $p \equiv 3 \pmod 4$. *Then either there exist* $a, b \in \mathbf{Z}$ *such that* $p = a^2 + 2b^2$ *or there exist* $a, b \in \mathbf{Z}$ *such that* $p = a^2 - 2b^2$.

As is well-known, this theorem does not tell the full story. If $p \equiv 1 \pmod 8$, then there exist $a, b \in \mathbf{Z}$ such that $p = a^2 + 2b^2$ and $c, d \in \mathbf{Z}$ such that $p = c^2 - 2d^2$. Our methods do not, however, yield either of these results.

To set the stage for our proof, let p be a prime congruent to 3 modulo 4. Let $\mathbf{F}_p$ denote the field with p elements and let $V = \mathbf{F}_p \times \mathbf{F}_p = \mathbf{F}_p^2$. Let $\mathbf{L}$ denote the set of one-dimensional subspaces of V. Thus $\mathbf{L}$ is the set of lines in V that contain the origin. Then it is easy to see that $|\mathbf{L}|$ (the cardinality of $\mathbf{L}$) $= p + 1$. Let $GL_2(\mathbf{F}_p)$ denote the set of invertible 2×2 matrices with entries in $\mathbf{F}_p$. Let

$$PGL_2(\mathbf{F}_p) = GL_2(\mathbf{F}_p) \Big/ \left\{ \begin{pmatrix} x & 0 \\ 0 & x \end{pmatrix} \Big| x \in \mathbf{F}_p^* \right\}$$

be the 2×2 projective general linear group over $\mathbf{F}_p$. If $C \in GL_2(\mathbf{F}_p)$, then we let $\overline{C}$ denote the element of $PGL_2(\mathbf{F}_p)$ determined by C. Let $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $R = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, $A = \begin{pmatrix} 0 & -1 \\ 2 & 0 \end{pmatrix}$, and $B = \begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix}$. Then $G = \{\overline{I}, \overline{R}, \overline{A}, \overline{B}\}$ is a subgroup of $PGL_2(\mathbf{F}_p)$ isomorphic to the Klein four-group.

If $l \in \mathbf{L}$ is the subspace spanned by $v \in V$, then we write $l = [v]$. For $\overline{C} \in PGL_2(\mathbf{F}_p)$, we define $\overline{C}(l) = [C(v)]$. Notice that $\overline{C}(l)$ depends neither on the element $C \in GL_2(\mathbf{F}_p)$ that represents $\overline{C}$ nor on the element $v \in V$ that spans l. Thus each element of $PGL_2(\mathbf{F}_p)$ defines a function from the set $\mathbf{L}$ to itself. Moreover, the group $PGL_2(\mathbf{F}_p)$ *acts* on the set $\mathbf{L}$. Specifically, for all $l \in \mathbf{L}$, $\overline{I}(l) = l$; and, for all $\overline{C}_1$, $\overline{C}_2 \in PGL_2(\mathbf{F}_p)$ and for all $l \in \mathbf{L}$, $(\overline{C}_1\overline{C}_2)(l) = \overline{C}_1(\overline{C}_2(l))$. We now focus on the action of the group G on the set $\mathbf{L}$.

Let

$$\begin{aligned}
\mathbf{L}(\overline{I}) &= \{l \in \mathbf{L} | l \text{ is fixed by } \overline{I} \text{ only}\} \\
&= \{l \in \mathbf{L} | g(l) \neq l \text{ for all } g \in G - \{\overline{I}\}\}, \\
\mathbf{L}(\overline{R}) &= \{l \in \mathbf{L} | l \text{ is fixed by } \overline{I} \text{ and } \overline{R} \text{ only}\}, \\
\mathbf{L}(\overline{A}) &= \{l \in \mathbf{L} | l \text{ is fixed by } \overline{I} \text{ and } \overline{A} \text{ only}\}, \\
\mathbf{L}(\overline{B}) &= \{l \in \mathbf{L} | l \text{ is fixed by } \overline{I} \text{ and } \overline{B} \text{ only}\}, \text{ and} \\
\mathbf{L}(G) &= \{l \in \mathbf{L} | l \text{ is fixed by } g \text{ for all } g \in G\}.
\end{aligned}$$

Let $\alpha(\overline{I})$, $\alpha(\overline{R})$, $\alpha(\overline{A})$, $\alpha(\overline{B})$, and $\alpha(G)$ denote respectively the cardinalities of these sets.

Intuitively we can think of the elements $\overline{R}$, $\overline{A}$, and $\overline{B}$ as reflections. The element R is a reflection in V about the line $\{(x, 0) | x \in \mathbf{F}_p\}$. The elements A and B are not themselves reflections in V. As elements of $PGL_2(\mathbf{F}_p)$, however, $\overline{A}$ and $\overline{B}$ have order 2. Our goal is to show that there exist $l_1, l_2 \in \mathbf{L}$ such that either $\overline{A}(l_1) = l_1$ or $\overline{B}(l_2) = l_2$. Thus either l_1 is left fixed by $\overline{A}$ or l_2 is left fixed by $\overline{B}$, and we can regard either $\overline{A}$ as a reflection in its action on $\mathbf{L}$ about the line l_1 or $\overline{B}$ as a reflection in $\mathbf{L}$ about the line l_2.

To show that $\overline{A}$ or $\overline{B}$ fixes a line in $\mathbf{L}$, we show that either $\alpha(\overline{A}) \neq 0$ or $\alpha(\overline{B}) \neq 0$. It follows that either $\mathbf{L}(\overline{A}) \neq \emptyset$ or $\mathbf{L}(\overline{B}) \neq \emptyset$. If $\mathbf{L}(\overline{A}) \neq \emptyset$, then we deduce that $p = a^2 + 2b^2$ for some $a, b \in \mathbf{Z}$. If $\mathbf{L}(\overline{B}) \neq \emptyset$, then we prove that $p = a^2 - 2b^2$ for some $a, b \in \mathbf{Z}$. Our proof requires the following lemmas.

LEMMA 1. *$\alpha(\bar{I})$ is divisible by* 4.

Proof. Let $l \in \mathbf{L}(\bar{I})$. Then l, $\bar{R}(l)$, $\bar{A}(l)$, and $\bar{B}(l)$ are distinct elements of $\mathbf{L}(\bar{I})$. Therefore, if $\mathbf{L}(\bar{I}) \neq \emptyset$, then $\mathbf{L}(\bar{I})$ can be partitioned into four-element subsets of this type. Thus, 4 divides $\alpha(\bar{I})$.

LEMMA 2. $\mathbf{L}(\bar{R}) = \{[(1,0)],[(0,1)]\}$.

Proof. If $l = [(u,v)] \in \mathbf{L}(\bar{R})$, then $(u, -v) = x \cdot (u,v)$ where $x \in \mathbf{F}_p$. If $u \neq 0$, then $x = 1$ and $v = 0$. If $u = 0$, then $x = -1$ and $v \neq 0$. In the first case, $l = [(u,0)] = [(1,0)]$ while in the second case, $l = [(0,v)] = [(0,1)]$.

LEMMA 3. $\mathbf{L}(G) = \emptyset$.

Proof. $\mathbf{L}(G) \subseteq \mathbf{L}(\bar{R}) = \{[(1,0)],[(0,1)]\}$ by Lemma 2. Since neither line in $\mathbf{L}(\bar{R})$ is fixed by $\bar{A}$, $\mathbf{L}(G) = \emptyset$.

Proof of the theorem. By the Lemmas 1–3,

$$|\mathbf{L}| = p + 1 = \alpha(\bar{I}) + \alpha(\bar{R}) + \alpha(\bar{A}) + \alpha(\bar{B}) + \alpha(G)$$
$$= 4k + 2 + \alpha(\bar{A}) + \alpha(B)$$

where $k \in \mathbf{Z}$. Reducing mod 4, we have

$$2 \equiv p - 1 \equiv \alpha(\bar{A}) + \alpha(\bar{B}) \pmod 4.$$

Therefore, either $\alpha(\bar{A}) \neq 0$ or $\alpha(\bar{B}) \neq 0$. We consider each case separately.

Case 1. Suppose $\alpha(\bar{A}) \neq 0$. Let $l \in \mathbf{L}(\bar{A})$ with $l = [(u,v)]$ where $u, v \in \mathbf{F}_p$. Let π: $\mathbf{Z}^2 \to \mathbf{F}_p^2 = V$ denote the group homomorphism that sends $\mathbf{Z}^2$ to V by reducing each coordinate mod p. Define

$$E = \{(x,y) \in \mathbf{Z}^2 | \pi(x,y) \in l\}.$$

E is a subgroup of $\mathbf{Z}^2$ that by the isomorphism theorems of group theory has index p in $\mathbf{Z}^2$. Thus, E is a lattice in $\mathbf{R}^2$.

Let $(b,a) \in E$ be chosen so that $(b,a) \neq (0,0)$ and the *norm of* $(b,a) = |(b,a)| = \sqrt{b^2 + a^2} \leqslant |(c,d)|$ for all $(c,d) \in E - \{(0,0)\}$. In other words, no nonzero element of E is closer to the origin than (b,a). Notice also that $(-a, 2b) \in E$.

Let $P = \{r(b,a) + s(-a,2b) | r, s \in \mathbf{R}$ and $0 \leqslant r, s < 1\}$. Thus P is the parallelogram in the plane whose vertices are $O = (0,0)$, $Y = (b,a)$, $X = (-a, 2b)$, and $Z = (b - a, a + 2b)$. We claim that P is a fundamental domain for the lattice E. To establish this assertion we must show that P contains no elements of E except for the origin.

Construct four circles each of radius $\sqrt{b^2 + a^2}$ centered at O, X, Y, and Z. Denote these circles by C_O, C_X, C_Y, C_Z. Since $|(a,2b)| = \sqrt{a^2 + 4b^2} < 2\sqrt{a^2 + b^2} = 2|(b,a)|$, the circle C_O intersects the circle C_X in the interior of P. Let A' be the point of intersection of C_O and C_X within P (FIGURE 1).

Similarly, C_Y and C_Z intersect in the interior of P at a point that we label C'. Finally, let B' and D' respectively denote the points within P at which the circles C_X and C_Z, and C_O and C_Y intersect.

Let Q be the region in P bounded by the arcs joining A' and B', B' and C', C' and D', and D' and A'. Since no elements of E lie closer to O than Y, no elements of E

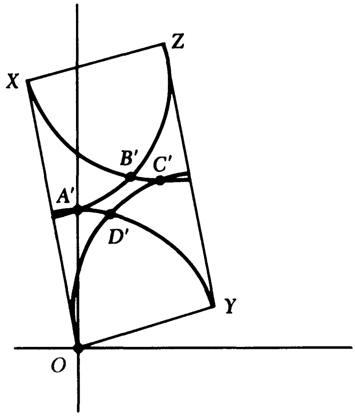


FIGURE 1

lie in the interiors of the circles C_O, C_X, C_Y, and C_Z except for the points O, X, Y, and Z. To complete the proof, we must show that $E \cap Q = \emptyset$.

Suppose $(c, d) \in E \cap Q$. Note that if $(c, d) = (1/2)Z$, then a and b must be even, which implies that $(1/2)(b, a)$ is a nonzero element of E closer to O than (b, a). Since $(c, d) \in E \cap Q$, $Z - (c, d) \in E \cap Q$. Thus we have two distinct elements of E in the region Q (FIGURE 2).

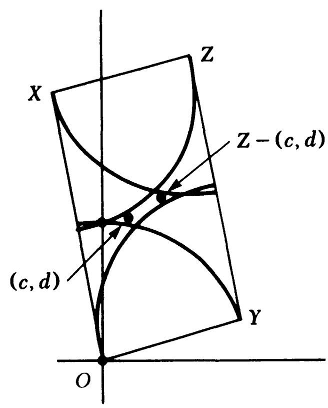


FIGURE 2

We now show that every segment with endpoints in Q has length less than $\sqrt{b^2 + a^2}$. Therefore $Z - 2(c, d)$ is a nonzero element of E that is closer to O than Y. This contradiction implies that $E \cap Q = \emptyset$, from which it follows that $E \cap P = \{0\}$.

By reasons of symmetry, $|A'B'| = |C'D'|$ and $|A'D'| = |B'C'|$. We claim that the longest diagonal of the parallelogram $A'B'C'D'$ is $A'C'$. Once this fact is established, we show that $|A'C'| < |Y|$.

To show that $A'C'$ is the longest diagonal, we show that $\angle A'D'C'$ is obtuse: Now $\angle A'D'C' = 2\pi - \angle A'D'O - \angle OD'Y - \angle YDC'$. Since triangle OYD' is equilateral, $\angle OD'Y = \pi/3$. Since $A'D'$ and $D'C'$ are chords of circles centered at O and Y respectively, $\angle AD'O < \pi/2$ and $\angle YD'C' < \pi/2$. Thus

$$\angle A'D'C' = 2\pi - \left(\frac{\pi}{3} + \frac{\pi}{2} + \frac{\pi}{2}\right) > \pi/2.$$

Finally, we prove that $|A'C'| < |Y|$. Let R and T be the midpoints of OX and YZ respectively. Construct the perpendicular bisectors of OX and YZ and label their respective intersections with YZ and OX by S and V (FIGURE 3). By its construction, *RSTV is a rectangle whose diagonal RT has length* $|Y|$. Notice that A' lies on RS and C' lies on VT; thus $|A'C'| \leqslant |RT| = |Y|$. Since A' lies in the interior of P, $A' \neq R$ and $A' \neq S$, and hence $|A'C'| < |Y|$.

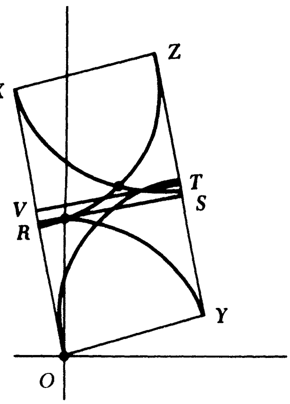


FIGURE 3

This completes the proof that $E \cap P = \{0\}$. Thus P is a fundamental domain for E. The area of P is $|\det\begin{pmatrix} b & a \\ -a & 2b \end{pmatrix}| = a^2 + 2b^2$. On the other hand, since E has index p in $\mathbf{Z}^2$, the area of P is p (see [4]). Therefore $p = a^2 + 2b^2$. By considering this equation mod 8, it follows that $p \equiv 3 \pmod 8$.

Case 2. Suppose $\alpha(\overline{B}) \neq 0$. Then an argument entirely analogous to that presented in Case 1 can be carried out. The lattice E is defined as in Case 1 and the parallelogram P has as its vertices $(0,0)$, (b, a), (a nonzero point in E closest to the origin), $(a, 2b)$, and $(a + b, a + 2b)$. As in Case 1, one shows that P is a fundamental domain for E whose area is $p = |a^2 - 2b^2|$. Thus, either $p = a^2 - 2b^2$ or $-p = a^2 - 2b^2$. But if $-p = a^2 - 2b^2$, then $p = (2b + a)^2 - 2(a + b)^2$. Therefore, in either case there exist $a, b \in \mathbf{Z}$ such that $p = a^2 - 2b^2$. Note that in this case, $p \equiv 7 \pmod 8$. The proof of Theorem 1 is now complete.

It is natural to ask if this technique can be extended to establish other quadratic form representation theorems for primes. Here is a specific question: Can this method be used to show that every prime congruent to 3 modulo 4 is a sum of four integer squares?

REFERENCES

1. M. Kraitchik, La Mathématique des Jeux ou Récréations Mathématiques, ch. XIII, *Le Problème des Reines*, Bruxelles, 1930, 300–356.
2. L. Larson, A theorem about primes proved on a chessboard, this MAGAZINE 50 (1977), 69–74.
3. G. Pólya, Über die "doppelt-periodischen." Lösungen des n-Damen Problems, *Math. Unterhalt. Spiele*, Dr. W. Ahrens, Zweiser Band, B. G. Teubner, Leipzig (1918), 364–374.
4. I. Stewart and D. Tall, *Algebraic Number Theory*, 2nd edition, Chapman and Hall, London, 1987.

Running Clubs—A Combinatorial Investigation

PHILLIP NISSEN
JOHN TAYLOR
University of the West Indies
Mona, Kingston 7, Jamaica

In the 1930s a running club was formed in Kuala Lumpur, Malaysia, which still meets to this day. It is known as the Hash House Harriers and has propagated similar clubs in many towns around the Pacific Basin and beyond. The original club meets every Monday evening, and one of its rules is that the route of the run should not have previously been traversed by the club.

We consider here a combinatorial problem based on the Hash House Harriers rules which generates investigations that could be used effectively at various levels of mathematics education.

A four-member running club The four founding members, Albert, Beryl, Cecil, and Daisy, each lived in a corner house around a village square (FIGURE 1). The running club they formed meets weekly and has the following rules:

Rule 1. A run must start and finish at someone's house.

Rule 2. The course must be along roads that form the village square.

Rule 3. No road can be traversed more than once.

Rule 4. No run is allowed to be repeated.

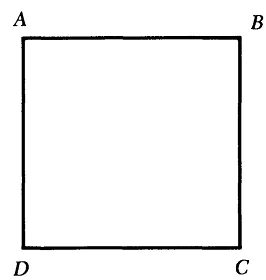


FIGURE 1

The problem we consider here is: *For how many weeks can the club meet before Rule 4 has to be broken?*

Beginning at Albert's house there are eight runs; first along one, then two, three, and finally four sides of the village square, in both a clockwise and counterclockwise direction. Therefore 32 weeks will elapse before Rule 4 is violated.

In graph-theoretic terminology (see, for example, [3]) an allowed run is an edge-sequence in which all the edges are distinct; that is, a trail. Thus we are enumerating the trails in the graph represented by FIGURE 1.

The road map of any (finite member!) running club is a finite graph whose vertices represent houses (or other points) at which runs must start or finish and whose edges represent the roads or tracks along which runs must follow. The resulting graph will be planar unless the club's neighbourhood contains bridges and underpasses.

Thus we can state our problem as follows: *Given a finite graph G, how many distinct trails does G contain?*

We should make two comments at this point. The first is that a trail is allowed to contain a given vertex more than once. Secondly, in order to maintain the spirit of the original running club problem, we shall count orientated trails; that is, trails with a given direction. (We assume that club members can detect the direction in which they run!)

For certain special classes of graphs the problem is easily solved. If G is a tree with n vertices, then the number of distinct runs is $n(n-1)$ since any ordered pair of distinct vertices determines a unique trail. The structure of the tree itself is immaterial. Presented with this investigation, we believe that students should readily discover these basic facts about trees.

If the graph is an n-sided polygon, the number of trails is $2n^2$. Here each ordered pair of (not necessarily distinct) vertices determines two trails—one in the clockwise and one in the counterclockwise direction.

Other examples of graphs to investigate are those formed by joining polygons together along common edges or common vertices. One soon discovers that the enumeration in these cases is quite complicated. As an illustration of the kind of analysis required, we now consider a class of such graphs.

The general linear club The general linear running club has $2n+2$ members and its road layout consists of n squares joined in a straight line. The graph G_n of the road layout is illustrated in FIGURE 2. Thus G_n has $2n+2$ vertices A_i and B_i for $i=0,1,\dots,n$, and $3n+1$ edges $(A_{i-1}\ A_i)$ and $(B_{i-1}\ B_i)$ for $i=1,\dots,n$, as well as $(A_i\ B_i)$ for $i=0,1,\dots,n$.

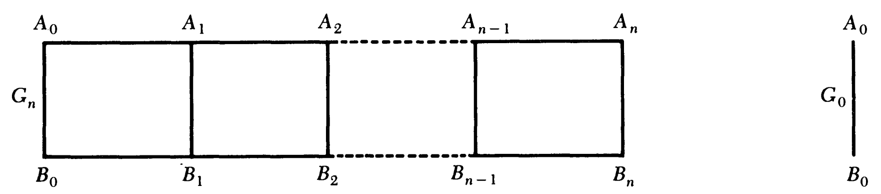


FIGURE 2

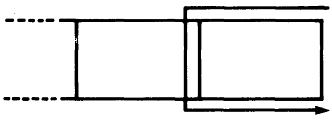
Let a_n denote the number of distinct, orientated trails in G_n. Our approach gives rise to a recursive formula for a_n, and involves consideration of certain special types of trails in G_n. With this in mind we make the following definitions:

Let

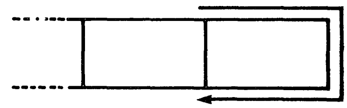
b_n = number of trails beginning at a given vertex of degree 2, say A_n,
c_n = number of trails beginning at A_n and ending at B_n,
d_n = number of trails that include the orientated edge $(A_n\ B_n)$ as an *internal* edge.

Note that all the trails enumerated in c_n are also counted in b_n, but that d_n enumerates a set of trails disjoint from those counted in b_n.

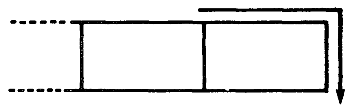
Examples. $(A_n\ A_{n-1})(A_{n-1}\ B_{n-1})(B_{n-1}\ B_n)$ is counted in both b_n and c_n.



$(A_{n-1}\ A_n)(A_n\ B_n)(B_n\ B_{n-1})$ is enumerated in d_n.



However, $(A_{n-1}\ A_n)(A_n\ B_n)$ is not counted in d_n since $(A_n\ B_n)$ is not an internal edge.



We obtain recursive formulae by imagining G_{n+1} being built from G_n by adding three edges in turn as shown in FIGURE 3.

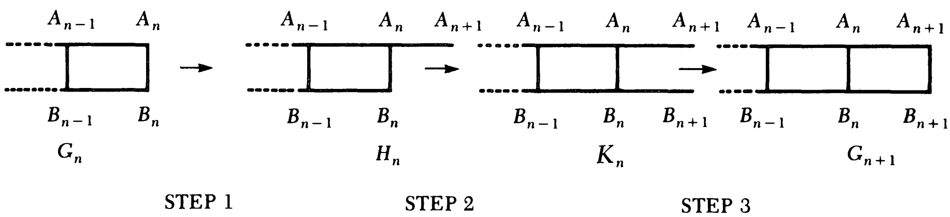


FIGURE 3

Let H_n and K_n denote the graphs formed by performing steps 1 and 2 respectively. We begin with a_n trails in G_n. Performing step 1 introduces $1 + b_n$ trails that begin at vertex A_{n+1}; namely the edge $(A_{n+1}\ A_n)$ itself and $(A_{n+1}\ A_n)$ followed by any trail beginning at A_n. Since we consider orientated trails, reversing orientation of each trail produces a new one. Therefore step 1 adds $2(1 + b_n)$ new trails, hence the graph H_n contains $a_n + 2(1 + b_n)$ trails in total.

Performing step 2 introduces $1 + b_n + c_n$ trails that begin at B_{n+1}. They are:

(i) $(B_{n+1}\ B_n)$ itself.
(ii) $(B_{n+1}\ B_n)$ followed by any trail in G_n beginning at B_n. There are b_n of these.
(iii) $(B_{n+1}\ B_n)$ followed by any trail in G_n from B_n to A_n followed by $(A_n\ A_{n+1})$. There are c_n of these.

Again the orientation of each trail may be reversed, so step 2 adds $2(1 + b_n + c_n)$ new trails. Therefore K_n contains

$$a_n + 4b_n + 2c_n + 4 \text{ distinct trails.}$$

Performing step 3 is more complicated. Each new trail introduced must include the added edge, otherwise it would be a trail in K_n. Clearly we only need to consider trails containing $(A_{n+1}\ B_{n+1})$ as opposed to the same edge with its orientation reversed. There are three possibilities for such a trail:

(i) the trail begins with $(A_{n+1}\ B_{n+1})$
(ii) the trail ends with $(A_{n+1}\ B_{n+1})$
(iii) the trail has $(A_{n+1}\ B_{n+1})$ as an internal edge.

Case (*i*). The trail is either $(A_{n+1}\ B_{n+1})$ itself or $(A_{n+1}\ B_{n+1})$ followed by a trail in K_n beginning at vertex B_{n+1}. From step 2 there are $1 + b_n + c_n$ of this latter type. Thus there are $2 + b_n + c_n$ trails in this case.

Case (*ii*). The trail must be a trail in K_n that ends at A_{n+1} followed by $(A_{n+1}\ B_{n+1})$. Again from step 2 there are $1+b_n+c_n$ of these by symmetry.

Case (*iii*). The trail must be of the form

$$(X)(A_n\ A_{n+1})(A_{n+1}\ B_{n+1})(B_{n+1}\ B_n)(Y)$$

where X and Y are (possibly empty) trails in G_n that end at A_n and begin at B_n respectively. Of course X and Y must also have no edge in common. There are four alternatives.

$X=\phi$, $Y=\phi$. There is only one trail in this case, namely, $(A_n\ A_{n+1})(A_{n+1}\ B_{n+1})(B_{n+1}\ B_n)$ itself.

$X=\phi$, $Y\neq\phi$. Since X is the empty trail Y can be any trail in G_n beginning at B_n, and there are b_n of these.

$X\neq\phi$, $Y=\phi$. Since Y is empty, X can be any one of the b_n trails in G_n ending at vertex A_n.

$X\neq\phi$, $Y\neq\phi$. The problem here is to ensure that the trails X and Y have disjoint sets of edges. We shall consider three separate subcases depending on whether X or Y is the single edge $(B_n\ A_n)$.

If X is the single edge $(B_n\ A_n)$ then Y is a trail in G_n that must begin with the edge $(B_n\ B_{n-1})$, since X and Y are disjoint. Therefore Y is $(B_n\ B_{n-1})$ itself [1 of these], or $(B_n\ B_{n-1})$ followed by a trail in G_{n-1} that begins at B_{n-1} [b_{n-1} of these], or $(B_n\ B_{n-1})$ followed by a trail in G_{n-1} from B_{n-1} to A_{n-1} followed by the edge $(A_{n-1}\ A_n)$ [c_{n-1} of these]. Hence the total number of trails of this type is $1+b_{n-1}+c_{n-1}$.

If Y is the single edge $(B_n\ A_n)$ then X is a trail in G_n that ends with the edge $(A_{n-1}\ A_n)$. By symmetry with the previous sub-case, there are $1+b_{n-1}+c_{n-1}$ possibilities for X.

Finally suppose that neither X nor Y is the edge $(B_n\ A_n)$. In this case, the trail we are considering must have the form

$$(X')(A_{n-1}\ A_n)(A_n\ A_{n+1})(A_{n+1}\ B_{n+1})(B_{n+1}\ B_n)(B_n\ B_{n-1})(Y')$$

where X' and Y' are (possibly empty) disjoint trails in G_{n-1} (FIGURE 4).

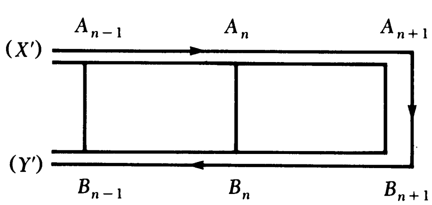


FIGURE 4

For each such trail there is a unique trail in G_n formed by replacing the edge sequence $(A_n\ A_{n+1})(A_{n+1}\ B_{n+1})(B_{n+1}\ B_n)$ by the edge $(A_n\ B_n)$. Therefore, trails of the type illustrated in FIGURE 4 are in bijective correspondence with the trails in G_n that have $(A_n\ B_n)$ as an internal edge. There are d_n trails of this kind. The aggregate for case (iii) is therefore

$$d_{n+1} = 3 + 2b_n + 2b_{n-1} + 2c_{n-1} + d_n. \tag{1}$$

Adding the totals for cases (i), (ii), and (iii) we obtain the number of new trails that include the orientated edge $(A_{n+1}\ B_{n+1})$ introduced by performing step 3. This number is

$$4b_n + 2c_n + 2b_{n-1} + 2c_{n-1} + d_n + 6.$$

As before, we can reverse the orientation of each such trail, so the total number of trails added by step 3 is

$$8b_n + 4c_n + 4b_{n-1} + 4c_{n-1} + 2d_n + 12.$$

Adding this to the number of distinct trails contained in K_n produces the recurrence relation

$$a_{n+1} = a_n + 12b_n + 6c_n + 4b_{n-1} + 4c_{n-1} + 2d_n + 16. \tag{2}$$

For this equation to be of any use we need to be able to evaluate the numbers b_n, c_n and d_n. It is an easy exercise to show that

$$c_{n+1} = 1 + c_n \quad \text{for } n \geqslant 0.$$

Since $c_0 = 1$ we have

$$c_n = n + 1.$$

Enumerating those trails added by steps 1, 2 and 3 that begin at vertex A_{n+1} gives

$$b_{n+1} = \underset{\text{Step 1}}{(1 + b_n)} + \underset{\text{Step 2}}{c_n} + \underset{\text{Step 3}}{(2 + b_n + c_n)}$$

so

$$b_{n+1} = 2b_n + 3c_n + 3. \tag{3}$$

Substituting for c_n, c_{n-1} and b_{n-1} in (3), (1) and (2) produces the following equations.

$$b_{n+1} = 2b_n + 3(n+2) \tag{4}$$

$$d_{n+1} = 3b_n + d_n - n \tag{5}$$

$$a_{n+1} = a_n + 14b_n + 2d_n + 4n + 16. \tag{6}$$

Since $a_0 = 2$, $b_0 = 1$ and $d_0 = 0$ we can compute the values for a_n, b_n and d_n recursively from (4), (5) and (6). The first few values of these are tabulated below.

n	a_n	b_n	d_n
0	2	1	0
1	32	8	3
2	170	25	26
3	596	62	99
4	1690	139	282
5	4232	296	695

Returning to the running clubs situation, we see that for the 10-member club whose road layout is the graph G_4 the rule that no run be repeated would not be violated for over 32 years!

Using the initial values $b_0 = 1$, $d_0 = 0$ and $a_0 = 2$ we may successively solve (4), (5) and (6) to obtain closed formulae for b_n, d_n and a_n. (For an account of the standard techniques for solving recursive equations see, for example, [**2**].) This yields the following formulae:

$$b_n = 5 \cdot 2^{n+1} - 3(n+3) \tag{7}$$

$$d_n = 15 \cdot 2^{n+1} - (5n^2 + 22n + 30) \tag{8}$$

$$a_n = 25 \cdot 2^{n+3} - \tfrac{2}{3}(5n^3 + 54n^2 + 192n + 297). \tag{9}$$

The complexity of both the analysis and the final result is somewhat surprising given the relatively straightforward situation we have considered. Our approach is, of course, not unique. It may be interesting to investigate whether an alternative method could produce equation (9) more efficiently.

Many related investigations suggest themselves. For instance, long distance runners may wish to know the length of the longest possible run. This turns out to be straightforward—in G_n the maximum length of a trail is $2n + 3$ edges. Enumerating such trails may prove to be an interesting problem. In addition, enumerating the trails in other classes of graphs, such as complete graphs, may be investigated. Alternatively, the number of distinct clubs of various types could be considered: For example, clubs whose graphs consist of squares joined together along common edges. This is closely related to the 'cell growth problem' [**1**, pp. 234–6].

Acknowledgement. We are grateful to the referees for many helpful comments and suggestions.

REFERENCES

1. F. Harary and E. M. Palmer, *Graphical Enumeration*, Academic Press, New York, 1973.
2. C. L. Liu, *Introduction to Combinatorial Mathematics*, McGraw-Hill, New York, 1968.
3. R. J. Wilson, *Introduction to Graph Theory*, 3rd edition, Longman, Harlow, England, 1985.

"Life is good for only two things, discovering mathematics and teaching mathematics."

—Siméon Poisson, 1781–1840

Evaluating Determinants via Generating Functions

SEYOUM GETU
Howard University
Washington, DC 20059

In this note several classes of determinants are evaluated by using generating functions. The matrices whose determinants we will be evaluating have either

A) all 1's on the super diagonal, 0's above, and identical entries on each diagonal below, perhaps with the exception of the first column, or

B) $-1, -2, -3, \dots$ on the super diagonal with 0's above and identical entries on each diagonal below including the first column.

Matrices are also constructed whose upper left corner determinants represent a given sequence. This technique gives a rather easy and systematic approach to these evaluations, often replacing clever inductions and lengthy or tricky calculations. The method evolved while the Howard University Combinatorics Group was working through the *Otto Dunkel Memorial Problem Book* [5].

Section 1 We begin with a typical example. All matrices in this section will have 1's on the super diagonal.

Example 1 [5]. Suppose

$$D = \begin{bmatrix} 1 & 1 & 0 & 0 & \cdot \\ -1 & 1 & 1 & 0 & \cdot \\ 1 & 0 & 1 & 1 & \cdot \\ -1 & 1 & 0 & 1 & \cdot \\ 1 & 0 & 1 & 0 & \cdot \\ -1 & 1 & 0 & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix} \tag{1}$$

We want to evaluate the upper left corner determinants,

$$|1|, \begin{vmatrix} 1 & 1 \\ -1 & 1 \end{vmatrix}, \begin{vmatrix} 1 & 1 & 0 \\ -1 & 1 & 1 \\ 1 & 0 & 1 \end{vmatrix}, \begin{vmatrix} 1 & 1 & 0 & 0 \\ -1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ -1 & 1 & 0 & 1 \end{vmatrix}, \dots .$$

The nth such determinant will be denoted D_n.

These upper left determinants will form the sequence $1, 2, 3, 5, \dots$ and it will be shown that we get the Fibonacci numbers.

One way to evaluate these determinants, say D_4, is as follows. Consider the system

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}, \tag{2}$$

where the right-hand side is the first column from the original matrix. By Cramer's Rule,

$$a_4 = \frac{\begin{vmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & -1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & -1 \end{vmatrix}}{1} = -D_4.$$

It is clear that

$$a_n = (-1)^{n-1} D_n, \qquad n \geqslant 1. \tag{3}$$

On the other hand, introducing generating functions for the columns of (1) and rewriting it as (2) in the general case, we get the system:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & \cdot \\ x & x & 0 & 0 & \cdot \\ 0 & x^2 & x^2 & 0 & \cdot \\ x^3 & 0 & x^3 & x^3 & \cdot \\ 0 & x^4 & 0 & x^4 & \cdot \\ x^5 & 0 & x^5 & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ \cdot \end{bmatrix} = \begin{bmatrix} 1 \\ -x \\ x^2 \\ -x^3 \\ x^4 \\ -x^5 \\ \cdot \end{bmatrix}. \tag{4}$$

[Note: $\Sigma_{n \geqslant 1} b_n x^n$ is defined as the ordinary generating function for the sequence $b_1, b_2, b_3, \ldots$ and $\Sigma_{n \geqslant 1} b_n (x^n/n!)$ as the exponential generating function for the same sequence $b_1, b_2, b_3, \ldots$. We use the notation O.G.F. for the former and E.G.F. for the latter.]

The sequence of entries in the first column in (1), which is now the right-hand side when rewritten as (2), has the generating function

$$1 - x + x^2 - x^3 + x^4 - x^5 + \cdots = \frac{1}{1+x}.$$

The 2nd, 3rd, 4th, ... columns of (1) that are now the 1st, 2nd, 3rd, ... columns of (4), excluding the 1's on the super diagonal, have generating functions

$$c(x) = x + x^3 + x^5 + x^7 + \cdots = \frac{x}{1-x^2},$$

$xc(x), x^2c(x), x^3c(x), \ldots,$ respectively. Let $A(x) = a_1 + a_2 x + a_3 x^2 + \cdots$ be the generating function for the sequence $a_1, a_2, a_3, \ldots$. Summing on both sides of (4), we get

$$A(x) + a_1(x + x^3 + x^5 + \cdots) + a_2 x(x + x^3 + x^5 + \cdots) + a_3 x^2(x + x^3 + x^5 + \cdots) + \cdots = \frac{1}{1+x}.$$

That is,

$$A(x) + a_1 \frac{x}{1-x^2} + a_2 x \frac{x}{1-x^2} + a_3 x^2 \frac{x}{1-x^2} + \cdots = \frac{1}{1+x}$$

and thus

$$A(x) + A(x)\frac{x}{1-x^2} = \frac{1}{1+x}.$$

Therefore,

$$A(x) = \frac{1-x^2}{(1+x)(1+x-x^2)} = \frac{1-x}{1+x-x^2}$$
$$= 1 - 2x + 3x^2 - 5x^3 + 8x^4 + \cdots .$$

For $n \geqslant 1$, $a_n = (-1)^{n-1}F_{n+1}$ where F_n is the nth Fibonacci number. We conclude that

$$D_n = F_{n+1}, \qquad n \geqslant 1. \tag{5}$$

The illustration in the above example leads to the following proposition:

PROPOSITION 1. *Consider the following infinite matrix with* 1*'s in the super diagonal.*

$$D = \begin{bmatrix} b_0 & 1 & 0 & 0 & \cdot \\ b_1 & c_1 & 1 & 0 & \cdot \\ b_2 & c_2 & c_1 & 1 & \cdot \\ b_3 & c_3 & c_2 & c_1 & \cdot \\ b_4 & c_4 & c_3 & c_2 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}. \tag{6}$$

Let $B(x) = \sum_{n=0}^{\infty} b_n x^n$, *and* $C(x) = \sum_{n=1}^{\infty} c_n x^n$ *be the generating functions for the sequences* $b_0, b_1, b_2, \ldots$ *and* $c_1, c_2, c_3, \ldots$, *respectively. If*

$$A(x) = \frac{B(x)}{1+C(x)} = \sum_{n=0}^{\infty} a_{n+1}x^n,$$

then $a_n = (-1)^{n-1}D_n$ *and* $1 + xA(-x)$ *is the generating function of* D_n.

Proof. As in (4) we have the system

$$\begin{bmatrix} 1 & 0 & 0 & \cdot \\ c_1x & x & 0 & \cdot \\ c_2x^2 & c_1x^2 & x^2 & \cdot \\ c_3x^3 & c_2x^3 & c_1x^3 & \cdot \\ c_4x^4 & c_3x^4 & c_2x^4 & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ \vdots \end{bmatrix} = \begin{bmatrix} b_0 \\ b_1x \\ b_2x^2 \\ b_3x^3 \\ b_4x^4 \\ \vdots \end{bmatrix} \tag{7}$$

and we get $A(x) + a_1C(x) + a_2xC(x) + a_3x^2C(x) + \cdots = B(x)$ and hence

$$A(x) = \frac{B(x)}{1+C(x)}.$$

If $x = 1$ in (7) for the $n \times n$ case we have, by Cramer's rule, as in (2), $a_n = (-1)^{n-1}D_n$ where D_n is the determinant of the $n \times n$ upper left corner matrix. The generating function of D_n is then $1 + xA(-x)$, where $a_0 = d_0 = 1$.

We now have a method of calculating the $n \times n$ determinants of the upper left corner of a matrix of the form (6) given $B(x)$ and $C(x)$ in closed form. Conversely, given any sequence, we can construct a matrix whose upper left-hand $n \times n$ determinants are the sequence elements. See Example 3, below.

Example 2. Suppose

$$D(t) = \begin{bmatrix} t & 1 & 0 & 0 & \cdot \\ 1 & 2t & 1 & 0 & \cdot \\ 0 & 1 & 2t & 1 & \cdot \\ 0 & 0 & 1 & 2t & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}. \tag{8}$$

Here $B(x) = t + x$ and $C(x) = 2tx + x^2$ and $A(x) = (t + x)/(1 + 2tx + x^2)$ and

$$1 + xA(-x) = 1 + \frac{tx - x^2}{1 - 2tx + x^2} = \frac{1 - tx}{1 - 2tx + x^2}$$

is the generating function for the Tchebycheff Polynomials of the first kind [4]. Setting $t = \cos\theta$ in (8), we get

$$D_n(\cos\theta) = \cos n\theta = \begin{vmatrix} \cos\theta & 1 & 0 & 0 & \cdot & 0 \\ 1 & 2\cos\theta & 0 & 0 & \cdot & 0 \\ 0 & 1 & 2\cos\theta & 0 & \cdot & 0 \\ 0 & 0 & 1 & 2\cos\theta & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & 1 & 2\cos\theta \end{vmatrix}_{n \times n}$$

See [4, pp. 232–233] for the derivation of $D_n(\cos\theta) = \cos n\theta$.

Example 3. Consider the sequence $1, 1, 2, 5, 14, 42, \ldots$ known as Catalan numbers, which count the number of ways to triangulate a regular n-gon, with generating function (see [7])

$$k(x) = \frac{1 - \sqrt{1 - 4x}}{2x}.$$

We want to construct a matrix whose $n \times n$ upper left corner determinant is the nth Catalan number.

By proposition 1, we want

$$A(x) = \frac{B(x)}{1 + C(x)} = k(x).$$

Clearly the choice of $B(x)$ and $C(x)$ is not unique. One choice would be

$$B(x) = \frac{1}{1 - x}k(x) = \frac{1 - \sqrt{1 - 4x}}{2x(1 - x)} = 1 + 2x + 4x^2 + 9x^3 + 23x^4 + 65x^5 + \cdots$$

and

$$C(x) = \frac{x}{1 - x} = x + x^2 + x^3 + x^4 + \cdots.$$

Hence the matrix, whose $n \times n$ upper left corner determinant (except for sign) is the nth Catalan number, will be

$$D=\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & \cdot \\ 2 & 1 & 1 & 0 & 0 & \cdot \\ 4 & 1 & 1 & 1 & 1 & \cdot \\ 9 & 1 & 1 & 1 & 1 & \cdot \\ 23 & 1 & 1 & 1 & 1 & \cdot \\ 65 & 1 & 1 & 1 & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}.$$

Another set of choices for $B(x)$ and $C(x)$ in Example 3, to get the Catalan numbers, is

$$B(x)=(1+x)\frac{1-\sqrt{1-4x}}{2x}=1+2x+3x^2+7x^3+19x^4+\cdots$$

and $C(x)=x$, which yields

$$D=\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & \\ 2 & 1 & 1 & 0 & 0 & \cdot \\ 3 & 0 & 1 & 1 & 0 & \cdot \\ 7 & 0 & 0 & 1 & 1 & \cdot \\ 19 & 0 & 0 & 0 & 1 & \\ & & \cdots & & & \end{bmatrix}.$$

We will now look at another class of determinants.

Section 2 In this section, matrices with $-1,-2,-3,\ldots$ in the super diagonals instead of 1's as in Section 1, are considered. A similar argument as in Section 1 leads, perhaps surprisingly, to a differential equation whose solution is a generating function. This method can be reversed so that if we have a sequence, we can set up a possibly interesting matrix whose upper left corner determinants are the sequence.

We start again with an example.

Example 4. Suppose

$$D=\begin{vmatrix} 0 & -1 & 0 & 0 & \cdot \\ 1 & 0 & -2 & 0 & \cdot \\ 1 & 1 & 0 & -3 & \cdot \\ 1 & 1 & 1 & 0 & \cdot \\ 1 & 1 & 1 & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{vmatrix}.$$

As in Example 1, consider the system

$$\begin{bmatrix} -1 & 0 & 0 & 0 & \cdot \\ 0 & -2x & 0 & 0 & \cdot \\ x^2 & 0 & -3x^2 & 0 & \cdot \\ x^3 & x^3 & 0 & -4x^3 & \cdot \\ x^4 & x^4 & x^4 & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}\begin{bmatrix} -f_1 \\ -f_2 \\ -f_3 \\ -f_4 \\ -f_5 \\ \cdot \end{bmatrix}=\begin{bmatrix} 0 \\ x \\ x^2 \\ x^3 \\ x^4 \\ \cdot \end{bmatrix}. \tag{9}$$

Let $F(x)=1+f_1x+f_2x^2+f_3x^3+f_4x^4+\cdots$ and $C'(x)=x+x^2+x^3+\cdots$ be ordinary generating functions. Summing on both sides of (9) we get a first-order differential equation

$$F'(x) - f_1(x^2 + x^3 + x^4 + \cdots) - f_2(x^3 + x^4 + \cdots) - \cdots = C'(x).$$

That is, $F'(x) - (F(x) - 1)C'(x) = C'(x)$ and so

$$F'(x) = F(x)C'(x). \tag{10}$$

Solving the differential equation with $F(0) = 1$ and $C(0) = 0$, we have $F(x) = e^{C(x)}$. Here

$$C'(x) = \frac{x}{1-x} \quad \text{and} \quad C(x) = -x - \ln(1-x).$$

So that

$$F(x) = \frac{1}{1-x} e^{-x},$$

which is the E.G.F. for the derangement numbers that count the number of permutations with no fixed elements [2]. Now consider the $n \times n$ upper left corner matrix of (9), let $x = 1$ and apply Cramer's rule to obtain

$$f_n = \begin{vmatrix} -1 & 0 & 0 & 0 & \cdot & 0 \\ 0 & -2 & 0 & 0 & \cdot & 1 \\ 1 & 0 & -3 & 0 & \cdot & 1 \\ 1 & 1 & 0 & -4 & \cdot & 1 \\ 1 & 1 & 1 & 0 & \cdot & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 1 & 1 & 1 & \cdot & -n \end{vmatrix}_{n \times n} = \frac{D_n}{n!}, \qquad n \geqslant 1.$$

As an exponential generating function,

$$F(x) = 1 + f_1 + 2!f_2\frac{x^2}{2!} + 3!f_3\frac{x^3}{3!} + 4!f_4\frac{x^4}{4!} + \cdots.$$

Therefore, $D_n = n!f_n$, the derangement numbers: $1, 0, 1, 2, 9, 44, \ldots$ (see [2]).

The following proposition is a modest generalization of Example 4.

PROPOSITION 2. *Suppose*

$$D = \begin{vmatrix} c_1 & -1 & 0 & 0 & \cdot \\ c_2 & c_1 & -2 & 0 & \cdot \\ c_3 & c_2 & c_1 & -3 & \cdot \\ c_4 & c_3 & c_2 & c_1 & \cdot \\ c_5 & c_4 & c_3 & c_2 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{vmatrix}. \tag{11}$$

Let $C'(x) = \sum_{n=1}^{\infty} c_n x^{n-1}$ and let $F(x) = 1 + \sum_{n=1}^{\infty} f_n x^n$ be the generating function for some numbers such that $F'(x) = C'(x)F(x)$. Then

$$D_n = \begin{cases} f_n & \text{if } F(x) \text{ is an E.G.F.} \\ n!f_n & \text{if } F(x) \text{ is an O.G.F.} \end{cases}$$

Reversing the argument, given a sequence, we can generate a matrix whose upper left corner determinants are the given sequence. (See Example 9.)

Example 5. $F(x) = e^{e^x - 1}$ is the E.G.F. for the Bell numbers. $F'(x) = e^x F(x)$ so that $C'(x) = e^x$. Let

$$D = \begin{vmatrix} 1 & -1 & 0 & 0 & \cdot \\ 1 & 1 & -2 & 0 & \cdot \\ \frac{1}{2!} & 1 & 1 & -3 & \cdot \\ \frac{1}{3!} & \frac{1}{2!} & 1 & 1 & \cdot \\ \frac{1}{4!} & \frac{1}{3!} & \frac{1}{2!} & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{vmatrix}.$$

Then the left corner determinants $D_n = f_n = B_n$ yield the Bell numbers: $1, 2, 5, 15, 52, \ldots$ (see [2]). The Bell numbers B_n count the number of ways n things can be partitioned into subsets.

Example 6. $F(x) = e^{x + (x^2/2)}$ is the E.G.F. for counting partitions of a set into blocks of size one or two. These are also known as the telephone numbers: $1, 1, 2, 4, 10, 26, 76, \ldots$. Here $F'(x) = (1 + x)e^{x + (x^2/2)} = (1 + x)F(x)$ so that $C'(x) = 1 + x$. Hence, if

$$D = \begin{vmatrix} 1 & -1 & 0 & 0 & \cdot \\ 1 & 1 & -2 & 0 & \cdot \\ 0 & 1 & 1 & -3 & \cdot \\ 0 & 0 & 1 & 1 & \cdot \\ 0 & 0 & 0 & 1 & \cdot \\ 0 & 0 & 0 & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{vmatrix},$$

then $D_n = f_n$.

Example 7. The E.G.F. for the number of permutations of a set with n elements is $F(x) = 1/(1 - x)$. Here

$$F'(x) = \frac{1}{(1 - x)^2} = \frac{1}{1 - x} F(x)$$

so that $C'(x) = 1/(1 - x)$. If

$$D = \begin{vmatrix} 1 & -1 & 0 & 0 & \cdot \\ 1 & 1 & -2 & 0 & \cdot \\ 1 & 1 & 1 & -3 & \cdot \\ 1 & 1 & 1 & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{vmatrix},$$

then $D_n = n! f_n = n!$ since $f_n = 1$.

Example 8. $F(x) = e^{-x^2 + 2tx}$ is the E.G.F. of the Hermite polynomials. $F'(x) = (2t - 2x)F(x)$ so that $C'(x) = 2t - 2x$. If

$$D(t)=\begin{vmatrix} 2t & -1 & 0 & 0 & \cdot \\ -2 & 2t & -2 & 0 & \cdot \\ 0 & -2 & 2t & -3 & \cdot \\ 0 & 0 & -2 & 2t & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{vmatrix},$$

then $D_n(t)=H_n(t)$, the Hermite polynomials [4].

Conversely, given a sequence we will construct a matrix with $-1, -2, -3, \ldots$ on the superdiagonal so that the $n\times n$ upper left corner determinants will be the given sequence.

Example 9. Consider the sequence $1,1,1,2,5,16,61,272,\ldots$ and let

$$F(x)=1+x+\frac{x^2}{2!}+\frac{2x^3}{3!}+\frac{5x^4}{4!}+\frac{16x^5}{5!}+\cdots$$

be the exponential generating function for the sequence. Then

$$1+F(x)=\sec x+\tan x$$

and

$$F'(x)=\sec x\,F(x).$$

Thus $C'(x)=\sec x$ and therefore, the required matrix is

$$D=\begin{bmatrix} 1 & -1 & 0 & 0 & 0 & \cdot \\ 0 & 1 & -2 & 0 & 0 & \cdot \\ \frac{1}{2!} & 0 & 1 & -3 & 0 & \cdot \\ 0 & \frac{1}{2!} & 0 & 1 & -4 & \cdot \\ \frac{5}{4!} & 0 & \frac{1}{2!} & 0 & 1 & \cdot \\ 0 & \frac{5}{4!} & 0 & \frac{1}{2!} & 0 & \cdot \\ \frac{61}{6!} & 0 & \frac{5}{4!} & 0 & \frac{1}{2!} & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}.$$

This technique can also be used to evaluate determinants of matrices like (11) with $1,2,3,\ldots$ on the superdiagonal. If we use $f_1, f_2, f_3, \ldots$ instead of $-f_1, -f_2, -f_3, \ldots$ in (9), then we solve the differential equation

$$F'(x)+F(x)C'(x)=2C'(x), \qquad F(0)=C(0)=0,$$

and $F(x)=2-e^{-C(x)}$.

Acknowledgement. The author wishes to thank Louis W. Shapiro for helpful suggestions in the preparation of this article and for posing the problem in the first place.

REFERENCES

1. D. Cohen, *Basic Techniques of Combinatorial Theory*, Wiley, New York, 1978.
2. L. Comtet, *Advanced Combinatorics*, Reidel, Boston, 1974.

3. T. Muir, *A Treatise on the Theory of Determinants*, Dover Publications Inc., New York, 1960.
4. A. Rabenstein, *Introduction to Differential Equations*, Academic Press, New York, 1972.
5. H. Eves and E. P. Starke, *The Otto Dunkel Memorial Problem Book*, Supplement to *Amer. Math. Monthly* 64, No 7, 1957.
6. J. Riordan, *Combinatorial Identities*, Wiley, New York, 1968.
7. L. W. Shapiro, A Catalan triangle, *Discrete Math.* 14 (1976), 83–90.

Infinite Series with Binomial Coefficients

COURTNEY MOEN
U.S. Naval Academy
Annapolis, MD 21402

As a result of Apery's 1978 announcement [4] of the irrationality of

$$\zeta(3)=\sum_{n=1}^{\infty}\frac{1}{n^3}$$

there arose interest in evaluating series of the form

$$S(k)=\sum_{n=1}^{\infty}\frac{1}{n^k\binom{2n}{n}}$$

and

$$T(k)=\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n^k\binom{2n}{n}}.$$

The special cases of $S(k)$, for $k=0,1,2,4$, and $T(k)$, for $k=0,1,2,3$ appear to have been known for some time [5]. More recently, evaluations of $S(3)$ and $S(5)$ have appeared [6].

In this note we give an alternative method for evaluating some of these sums, as well as some other series involving binomial coefficients in the denominator. We also show how the problem becomes much easier if we alter it by adding some terms to the series.

Recall the standard identity [1]

$$\frac{\Gamma(s)\Gamma(t)}{\Gamma(s+t)}=B(s,t) \tag{1}$$

where

$$\Gamma(s)=\int_0^{\infty}e^{-u}u^{s-1}\,du,\qquad s>0,$$

and

3. T. Muir, *A Treatise on the Theory of Determinants*, Dover Publications Inc., New York, 1960.
4. A. Rabenstein, *Introduction to Differential Equations*, Academic Press, New York, 1972.
5. H. Eves and E. P. Starke, *The Otto Dunkel Memorial Problem Book*, Supplement to *Amer. Math. Monthly* 64, No 7, 1957.
6. J. Riordan, *Combinatorial Identities*, Wiley, New York, 1968.
7. L. W. Shapiro, A Catalan triangle, *Discrete Math.* 14 (1976), 83–90.

Infinite Series with Binomial Coefficients

COURTNEY MOEN
U.S. Naval Academy
Annapolis, MD 21402

As a result of Apery's 1978 announcement [4] of the irrationality of

$$\zeta(3)=\sum_{n=1}^{\infty}\frac{1}{n^3}$$

there arose interest in evaluating series of the form

$$S(k)=\sum_{n=1}^{\infty}\frac{1}{n^k\binom{2n}{n}}$$

and

$$T(k)=\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n^k\binom{2n}{n}}.$$

The special cases of $S(k)$, for $k=0,1,2,4$, and $T(k)$, for $k=0,1,2,3$ appear to have been known for some time [5]. More recently, evaluations of $S(3)$ and $S(5)$ have appeared [6].

In this note we give an alternative method for evaluating some of these sums, as well as some other series involving binomial coefficients in the denominator. We also show how the problem becomes much easier if we alter it by adding some terms to the series.

Recall the standard identity [1]

$$\frac{\Gamma(s)\Gamma(t)}{\Gamma(s+t)}=B(s,t) \tag{1}$$

where

$$\Gamma(s)=\int_0^{\infty}e^{-u}u^{s-1}\,du, \qquad s>0,$$

and

$$B(s,t)=\int_0^1 u^{s-1}(1-u)^{t-1}\,du,\qquad s,t>0,$$

are the well-known gamma and beta functions respectively. We illustrate our method by means of an example.

$$\sum_{n=1}^{\infty}\frac{1}{n\binom{2n}{n}}=\sum_{n=1}^{\infty}\frac{1}{n}\frac{n!n!}{(2n)!}=\sum_{n=1}^{\infty}\frac{\Gamma(n)\Gamma(n+1)}{\Gamma(2n+1)}$$
$$=\sum_{n=1}^{\infty}B(n,n+1)=\sum_{n=1}^{\infty}\int_0^1 u^n(1-u)^{n-1}\,du=\int_0^1 u\sum_{n=1}^{\infty}(u(1-u))^{n-1}\,du$$
$$=\int_0^1\frac{u}{u^2-u+1}\,du=\frac{\pi}{3\sqrt{3}}.$$

$\sum_{n=1}^{\infty}1/\binom{2n}{n}$ is handled in a similar fashion. For the series

$$\sum_{n=1}^{\infty}\frac{1}{n^2\binom{2n}{n}},$$

we arrive at the integral

$$-\int_0^1\frac{\log(u^2-u+1)}{u}\,du,\tag{2}$$

which appears in the study of the dilogarithm. Consulting the standard reference [3], we find that the integral (2) equals $2\mathrm{Li}_2(1,\pi/3)=\pi^2/18$.

A standard method for evaluating these series (see [2]) begins with the expansion

$$\frac{2x\sin^{-1}(x)}{\sqrt{1-x^2}}=\sum_{m=1}^{\infty}\frac{(2x)^{2m}}{m\binom{2m}{m}},\qquad |x|<1.$$

This formula can easily be derived from (1) by the same method we used to evaluate $\sum_{n=1}^{\infty}1/n\binom{2n}{n}$.

A typical series that this method can evaluate and that we have not seen previously is

$$\sum_{n=1}^{\infty}\frac{1}{n\binom{3n}{n}}.$$

This series leads to the integral

$$\int_0^1\frac{u^2}{u^3-u^2+1}\,du,$$

which can be evaluated explicitly with partial fractions. The final result is given in terms of the real zero of u^3-u^2+1, but the expression is unenlightening and too complicated to reproduce.

Our second remark is that if one considers not just series with the central binomial coefficient $\binom{2n}{n}$ in the denominator but series involving all possible entries $\binom{n}{k}$ in

Pascal's triangle, the problem becomes much easier.

To be specific, consider the series

$$\sum_{k=2}^{\infty} \sum_{n=k}^{\infty} \frac{1}{k^m \binom{n}{k}}, \qquad m \geqslant 2. \tag{3}$$

We neglect the terms involving $\binom{n}{0}$ and $\binom{n}{1}$ to insure convergence. Equation (1) and the above method give an easy proof that for $k \geqslant 2$,

$$\sum_{n=k}^{\infty} \frac{1}{\binom{n}{k}} = \frac{k}{k-1}. \tag{4}$$

If $m = 2$, the sum (3) therefore equals

$$\sum_{k=2}^{\infty} \frac{1}{k^2} \frac{k}{k-1} = 1.$$

If $m \geqslant 3$, we use partial fractions again to see that (3) equals

$$1 + \sum_{j=2}^{m-1} (1 - \zeta(j)),$$

where $\zeta(j) = \sum_{n=1}^{\infty} 1/n^j$. For $m = 0$ and $m = 1$ we must exclude more terms to have the series converge. The analogous results in these cases are

$$\sum_{k=2}^{\infty} \sum_{n=k+1}^{\infty} \frac{1}{k\binom{n}{k}} = 1$$

and

$$\sum_{k=2}^{\infty} \sum_{n=k+2}^{\infty} \frac{1}{\binom{n}{k}} = \frac{3}{2}.$$

REFERENCES

1. R. G. Bartle, *Real Analysis*, 2nd edition, Wiley, New York, 1976, 283.
2. D. H. Lehmer, Interesting series involving the central binomial coefficient, *Amer. Math. Monthly* 92 (1985), 449–457.
3. L. Lewin, *Dilogarithms and Associated Functions*, MacDonald, London, 1958, 4, 117.
4. A. J. van der Poorten, A proof that Euler missed, *Math. Intelligencer* 1 (1979), 195–209.
5. ______, Some wonderful formulas, *Queen's Papers in Pure and Appl. Math.* 54 (1979), 269–286.
6. I. J. Zucker, On the series $\sum_{k=1}^{\infty} \binom{2k}{k}^{-1} k^{-n}$ and related sums, *J. Number Theory* 20 (1985), 92–102.

Buffon's Needle Problem on Radial Lines

H. J. KHAMIS
Wright State University
Dayton, OH 45435

Duncan [1] has worked out a variation of the classical Buffon Needle Problem [2] whereby rays with uniform angular spacing $2\pi/n$ are drawn from the point O on a board, and a needle of length $2L$ and with midpoint fixed at point M is randomly thrown onto the board.

If $R=\overline{OM}$ and

$$L \leqslant R \sin \frac{\pi}{n}, \tag{1}$$

then the probability that the needle crosses a line, p, is given by

$$p=\frac{n}{\pi^2}\int_0^{\pi}\arctan\left(\frac{L\sin\phi}{R-L\cos\phi}\right)d\phi,$$

and an approximation is given by

$$p=\frac{n}{\pi^2}\log\left(\frac{R+L}{R-L}\right)+O\left(\frac{1}{n^2}\right).$$

See FIGURE 1. Note that for the case $L>R\sin(\pi/n)$, the needle may cross either or both rays forming the angle $2\pi/n$, which makes the problem much more complicated.

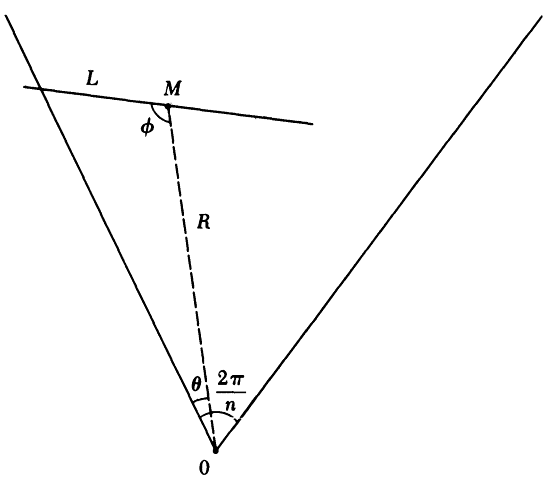


FIGURE 1
Needle of length $2L$ on set of radial lines.

This formulation of the problem involves a fixed distance, R. A problem of more interest perhaps, would be to assume that R is random such that $R \geqslant L\csc(\pi/n)$; see (1). Then, just as in Duncan's variation [1], we would let θ be the angle between OM

and the nearest line and let ϕ be the smaller angle between OM and that half of the needle that intersects the interior of the angle θ. The needle intersects one of the lines if and only if $R\sin\theta \leqslant L\sin(\theta+\phi)$ or, equivalently, if

$$\theta \leqslant \arctan\left(\frac{L\sin\phi}{R-L\cos\phi}\right).$$

We assume that θ, ϕ, and R are independent random variables, where θ is uniformly distributed on $[0, \pi/n]$ and ϕ is uniformly distributed on $[0, \pi]$.

Any realistic distribution may be attributed to the random variable R. It seems reasonable to assign a large probability to the event that the needle falls close to the point O (small R), with the probability decreasing exponentially as the needle gets farther from O (large R). The distribution describing this behavior, and one having nice limiting properties related to the classical Buffon Needle Problem, is the exponential distribution that is truncated according to (1):

$$f(R) = \frac{\lambda}{c_n}e^{-\lambda R}, \qquad R \geqslant L\csc\frac{\pi}{n},$$

with $c_n \equiv e^{-\lambda L\csc(\pi/n)}$. The parameter λ represents the reciprocal of the mean distance beyond $L\csc(\pi/n)$ attained by R.

Then, the probability that the needle intersects a radial line becomes

$$p_n = \frac{\lambda n}{c_n\pi^2}\int_{L\csc(\pi/n)}^{\infty}\left[\int_0^{\pi}\arctan\left(\frac{L\sin\phi}{R-L\cos\phi}\right)d\phi\right]e^{-\lambda R}\,dR. \tag{2}$$

For $n \geqslant 4$, we must have $L\sin\phi/(R-L\cos\phi) \leqslant 1$, so the Maclaurin series expansion for the arctangent term can be written

$$\arctan\left(\frac{L\sin\phi}{R-L\cos\phi}\right) = \frac{L\sin\phi}{R-L\cos\phi} + R_3,$$

where the remainder is

$$R_3 = \left(\frac{L\sin\phi}{R-L\cos\phi}\right)^3\left(3\delta^2\left(\frac{L\sin\phi}{R-L\cos\phi}\right)^2-1\right)\Big/3\left(1+\delta^2\left(\frac{L\sin\phi}{R-L\cos\phi}\right)^2\right)^3,$$
$$0<\delta<1.$$

Since

$$\left|3\delta^2\left(\frac{L\sin\phi}{R-L\cos\phi}\right)^2-1\right| \leqslant 2 \quad \text{for } n \geqslant 4,$$

we have

$$|R_3| \leqslant \left(\frac{L}{R-L}\right)^3\cdot\frac{2}{3} \leqslant \left(\frac{L}{R-L}\right)^3$$
$$\leqslant \left(\frac{\sin(\pi/n)}{1-\sin(\pi/n)}\right)^3 = O\left(\frac{1}{n^3}\right) \qquad \left(R \geqslant L\csc\frac{\pi}{n}\right).$$

Then, from equation (2),

$$p_n = \frac{\lambda n}{c_n \pi^2} \int_{L \csc(\pi/n)}^{\infty} \int_0^{\pi} \left(\frac{L \sin\phi}{R - L\cos\phi} + O\left(\frac{1}{n^3}\right) \right) e^{-\lambda R}\, d\phi\, dR$$

$$= \frac{\lambda n}{c_n \pi^2} \int_{L \csc(\pi/n)}^{\infty} \left(\log\left(\frac{R+L}{R-L}\right) + O\left(\frac{1}{n^3}\right) \right) e^{-\lambda R}\, dR.$$

And, using integration by parts (set u equal to the logarithmic term and dv equal to the exponential term), we get

$$p_n = \frac{n}{\pi^2} \log\left(\frac{1+\sin(\pi/n)}{1-\sin(\pi/n)} \right) - \frac{2\,nL}{\pi^2} e^{\lambda L \csc(\pi/n)} \int_{L\csc(\pi/n)}^{\infty} \frac{e^{-\lambda R}}{R^2 - L^2}\, dR + O(1/n^2). \tag{3}$$

Unfortunately, the answer involves a definite integral that must be approximated by a numerical quadrature routine.

A curious result is obtained upon taking the limit as n approaches infinity. By using L'Hôpital's Rule and the Fundamental Theorem of Calculus, the second term on the right side of (3) becomes zero and the first term becomes $2/\pi$. Hence,

$$\lim_{n \to \infty} p_n = 2/\pi.$$

This is just the probability that one obtains in the classical Buffon Needle Problem when the length of the needle is identical to the distance between parallel lines. This fact is easily verified by noting that $E(R) = L\csc(\pi/n) + 1/\lambda$, and $E(R) \approx nL/\pi$ as $n \to \infty$. Consequently, as n tends to infinity the set of radial lines approaches a set of parallel lines with uniform spacing that is approximated by the arc length $nL/\pi \cdot 2\pi/n = 2L$, which is just the length of the needle.

REFERENCES

1. R. L. Duncan, A variation of the Buffon Needle Problem, this MAGAZINE 40 (1967), 36–38.
2. B. V. Gnedenko, *Theory of Probability*, Chelsea, New York, 1962.

PROBLEMS

LOREN C. LARSON, *editor*
St. Olaf College

GEORGE GILBERT, *associate editor*
Texas Christian University

Proposals

To be considered for publication, solutions should be received by July 1, 1991.

1363. *Proposed by Daniel B. Shapiro, The Ohio State University, Columbus, Ohio.*

Let a, b be positive integers and perform Euclid's algorithm as follows, where $r_0 = a$ and $r_1 = b$:

$$\begin{aligned} r_0 &= r_1 q_1 + r_2 \\ r_1 &= r_2 q_2 + r_3 \\ &\vdots \\ r_{n-2} &= r_{n-1} q_{n-1} + r_n \\ r_{n-1} &= r_n q_n. \end{aligned}$$

Then $r_n = (a, b)$ is the greatest common divisor of a and b.

It is easy to show that $\sum_{j=1}^n r_j q_j = a + b - (a, b)$, and $\sum_{j=1}^n r_j^2 q_j = ab$. Show that if $g(x)$ is a polynomial such that the sum

$$\sum_{j=1}^{n} g(r_j) q_j \equiv S_g(a, b)$$

is a polynomial in a, b, and (a, b), then $g(x)$ is a linear combination of x and x^2.

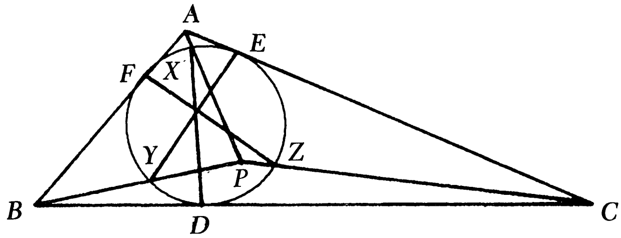
 We invite readers to submit problems believed to be new and appealing to students and teachers of advanced undergraduate mathematics. Proposals should be accompanied by solutions, if at all possible, and by any other information that will assist the editors and referees. A problem submitted as a Quickie should have an unexpected, succinct 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 mailed in duplicate to Loren C. Larson, Department of Mathematics, St. Olaf College, Northfield, MN 55057.

1364. *Proposed by Stanley Rabinowitz, Westford, Massachusetts.*

The incircle of triangle ABC touches sides BC, CA, and AB at points D, E, and F, respectively. Let P be any point inside triangle ABC. Line PA meets the incircle at two points; of these, let X be the point that is closer to A. In a similar manner, let Y and Z be the points where PB and PC meet the incircle respectively. Prove that DX, EY, and FZ are concurrent.



1365. *Proposed by Sidney H. Kung, Jacksonville University, Jacksonville, Florida.*

Prove that for $0 < a < b$,

$$\left(\frac{\sqrt{a}+\sqrt{b}}{2}\right)^2 < \frac{1}{e}\left(\frac{b^b}{a^a}\right)^{1/(b-a)}.$$

1366. *Proposed by Howard Morris, Chatsworth, California.*

Let $\sum_{i=1}^{\infty} a_i$ be a convergent series of positive real numbers, and x an arbitrary positive number. Show there exist uncountably many nondecreasing infinite sequences (b_n) of non-negative integers, such that $\sum_{i=1}^{\infty} a_i b_i = x$.

1367. *Proposed by John O. Kiltinen, Northern Michigan University, Marquette, Michigan.*

Let G be a group with the property that whenever H and K are subgroups of G, so is $HK = \{hk\colon h \in H \text{ and } k \in K\}$. Does it follow that every subgroup of G is normal? (If every subgroup of G is normal and G is nonabelian, then G is called a Hamiltonian.)

Quickies

Answers to the Quickies are on pages 66–67.

Q772. *Proposed by Irl C. Bivens, Davidson College, Davidson, North Carolina.*

Let E denote any set of n irrational numbers. Prove that there exists a nested family

$$E_1 \subset E_2 \subset \cdots \subset E_{n-1} \subset E$$

of subsets of E such that for each value of $j = 1, 2, \ldots, n-1$, E_j has order j and the sum of the elements of E_j is an irrational number.

Q773. *Proposed by Frank W. Schmidt, Bryn Mawr, Pennsylvania, and Rodica Simion, George Washington University, Washington, DC.*

Determine the largest value c for which the power series for e^{x-cx^2} has all its coefficients non-negative.

Q774. *Proposed by Murray S. Klamkin, University of Alberta, Edmonton, Alberta, Canada.*

If $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ are distinct coplanar vectors with equal lengths such that $\mathbf{A}\cdot\mathbf{B}+\mathbf{C}\cdot\mathbf{D} = \mathbf{A}\cdot\mathbf{D}+\mathbf{B}\cdot\mathbf{C}$, show that $\mathbf{A}\cdot\mathbf{B}+\mathbf{C}\cdot\mathbf{D}=\mathbf{0}$.

Solutions

Rate of convergence of a difference sequence **February 1990**

1338. *Proposed by Andrew Cusumano, Great Neck, New York.*
Evaluate

$$F(k) \equiv \lim_{n\to\infty} n^k\left[\left(1+\frac{1}{n+1}\right)^{n+1}-\left(1+\frac{1}{n}\right)^n\right].$$

I. *Solution by the Western Maryland College Problems Group, Western Maryland College, Westminster, Maryland.*

$$F(k) = \begin{cases} 0 & \text{if } k<2, \\ e/2 & \text{if } k=2, \\ \infty & \text{if } k>2. \end{cases}$$

By the Mean Value Theorem, for each $n>1$ there is a real number x with $n \leqslant x \leqslant n+1$ such that

$$\left(1+\frac{1}{n+1}\right)^{n+1}-\left(1+\frac{1}{n}\right)^n = \left(1+\frac{1}{x}\right)^x\left[\ln\left(1+\frac{1}{x}\right)-\frac{1}{x+1}\right].$$

Now

$$\begin{aligned} F(k) &= \lim_{n\to\infty} n^k\left(1+\frac{1}{x}\right)^x\left[\ln\left(1+\frac{1}{x}\right)-\frac{1}{x+1}\right] \\ &= \lim_{n\to\infty}\left(\frac{n}{x}\right)^k\left(1+\frac{1}{x}\right)^x x^k\left[\frac{(x+1)\ln\left(1+\frac{1}{x}\right)-1}{x+1}\right]. \end{aligned}$$

Noting that $(1+1/x)^x \to e$ and $(n/x)^k \to 1$, and setting $y=1/x$, we obtain

$$F(k) = e\lim_{y\to 0^+}\frac{1}{y^k}\left[\frac{\left(\frac{1}{y}+1\right)\ln(1+y)-1}{\frac{1}{y}+1}\right]$$

$$= e \lim_{y \to 0^+} \frac{(1+y)\ln(1+y) - y}{y^k + y^{k+1}}.$$

Applying L'Hôpital's rule an appropriate number of times (either once or twice), we find the answer given above.

II. *Solution by Jean-Marie Monier, Lyon, France.*

We have

$$\left(1+\frac{1}{n}\right)^n = e^{n\ln(1+1/n)} = e^{1-1/(2n)+1/(3n^2)+o(1/n^2)}$$
$$= e\left(1 - \frac{1}{2n} + \frac{11}{24n^2} + o\left(\frac{1}{n^2}\right)\right),$$

and, therefore,

$$\left(1+\frac{1}{n+1}\right)^{n+1} - \left(1+\frac{1}{n}\right)^n = e\left(\frac{1}{2n(n+1)} - \frac{11}{24}\frac{2n+1}{n^2(n+1)^2} + o\left(\frac{1}{n^2}\right)\right)$$
$$\sim \frac{e}{2n^2}, \text{ as } n \to \infty.$$

From this, we get the desired limit, as described in the preceding solution.

III. *Solution by David Callan, University of Bridgeport, Bridgeport, Connecticut.*

Rearranging the bracketed expression,

$$F(k) = \lim_{n\to\infty} n^k \left[\left(1 - \frac{1}{(n+1)^2}\right)^{n+1}\left(1 - \frac{1}{n+1}\right)^{-1} - 1\right]\left(1+\frac{1}{n}\right)^n.$$

Consider the related limit

$$G(k) = \lim_{x\to 0^+} \frac{\left[(1-x^2)^{1/x}(1-x)^{-1} - 1\right]}{x^k}.$$

The numerator here is

$$e^{(\ln(1-x^2))/x}(1-x)^{-1} - 1 = (1 - x + x^2/2 + \cdots)(1 + x + x^2 + \cdots) - 1$$
$$= \frac{x^2}{2} + O(x^3).$$

Thus, $G(k) = 0, 1/2, \infty$ accordingly as $k < 2$, $k = 2$, $k > 2$, respectively. Since $(1+1/n)^n \to e$ and $n/(n+1) \to 1$, we obtain $F(k) = eG(k)$.

Also solved by Robert A. Agnew, Barry Brunson, Paul F. Byrd, Onn Chan (student), Joseph E. Chance, Chico Problem Group, Con Amore Problem Group (Denmark), Fred Dodd, W. O. Egerland and C. E. Hansen, Doug Faires, Mordechai Falkowitz (Israel), H. Guggenheimer, Russell Jay Hendel, Joe Howard, Ole Jørsboe (Denmark), Hans Kappus (Switzerland), David W. Koster, Kee-Wai Lau (Hong Kong), Peter W. Lindstrom, I. E. Leonard and J. R. Pounder (Canada), Pui-Fai Leung (Singapore), Eugene Levine and Harry Ruderman, Stephen Noltie, Patrick Dale McCray, Reiner Martin, Sean Matz, S. L. Paveri-Fontana (Italy), Allan Pedersen (Denmark), Richard E. Pfiefer, Ray Rosentrater, Harry D. Ruderman, Heinz-Jürgen Seiffert (West Germany), Nick Singer, John S. Sumner and Kevin Dove, Nora S. Thornber, University of Wyoming Problem Circle, Xavier Mathematics Problem League, Michael Vowe (Switzerland), and Edward T. H. Wang (Canada). There were two incorrect solutions and one correct but unidentified solution.

Brunson notes that the enhanced version of *Mathematica*, running on a Macintosh IIcx, took five minutes to evaluate $f(1000)$ (≈ 1.355) and 19 minutes to evaluate $f(2000)$ (≈ 1.357), where $f(n)$ is the expression, with $k=2$, whose limit is desired. The function "Limit[·]" failed to converge and the kernel crashed, available memory exhausted. Single precision on a VAX 11/780 gave fluctuating positive and negative values for n around 1000.

There were several generalizations. *Ruderman* proved that $\lim_{n\to\infty} n^2\Sigma_{i=1}^{K} a_i(1+1/(n+\alpha_i))^{L(n+\alpha_i)} = (Le^L/2)\Sigma_{i=1}^{K} a_i\alpha_i$, where $a_1+a_2+\cdots+a_K=0$. The solution given by an unidentified source showed that if k is replaced with $k(n)=2+b/(\ln(n))+o(1/(\ln(n)))$, where b is an arbitrary fixed scalar, the resulting limit is $e^{b+1}/2$.

A triple congruence — February 1990

1339. *Proposed by George Gilbert, St. Olaf College, Northfield, Minnesota.*

Find all integer triples (x,y,z), $2\leqslant x\leqslant y\leqslant z$, such that

$$xy\equiv 1 \pmod z$$
$$xz\equiv 1 \pmod y$$
$$yz\equiv 1 \pmod x.$$

Solution by Brian D. Beasley, Presbyterian College, Clinton, South Carolina.

The only solution is $(x,y,z)=(2,3,5)$.

First, we note that x, y, and z must be pairwise relatively prime. Thus, $2\leqslant x<y<z$. Next, we combine the three equations to get

$$xy+xz+yz-1\equiv 0 \ (\text{mod } x,\ y, \text{ and } z).$$

Since x, y, and z are pairwise relatively prime, this implies

$$xy+xz+yz-1\equiv 0 \pmod{xyz}.$$

Hence, $xy+xz+yz-1=k(xyz)$ for some integer $k\geqslant 1$. Dividing by xyz yields $1/z+1/y+1/x=1/xyz+k>1$. For $2\leqslant x<y<z$, this implies $(x,y,z)=(2,3,4)$ or $(2,3,5)$. But 2 and 4 are not relatively prime, so the unique solution is $(x,y,z)=(2,3,5)$, as $1/5+1/3+1/2=1/30+1$.

Also solved by David Callan, Richard Carpenter, Onn Chan (student), Con Amore Problem Group (Denmark), George W. Dinolt, Fred Dodd, F. J. Flanigan, Kevin Ford (student), Lorraine L. Foster, John F. Goehl, Jr., Francis M. Henderson, David W. Koster, Peter W. Lindstrom, Helen M. Marston, Reiner Martin (student), Jean-Marie Monier (France), M. Parmenter (Canada), Allan Pedersen (Denmark), Mike Pinter, R. Bruce Richter (Canada), Harvey Schmidt, Jr., Heinz-Jürgen Seiffert (West Germany), Man Keung Siu and Kai Man Tsang (Hong Kong), John S. Sumner and Kevin Dove, Texas Academy of Mathematics and Science Problem Solving Group, William P. Wardlaw, Xavier Mathematics Problem League, Paul Yiu, and the proposer.

Ford notes that this problem appears as Problem #32, p. 292, in Hillman and Alexanderson, *A First Undergraduate Course in Abstract Algebra*, 4th edition, Wadsworth Publishing Company, Belmont, CA, 1988.

Constructible angles of prime degree — February 1990

1340. *Proposed by Richard L. Francis, Southeast Missouri State University, Cape Girardeau, Missouri.*

Show that the 3-degree angle is the only constructible angle of prime degree measure.

Solution by Stephen I. Gendler, Clarion University, Clarion, Pennsylvania.

We show more generally that any constructible angle of whole number degree is constructible if and only if it has measure that is a multiple of 3°.

First note that the sum, difference, and integral multiple of angle measures that are constructible are also constructible. Since $3° = 108° - (45° + 60°)$, all of which are well-known as the measure of constructible angles, it is also constructible, as are its multiples.

On the other hand, if any $n°$ angle were constructible, and 3 did not divide n, then the greatest common divisor of 3 and n would be equal to 1, so by the Euclidean algorithm, $1° = (3x)° + (ny)°$ would be constructible. Hence all integral measure angles could be constructed. But it is well-known that a 20° angle cannot be constructed, since it is impossible to trisect a 60° angle. Hence for any $n°$ angle that can be constructed, 3 divides n.

Also solved by S. F. Barger, David Callan, Stephen R. Cavior, Con Amore Problem Group (Denmark), Timothy V. Craine, Arthur H. Foss, Kevin Ford (student), Anthony Gaglione and William Wardlaw, Lee O. Hagglund, Man Keung Siu (Hong Kong), David W. Koster, Kee-Wai Lau (Hong Kong), Hugh Noland, Stephen Noltie, Bob Prielipp, Allan Pedersen (Denmark), Volkhard Schindler (East Germany), Gary E. Stevens, John S. Sumner and Kevin Dove, Texas Academy of Mathematics and Science Problem Solving Group, University of Wyoming Problem Circle, Xavier Mathematics Problem League, Dick Wood, and the proposer.

Labelings of planar graphs **February 1990**

1341. *Proposed by Bernardo Recamán, Universidad de los Andes, Bogotá, Colombia.*

a. Determine all positive integers n for which it is always possible to label the vertices of a planar graph on n vertices with the first n positive integers, so that adjacent vertices will have labels that are relatively prime.

b*. Is it always possible to label the vertices of a planar graph on n vertices with the first n *odd* positive integers, so that adjacent vertices will have labels that are relatively prime?

Solution by Jerrold W. Grossman, Oakland University, Rochester, Michigan.

a. Admissible labelings are always possible if and only if $n = 1, 2, 3, 5$. For $n \leqslant 3$ the situation is trivial. For $n = 4$ the complete graph on four vertices, K_4, cannot be labeled, since 2 and 4 are not relatively prime. For $n = 5$ we note that since K_5 is not planar, any planar graph on five vertices must contain at least one pair of nonadjacent vertices. We can label this graph by putting 2 and 4 on a pair of vertices not joined by an edge; all other pairs of numbers from 1 to 5 are relatively prime. Finally, for any $n \geqslant 6$ we can form a graph consisting of $\lfloor n/4 \rfloor$ disjoint copies of K_4, with the remaining three or fewer vertices (if any) in a separate complete component. This graph has $\lceil n/4 \rceil$ components in all, and there is room for only one even label in each. Since there are $\lfloor n/2 \rfloor$ even labels to be used, and since $\lceil n/4 \rceil < \lfloor n/2 \rfloor$ as long as $n \geqslant 6$, no labeling is possible.

b. No—for all n other than a few small exceptions, there are planar graphs on n vertices that cannot be admissibly labeled by the first n odd integers. Consider the graph consisting of $\lfloor n/4 \rfloor$ disjoint K_4's, with the remaining three or fewer vertices (if any) in a separate complete component; this graph thus has $\lceil n/4 \rceil$ components in all. In an admissible labeling, each component can have at most one multiple of 3; therefore at most $\lceil n/4 \rceil$ multiples of 3 can be used. But the first n odd positive integers contain $\lfloor (n+1)/3 \rfloor$ multiples of 3. Since the latter quantity exceeds the former for all $n > 13$, no admissible labeling is possible for any $n > 13$.

Also solved by Con Amore Problem Group (Denmark; part a); Allen J. Schwenk, John S. Sumner and Kevin Dove, and the Xavier Mathematics Problem League.

Sumner and *Dove* proved that labelings as in (b) need only exist when $n \leqslant 7$, $n = 9$, 10, or 13. *Schwenk* mentions the generalization to labeling from an arbitrary arithmetic progression, and suggests $1, 7, 13, 19, \ldots$ as a possibility for labeling all planar graphs.

Inverse of matrix with binomial entries **February 1990**

1342. *Proposed by Greg Fredricks and Harvey Schmidt, Jr., Lewis and Clark College, Portland, Oregon.*

Let $\mathbf{A} = (a_{i,j})$ be the upper triangular $n \times n$ matrix where

$$a_{i,j} = (-1)^{i+j}\binom{n-i+1}{j-i}$$

for $i \leqslant j$. Find the inverse of $\mathbf{A}$.

Solution by Barry Brunson, Western Kentucky University, Bowling Green, Kentucky.

The inverse of $\mathbf{A}$ is the matrix $\mathbf{B} = (b_{i,j})$ for which $b_{i,j} = |a_{i,j}|$ for every i and j. Consider the element of the product $\mathbf{AB} = (c_{i,j})$,

$$c_{i,j} = \sum_{k=1}^{n} a_{i,k} b_{k,j} = \sum_{k=i}^{j} (-1)^{i+k}\binom{n-i+1}{k-i}\binom{n-k+1}{j-k}.$$

Now use a combinatorial identity (which amounts to writing the binomial coefficients with factorials, factoring out factors that don't involve k, including half of a new $(j-i)!/(j-i)!$ factor), and introduce $m = k - i$, to get

$$c_{i,j} = \binom{n-i+1}{j-i}\sum_{m=0}^{j-i}(-1)^m\binom{j-i}{m},$$

and the summation is well-known to be zero unless $j = i$, in which case it is unity.

Remark on an appropriate use of technology: The above *confirms* an answer. Knowing what answer to try to confirm could be discovered using a symbolic algebra system to find and display both $\mathbf{A}$ and its inverse for n up to 8; the pattern is hard to miss.

Also solved by Brian D. Beasley, David Callan, Onn Chan (student), Joseph E. Chance, Chico Problem Group, Con Amore Problem Group (Denmark), David Doster, Robert Doucette, Ralph P. Grimaldi, Hans Kappus (Switzerland), Y. H. Harris Kwong, Peter W. Lindstrom, Norman F. Lindquist, Reiner Martin (student), Jean-Marie Monier (France), Volkhard Schindler (East Germany), Heinz-Jürgen Seiffert (West Germany), Nick Singer, Regis A. Smith, John S. Sumner, Gerald Thompson and Linda Key (student), University of Wyoming Problem Circle, William P. Wardlaw, Western Maryland College Problems Group, Michelle Wilson and Vincia Francis (students), Xavier Mathematics Problem League, and the proposers.

Comment

Q761. In Vol. 63, No. 2, April 1990, p. 133, *Murray Klamkin* offered the following generalization to his Quickie **(Q761)**: "For positive integers n and p and numbers $x_1, \ldots, x_n > 0$,

$$x_1^p + x_2^p + \cdots + x_n^p \geqslant x_1 x_2 \cdots x_p + x_2 x_3 \cdots x_{p+1} + \cdots + x_n x_1 \cdots x_{p-1},$$

where each subscript on the right is understood to be reduced modulo n to one of

$1, 2, \ldots, n$." *Peter D. Johnson, Jr., Auburn University*, offers the following generalization.

We start with the well-known rearrangement inequality (e.g., see *Inequalities*, Hardy, Littlewood, and Polya, Chapter 10): If $a_1 \leqslant \cdots \leqslant a_n$ and $b_1 \leqslant \cdots \leqslant b_n$, and π is a permutation of $\{1, 2, \ldots, n\}$, then $\Sigma_i a_i b_i \geqslant \Sigma_i a_i b_{\pi(i)}$.

For non-negative sequences this generalizes, by induction on p, to the following. If $p \geqslant 2$ and n are positive integers, and $\mathbf{A} = (a_{ij})$ is a $p \times n$ matrix of non-negative numbers with non-decreasing rows, then for any permutations $\pi_2, \pi_3, \ldots, \pi_p$ of $\{1, 2, \ldots, n\}$,

$$\sum_{j=1}^{n} a_{1j} a_{2j} \cdots a_{pj} \geqslant \sum_{j=1}^{n} a_{1j} a_{2\pi_2(j)} \cdots a_{p\pi_p(j)}.$$

Klamkin's inequality is obtainable from this by taking the rows of $\mathbf{A}$ to be equal (to a non-decreasing rearrangement of $x_1, \ldots, x_n$) and the permutations $\pi_2, \ldots, \pi_p$ to be successive powers of a certain cycle (π_2).

Answers

Solutions to the Quickies on p. 60.

A772. Clearly, it suffices to prove that in any set of $n \geqslant 2$ irrational numbers there is a subset of $n-1$ numbers with irrational sum. Suppose then there exists a set $E = \{x_1, x_2, \ldots, x_n\}$ of n irrational numbers such that the sum of any $n-1$ of the numbers is rational. Then the number $S = (n-1)(x_1 + x_2 + \cdots + x_n) = (n-1)[x_1 + (x_2 + \cdots + x_n)]$ is irrational since the number inside the brackets is the sum of an irrational number and a rational number. On the other hand,

$$S = \sum_{j=1}^{n} \left(x_1 + x_2 \cdots + \widehat{x_j} + \cdots + x_n \right)$$

is the sum of n rational numbers and consequently rational (the "hat" over the variable means the term is not included in the sum). This contradiction establishes the claim.

A773. *Solution* 1. The answer is $c = 0$. Suppose that $c > 0$. The power series for $f(x) \equiv e^{x - cx^2}$ has the form $1 + g(x)$. If all the coefficients of $g(x)$ are non-negative, then $g(x) \geqslant 0$ for all $x > 0$, and therefore $f(x) \geqslant 1$ for all $x > 0$. But this is impossible because $x - cx^2 \to -\infty$ as $x \to \infty$, and therefore for large x, $f(x) < 1$.

Solution 2. Suppose that all the coefficients in the power series of $f(x)$ are non-negative. Then the same is true of the power series of $f'(x)$, and therefore $f'(x) \geqslant 0$ for $x > 0$. But $f'(x) = e^{x - cx^2}(1 - 2cx)$. This expression shows that if $c > 0$, then $f'(x)$ is negative for large x, contrary to the previous conclusion. Thus, $c = 0$.

Solution 3. Suppose that $c > 0$. Complete the square in the exponent so that

$$e^{-c(x - 1/(2c))^2 + 1/(4c)} = \sum_{n=0}^{\infty} a_n x^n$$

with $a_n \geqslant 0$.

Let $y = x - 1/(2c)$ to get

$$e^{-cy^2} = e^{-1/(4c)} \sum_{n=0}^{\infty} a_n \left(y + \frac{1}{2c}\right)^n = \sum_{n=0}^{\infty} b_n y^n$$

with $b_n \geqslant 0$, since $e^{-1/(4c)} > 0$ and all expansion terms are positive.

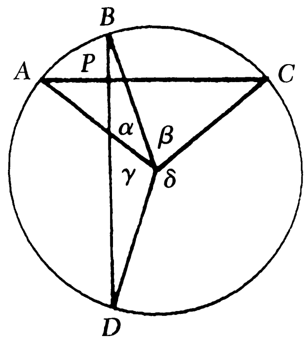
However, this is a contradiction because

$$e^{-cy^2} = 1 - cy^2 + \cdots$$

has a negative coefficient.

(Note: The same reasoning shows that if $p(x)$ is a polynomial such that $e^{p(x)}$ has non-negative coefficients and $p'(x_0) = 0$, with $x_0 > 0$, then $p''(x_0) > 0$.)

A774. If O is the origin of the vectors, then their endpoints A, B, C, D, respectively, lie on a circle centered at O. Since $(\mathbf{A} - \mathbf{C}) \cdot (\mathbf{B} - \mathbf{D}) = 0$, $AC \perp BD$. Hence, A, B, C, D are in consecutive order on the circle and do not lie on any semicircle.



Referring to the preceding figure, we have to show that $\cos\alpha + \cos\delta = 0$. Since $90° = \angle APB = \frac{1}{2}$ (arc AB + arc CD), $\alpha + \delta = 180°$ and we are done.

Conference on History, Geometry, and Pedagogy
At the University of Central Florida
Orlando, FL
In honor of the 80th birthday of Howard Eves

Howard Eves' career interests in teaching, history, and geometry provide an ideal setting within which mathematics teachers and university professors can discuss their experiences and research. It is a fitting tribute in the year that marks the 80th birthday of Howard Eves that a conference be organized which brings together representatives of these groups to discuss their common interest so each can learn from the perspectives of the others. Major speakers will include Professors Clayton Dodge, Peter Hilton, Murray Klamkin, Bruce Meserve, Fred Rickey, Marjorie Senechal, and, of course, Howard Eves. There also will be parallel sessions for contributed papers and workshops.

For more information, address inquiries to Professor Joby Anthony, Department of Mathematics, University of Central Florida, Orlando, FL 32816-6990. Phone (407) 823-2700 or FAX (407) 281-5156.

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.

Kolata, Gina, Solution to old puzzle: How short a shortcut? *New York Times* (30 October 1990) (National Edition) B6. Cipra, Barry, In math, less is more—up to a point, *Science* 250 (23 November 1990) 1081–1082.

The Steiner problem asks, What is the shortest network that connects a set of planar points? For some sets of points, adding "hub" points can reduce the length of the network. How low can the length go? Frank Hwang (Bell Labs) and Ding Xhu Du (Princeton) have shown that using hubs cannot reduce the length of the network by a factor of more than $1 - \sqrt{3}/2 \approx 13\%$. This specific factor comes from the situation of three points forming an equilateral triangle, with the center of the triangle added as a hub. The proof settles a 22-year-old conjecture of Henry Pollak and Edgar Gilbert of Bell Labs. The problem of how to pick the optimal hub points was shown to be NP-complete in 1977 by Ron Graham, Michael Garey, and David Johnson (also all from Bell Labs).

Pickover, Clifford A., *Computers, Pattern, Chaos and Beauty: Graphics from an Unseen World*, St. Martin's, 1990; xvi + 397 pp, $29.95. ISBN 0-312-04123-3

"Sometimes I consider myself a fisherman. Computer programs and ideas are the hooks, rods and reels. Computer pictures are the trophies and delicious meals." This book, "a catalog of the author's published works," is jam-packed with fascinating computer graphics arising from all kinds of areas: Fourier transforms, human speech, image processing of the Shroud of Turin, genesis equations, iterates of complex functions, 3-D strange attractors, exotic symmetries from Pascal's triangle, patterns in hailstone ($3n + 1$) numbers, spirals from nature, synthesizing ornamental textures, tessellation automata, and Chebyshev polynomials. Some simple algorithms are included, as are extensive references and suggestions for experiments. Sophisticated mathematical concepts and equations occur where they need to, but readers of varying mathematical background will each be able to marvel and take part in this feast.

Graham, Ronald L., and Joel H. Spencer, Ramsey theory, *Scientific American* (July 1990) 112–117.

"...Ramsey theory states that any structure will necessarily contain an orderly substructure." (A more general principle along these lines needs to be impressed on members of the U.S. public who are demanding that NASA investigate features on Mars that resemble human faces.) Ramsey theory, which began with a paper by Ramsey in 1925, languished until Esther Klein, a student at the University of Budapest, brought up the problem: "If five points lie in a plane so that no three points form a straight line, prove that four of the points will always form a convex quadrilateral." The problem generalizes to the conjecture that from $2^{k-2} + 1$ points in the plane, no three in a straight line, one can always select k points to form a convex k-gon. In an article with special appeal to the numerical and geometrical intuitions of the general educated public, Graham and Spencer go on to relate the further history of Ramsey theory and its connections with arithmetic progressions, tic-tac-toe, and

Ackermann functions.

Haskins, Loren, and Kirk Jeffrey, *Understanding Quantitative History*, MIT Press, 1990; xxv + 366 pp, $27.50. ISBN 0-262-08190-3

Reprints brief excerpts from quantitative papers on U.S. history, accompanied by extensive questions (with answers). The sections cover summarizing information, statistical arguments, and relationships among variables. The book (or a course based on it) is not a substitute for history students taking a statistics course. However, at some institutions, the mathematics department course in elementary statistics is accompanied by one-credit-hour applications seminars led by professors in applications areas, with different sections for different disciplines (the purpose of the seminar being to try to motivate students who don't feel they will need statistics); at many other institutions, the elementary statistics course is followed by a course in the applications discipline's research methods. This book would be *ideal* for either of those situations, for students interested in history or other social sciences. (The problem then becomes how to motivate prospective majors in history to enroll in the elementary statistics course in the first place.)

Berreby, David, Math in a million dimensions, *Discover* 11(10) (October 1990) 58–66.

Profile of Neil Sloane (Bell Labs), focusing on his work in sphere-packing, its connections to error-correcting codes, and his *Handbook of Integer Sequences*. Actually, Sloane has investigated packings in $10^{15,000}$ dimensions.

Stewart, Ian, *Game, Set, and Math: Enigmas and Conundrums*, Blackwell, 1989; viii + 191 pp, $19.95. ISBN 0-631-17114-2

In the French edition of *Scientific American*, entitled *Pour la Science*, the inspiration of Martin Gardner's "Mathematical Recreations" column has lived on, under the title "Visions mathématiques," with Ian Stewart as the columnist. This volume brings together a dozen of Stewart's columns (in his original English). Stewart tends more toward dialogues than Gardner did; but the columns are first-rate, including the figures. Too bad the American edition of *Scientific American* didn't pick up on Stewart; but at least we have this collection.

Kappraff, Jay, *Connections: The Geometric Bridge between Art and Science*, McGraw-Hill, 1991; xxi + 471 pp, (P). ISBN 0-07-034250-4, 0-07-034251-2 (pbk.)

An introduction to the interdisciplinary field of *design science*, whose origins the author attributes to Buckminster Fuller. Chapters cover proportion in architecture, similarity, the golden mean, graphs, tilings, networks and lattices, polyhedra (four chapters), isometries, and plane symmetries. The book is diverse, highly interesting, and well-illustrated. Though not written as a textbook (e.g., there are few exercises), it could be ideal for a course in liberal arts mathematics; the author also has written a supplementary manual of exercises, problems, projects, and instructor's guide.

Dewdney, A.K., *The Magic Machine: A Handbook of Computer Sorcery*, Freeman, 1990; xviii + 357 pp, 9 color plates, $23.95. ISBN 0-7167-2125-2

Second collection of Dewdney's splendid "Computer Recreations" columns from *Scientific American*. "Let this book then serve as a manual of magic, an entrée to the occult world of programming and to esoteric knowledge of the machine."

Schattschneider, Doris, *Visions of Symmetry: Notebooks, Periodic Drawings, and Related Works* *of M.C. Escher*, Freeman, 1990; xiii + 354 pp, \$39.95. ISBN 0-7167-2126-0

This is not just another Escher picture book, but a careful account of his analysis of symmetry and regular division, showing *how he did it.* Includes 180 illustrations that have never been published, with the complete set of numbered symmetry drawings and two Escher notebooks (1941-1942).

Senechal, Lester, ed., *Models for Undergraduate Research in Mathematics*, MAA, 1990; viii + 200 pp, (P). ISBN 0-88385-070-2

Contains descriptions of successful summer and academic year undergraduate research programs, together with students' accounts of the experiences and samples of papers that have resulted.

Sandefur, James T., *Discrete Dynamical Systems: Theory and Applications*, Oxford U Pr, 1990; xiii + 445 pp, \$39.95. ISBN 0-19-853384-5

Requiring only a minimum of calculus, this text in discrete dynamics is in effect a contemporary text on difference equations and their applications. A course based on it would make an intriguing follow-up to Calculus I, broadening students' exposure to mathematics much better than traditional Calculus II does. Included are Markov chains and phase plane analysis, as well as applications to genetics, harvesting, economics, and chaos. "After reading this text, you should be able to apply discrete dynamics to any field in which things change, which is most fields."

Banchoff, Thomas F., *Beyond the Third Dimension: Geometry, Computer Graphics, and Higher Dimensions*, Scientific American Library, 1990; ix + 210 pp. ISBN 0-7167-5025-2

Stunningly illustrated coffee-table book on geometry, treated through the theme of dimension. Of course, the fourth dimension is here; but also are configuration spaces in rehabilitation therapy, perspective and animation, and the origins of Egyptian and Mayan geometry.

Tucker, Thomas W., ed., *Priming the Calculus Pump: Innovations and Resources*, MAA, 1990; 321 pp, (P). ISBN 0-88385-067-2

Report of the CUPM Subcommittee on Calculus Reform and the First Two Years, including reports of NSF-funded efforts on 10 campuses (with sample materials), abstracts of projects at 100 more, and references on resources (graphing calculators, software, the history of calculus, and more). So, now, will anything change?

Fowler, D.H., *The Mathematics of Plato's Academy: A New Reconstruction*, Oxford U Pr, 1987; xxii + 401 pp. ISBN 0-19-853912-6

Places the study of continued fractions and ratio theories at the center of the development of Platonic mathematics, in an engaging study that features four Socratic dialogues composed by the author. Why do the historical sources not make clear that these were the central ideas of Greek mathematics? "[J]ust as Gauss rewrote number theory without continued fractions and thus may have prompted the decline in their study, so Eudoxus may have rewritten anthyphairetic and astronomical ratio theories as proporiton theory with the same effect; ... both worked with a deep understanding of what they were omitting."

NEWS AND LETTERS

Editorial Policy
(adapted from policy statement, this MAGAZINE, 54 (1981) 44–45)

Mathematics Magazine is a journal of collegiate mathematics designed to enrich undergraduate study of the mathematical sciences. Articles in the *Magazine* are intended to be used to supplement undergraduate courses or by instructors to enliven and supplement their lessons. They should be attractive and accessible to undergraduates.

Manuscripts submitted to the *Magazine* should be written in a clear and lively expository style. The *Magazine* is not a research journal, so papers written in a terse "theorem-proof-corollary-remark" style generally will be unsuitable for publication. The best contributions contain examples, historical background, and illustrations. Manuscripts on history are especially welcome, as are those showing relationships between various branches of mathematics and other disciplines. The editors welcome papers providing insight into applications of mathematics and the interplay of mathematics and computer science. Articles on pedagogy alone, unaccompanied by interesting mathematics, are not suitable. Neither are articles consisting mainly of computer programs unless these illustrate interesting mathematics. We encourage the contribution of anecdotes, quotations, cartoons, proofs without words, and mathematical games appropriate for end-of-article fillers.

Writing and Revising

Unlike a research journal, *Mathematics Magazine* is responsible first to its readers, then to its authors. The suitability of a manuscript for publication depends as much on the quality of its exposition as on the significance of its contents.

Our advice to authors is simple: Say something new in an appealing way or say something old in a refreshing way. But, say it intelligibly, presuming a minimum of technical vocabulary.

Good writing should be vigorous and informal, written in the active voice, and stocked with appropriate examples. Eschew computation; stress instead insight and motivation. Illustrate your ideas with appealing graphics: diagrams, tables, drawings, photographs. Good illustrations help convey ideas and enhance the visual appeal of the article.

The opening sentence or paragraphs of an article are the most important. These should provide an inviting introduction to the entire paper, explaining its importance, its purpose, and its appeal.

Effective transitions between different parts of your article will keep the reader's interest. Carefully chosen connective words can have a critical effect on the article's readability.

In choosing a title, authors should be brief, direct, and descriptive of the article's contents.

Referees for *Mathematics Magazine* are instructed to check for mathematical accuracy, but also to give detailed suggestions on stylistic matters. Most papers require redrafting by authors, followed by a thorough editing in our office.

Be generous in providing references for related reading since one of our objectives is to encourage readers to pursue ideas further. Bibliographies may contain suggested reading in addition to the sources actually used.

Authors may consult several excellent guides to mathematical exposition. These are listed as references at the end of this policy statement.

Format of Manuscript

ARTICLES are generally survey papers, longer than 2000 words. They should be sectioned into a few subsections with carefully chosen subtitles typed on lines above the paragraph they introduce. NOTES need not be sectioned; the normal connectives of English prose should be sufficient to maintain the logical flow of ideas in such a paper that is generally shorter than 2000 words.

Manuscripts should be typewritten, double-spaced with one-inch margins, with title, author, and author's address at the top of the first page. Pages should be numbered. Typewriter and printer ribbons should produce dark copy and paper (8 1/2″ × 11″) should be of medium to high quality. Do not staple pages together. Use paper clips.

Words being defined should be underscored by a black wavy line since *Mathematics Magazine* uses boldface for words being defined. List references either alphabetically or in the order in which they are cited in the text; consistency is essential with either choice. Authors should be listed with first name or initials followed by last name. Titles of books and journals should be underlined or italicized; all important words begin with a capital letter. Journal titles should be given in the abbreviated form found in *Mathematical Reviews*. Titles of journal articles should *not* be underlined or in quotation marks; only the first letter of the title of an article is capitalized. The following illustrations show our printed format for references.

2. J. L. Paul, On the sum of the kth powers of the first n integers, *Amer. Math. Monthly* 78 (1971), 271–272.

14. D. Hilbert and S. Cohn-Vossen, *Geometry and the Imagination*, Chelsea, New York, 1952.

A Reference Style Sheet is available from the *Magazine* office.

Illustrations should not be interspersed with text, but should be carefully prepared on separate sheets of paper in black ink—the original without lettering, and the copies with lettering added. Do not use ball-point pen. Fine-point nylon tip pens or artist's drawing pens are best for accurate reproduction. Each illustration should be numbered and referred to by number in the text.

Authors should submit the original manuscript and two copies (keeping one copy as protection against possible loss). Manuscripts and other correspondence should be sent to:

Martha J. Siegel, Editor
Mathematics Magazine
Towson State University
Towson, MD 21204

REFERENCES FOR AUTHORS

1. *A Manual for Authors of Mathematical Papers*, 8th edition, American Mathematical Society, 1984.
2. R. P. Boas, Can we make mathematics intelligible, *Amer. Math. Monthly* 88 (1981), 727–731.
3. Harley Flanders, Manual for *Monthly* authors, *Amer. Math. Monthly* 78 (1971), 1–10.
4. Leonard Gillman, *Writing Mathematics Well*, MAA, 1987.
5. N. E. Steenrod, P. R. Halmos, M. M. Schiffer, and J. A. Dieudonné, *How to Write Mathematics*, American Mathematical Society, 1973.
6. Ellen Swanson, *Mathematics into Type*, revised edition, American Mathematical Society, 1979.

CONTENTS

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS

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